Eye-tracking multi-word units: some methodological questions

Gareth Carrol
University of Nottingham

Kathy Conklin
University of Nottingham

Eye-tracking in linguistics has focused mainly on reading at the level of the word or sentence. In this paper we discuss how the phenomenon of formulaic language might best be examined using this methodology. Formulaic language is fundamentally multi-word in nature, therefore an approach to eye-tracking that considers the "word" as the basic unit of analysis may require re-evaluation. We review the existing literature on single word and sentence processing, and also those studies that have used eye-tracking as a way of investigating formulaic language to date. We discuss how eye-tracking might elucidate the "added extra" processing advantage for formulaic language. We conclude with some suggestions about the best way to utilise eye-tracking within this sub-field of linguistic investigation.

Introduction

When is a word not a word?

Eye-tracking has provided an invaluable tool in the armoury of the modern psycholinguist. For those concerned with the structure of the mental lexicon, it provides an online way to examine how words are recognised, processed and integrated into sentence structures, and to explore the various factors that affect these processes such as frequency, length, ambiguity and other variables. Eye-tracking has therefore been essential for our models of single word processing, but, as pointed out by Clifton, Staub and Rayner (2007), as the length of critical regions of interest increases, it becomes much harder to see precisely where an effect might occur within that region. For this reason, it is less straightforward to use eye-tracking for investigating "formulaic language" – sequences of more than one word that behave like "single choices" (Sinclair, 1991). A good example is an idiom like *kick the bucket*, which is made up of three lexical items but which has a single "formulaic" meaning, namely "die". Despite the fundamentally multi-word construction of such items, both semantically and

syntactically they can be considered as single units (*kick the bucket* would be better analysed as a single intransitive verb than as a sequence of verb + object), and many models of formulaic language reflect this "whole form" storage (c.f. Wray, 2002; Wray & Perkins, 2000; Van Lancker Sidtis, 2012). This leads to the question of what our unit of analysis for formulaic language should be. The following discussion explores the notion of the "word" and proposes different approaches that we could adopt in eye-tracking research based on those studies that have so far used this methodology to investigate multi-word units (MWUs).

The theoretical basis of eye-tracking as an approach to linguistic investigation is generally quite straightforward. As with other methods such as measurement of reaction times to a given stimulus, eye-tracking considers the amount of time spent on an item to be a reflection of the cognitive effort required to process it. Two assumptions are key to this: a principle of immediacy/incremental processing as each lexical item is encountered, and some degree of eye-mind equivalence, whereby it is assumed that what is being looked at is what is being processed (Pickering, Frisson, McElree & Traxler, 2004, but see also the discussion within this pa-

per relating to how higher-level processes can call this assumption into question in certain contexts). Although different models of eye-movement control in reading vary in their predictions about specific features such as serial vs. parallel allocation of attention (see, for example, the predictions of the E-Z Reader (Reichle, Rayner & Pollatsek, 2003) and SWIFT (Engbert, Nuthmann, Richter & Kliegel, 2005) models), one common theme is that the analysis generally considers the "word" as the primary unit of analysis. Fixations (or skipped fixations) are assigned to a single lexical item, and measurements have traditionally been separated into "early" indicators – metrics like first fixation duration, first pass reading time/gaze duration and likelihood of skipping a given word – which are often taken to be a reflection of automatic processes, and "late" measures – total reading time, total number of fixations and re-reading patterns – which can be seen as largely reflecting the more strategic, controlled processes involved in reading comprehension (Altarriba, Kroll, Scholl & Rayner, 1996; Inhoff, 1984; Paterson, Liversedge & Underwood, 1999; Staub & Rayner, 2007). This preference for treating each word as an individual unit of analysis is justified in Pickering et al. (2004), who suggest that long regions of interest are problematic for several reasons, not least that early effects such as first pass reading time become harder to interpret. The authors state that "our preference has always been to define one word critical regions where possible. Under such conditions, first-pass time, like first-fixation time, is spatially well-localized." (p.5).

Such an approach presupposes one key aspect: that the identification of a "word" is a simple process. However, as argued by Reichle, Liversedge, Polatsek and Rayner (2009), amongst others, this seemingly straightforward assumption can be deceptively hard to implement. They adopt a working definition of a word as "any sequence of letters that are separated by spaces and that have an accepted pronunciation and meaning in the language" (p.116), but take pains to point out the potential pitfalls for languages other than English where orthographic conventions might make it much harder to identify clear boundaries in this way. A further objection to this definition of the word is taken up by Cutter, Drieghe and Liversedge (2014), who propose that an approach based on this definition of the "basic lexical unit" (p.1778) is potentially vastly underspecified when we consider those items that are considered under the broad heading of formulaic language. This echoes a recent discussion by Wray (2014), who asks how we can even be sure that we know what a "word" is, and who further argues that any vagueness in our definitions reflects the inherent problem that orthography imposes boundaries that do not always reflect any psychological validity.

Such calls for a rethink on how we might best describe a "word" are in themselves reflections of the position taken by multiple researchers within the field of formulaic language, where strong evidence has been presented for the representation of (semi) fixed sequences as single entries that are retrieved directly from the mental lexicon (c.f. Arcara et al., 2010; Libben & Titone, 2008; Rommers, Dijkstra & Bastiaansen, 2013; Sprenger, Levelt & Kempen, 2006; Titone & Connine, 1999; Titone, Columbus, Whitford, Mercier, & Libben, 2015). Especially in the case of idioms, a common feature of models is a "hybrid" approach, whereby a familiar phrase exists both as a whole unit and as the individual component parts. For example, Sprenger et al. (2006) proposed that idioms exist as "superlemmas" – lexical-conceptual entries that represent the phrase as a whole but which are reciprocally linked to each of the component words. So in an example like *kick the bucket*, each individual word would exist as a single lexical item (*kick, the, bucket*) but a whole form "superlemma" for *kick the bucket* would also exist at some level of representation; this is linked to each component word and also to the phrase level meaning ("die"). In this way idioms display a high level of internal constituency and flexibility but can also be characterised as whole units that are retrieved/accessed directly.

Given that idioms, and other forms of formulaic language, may therefore be represented at some level, an important question is how can we use eye-tracking to investigate the processing of such linguistic forms? The intention of this article is to discuss how eye-tracking might best be utilised to explore this. To begin, we briefly review the existing literature on the lexical and contextual factors that have been investigated to date in eye-tracking research to see what each might tell us about the processing of formulaic language.

What can single word processing tell us about formulaic language?

Cutter et al. (2014) suggest that (ease of) lexical processing is the main determinant of when the eyes move from one word to another. Or, as Clifton et al. (2007, p.

348) put it, "how long readers look at a word is influenced by the ease or difficulty associated with accessing the meaning of the word", and this is an effect that emerges most clearly in early measures. Staub and Rayner (2007, p. 330) outline the "intrinsic lexical factors" that affect the reading of individual words. Frequency is a primary determinant of fixation duration (Rayner & Duffy, 1986; Inhoff & Rayner, 1986) and likelihood of skipping (Rayner, Warren, Juhasz, & Liversedge, 2004), but in addition morphological structure (Andrews et al., 2004; Pollatsek, Hyona & Bertram, 2000; Juhasz, Starr, Inhoff & Placke, 2003) and meaning ambiguity leading to competition between lexical representations (Duffy, Morris & Rayner, 1988; Sereno, O'Donnell & Rayner, 2006) both show significant effects on single word reading patterns.

One of the main considerations here is the way in which formulaic language complicates many of these factors. Single word frequency is undoubtedly important, but for multi-word units we might also usefully consider whole phrase frequency and corpus-derived metrics such as mutual information (a measurement of observed unit frequency compared to the expected co-occurrence based on the individual word frequencies and the size of the sample they appear in) or transitional probability (the likelihood of seeing word B once word A has been encountered). It is clear that any given word can become significantly easier to process when it is used as part of a formulaic sequence (c.f. Conklin & Schmitt, 2008; Gibbs, 1980; Libben & Titone, 2008; Swinney & Cutler, 1979; Tabossi, Fanari & Wolf, 2009). This occurs despite the fact that idioms often use low frequency words (e.g. *bury the hatchet*), sometimes display non-standard morphology (e.g. *toing and froing*), can be inherently ambiguous (*drop the ball*), and often demonstrate highly context-specific meanings (e.g. *spill the beans*, where *beans* acquires a specific figurative meaning that is not assigned to it in any other context). When investigating formulaic language, other factors not relevant to individual words must also be taken into consideration. For example, previous studies on single word processing have generally shown unreliable $n + 2$ preview effects (benefit derived from a parafoveal preview of the word two words further on from the point of fixation); when such effects exist they are generally limited to sequences where both n and $n + 1$ are very short and highly frequent (Kligel, Risse & Laubrock, 2007; Radach, Inhoff, Glover, & Vorstius, 2013). However, a recent study by Cutter at al. (2014)

investigating spaced compounds provided what they considered to be "one of the strongest pieces of evidence thus far in favour of MWUs [multi-word units] having unified lexical entries" (p.1784). They found an $n + 2$ preview benefit, demonstrated in shorter fixation times for word $n + 1$, when $n + 1$ and $n + 2$ were constituents of a spaced compound (e.g. *teddy bear*), which they took as evidence that both words were being processed as part of a larger MWU. Crucially, $n + 2$ effects were only seen when $n + 1$ "licensed" the whole form, leading to an advantage that was not seen for any other combination (when either $n + 1$, $n + 2$ or both were nonwords). Cutter et al. (2014) argued that the increased length and lower frequency of the $n + 1$ items in their study (compared to previous investigations) was evidence of this effect being driven by lexical rather than perceptual factors.

Juhasz, Pollatsek, Hyönä, Drieghe and Rayner (2009) also found $n + 2$ preview effects for spaced compounds as well as for novel adjective + noun combinations; they suggested that for their stimuli the high syntactic predictability of the final noun was responsible for the effect in both spaced compounds and novel pairs. However, Cutter et al. (2014) argued that the predictability of word $n + 2$ was not on its own a good explanation for their results: $n + 2$ only became strongly predicted once $n + 1$ had been seen, meaning that $n + 1$ would have to be fully identified and integrated during fixations on word n if predictability was driving the effect. A similar finding emerged from a study by Siyanova-Chanturia, Conklin and van Heuven (2011), who looked at reading times for binomials: sequences of *X and Y* where one order of components is strongly preferred (e.g. *bride and groom*). They found an advantage for binomials over their corresponding reversed forms (e.g. *groom and bride*) that was not solely attributable to predictability (as measured by a phrase completion test). They concluded that the processes involved in speeded reading of the binomials reflected something over and above simple predictability, and that the phrasal configuration itself played a crucial role.

Clearly predictability is a key component of the formulaic advantage. Previous research on predictability for single words has shown strong effects in terms of shorter first fixation durations and greater likelihood of skipping for more predictable words (Ashby, Rayner & Clifton, 2005; Rayner & Well, 1996), but formulaic language seems to show some level of "added extra" that goes beyond simple predictability. The question is therefore how

eye-tracking might best be used to reveal the mechanism underlying this.

Cutter et al. (2014) do a good job of demonstrating how eye-tracking can usefully be applied to MWUs such as spaced compounds, but longer formulaic items would present considerably more of a challenge. Even for idioms of the common form V-det-N (*kick the bucket, spill the beans, chew the fat*), the presence of the determiner and the consequent extension of the unit to three words immediately raises the question of what we should be treating as our unit of analysis. The few studies that have used eye-tracking to look at idioms have broadly taken the same approach; that is, an idiom (e.g. *a pain in the neck*) is compared to a control phrase (e.g. *a pain in the back*) and the reading times are compared, either for the phrase as a whole or specifically for the final word (Siyanova-Chanturia et al., 2011; Underwood, Schmitt & Galpin, 2004). This line of enquiry is an extension of other methodologies that have compared formulaic and novel phrases through, for example, phrase acceptability judgements (Swinney & Cutler, 1979; Tabossi, Fanari & Wolf, 2009) and self-paced reading studies (Conklin & Schmitt, 2008; Libben & Tione, 2008). The advantage offered by eye-tracking is that both phrase level and word level patterns can be examined in the same study. In this way Siyanova-Chanturia et al. (2011) were able to analyse idioms in terms of both whole phrase reading and sub-part reading (before and after the recognition point or "idiom key" – Cacciari & Tabossi, 1988). They found an advantage for idioms (e.g. *at the end of the day*) vs. controls items (e.g. *at the end of the war*) for whole phrase reading times in late measures but not early measures, and found no effects for sub-part analysis for native or non-native speakers. Other studies (e.g. Underwood et al., 2004) have found patterns at the single word level. Carrol and Conklin, (submitted) found facilitation for both whole phrase and final word reading of idioms compared to control phrases. Native speakers were more likely to skip the final word or read it more quickly, and overall spent less time reading and re-reading the whole phrase for familiar idioms. In a separate study, Carrol and Conklin (in preparation) found that short idioms of the form V-det-N show a robust advantage for the whole phrase (e.g. *spill the beans* vs. *drop the beans* vs. *spill the chips*) in terms of first pass and total reading times. Further examination showed that the locus of this advantage is clearly the final noun, which is read more quickly in the idiom condition for all measures.

The discrepancy between Siyanova-Chanturia et al. (2011) finding effects only for the whole phrase and other studies finding effects for specific words underlines the need to adopt an approach that captures both the macro and micro features of formulaic units. An additional argument for such a dual approach is that it provides a way to accommodate skipping behaviour into analyses. Traditionally, duration measures on single words are only considered for those items that are not skipped during first pass reading. For formulaic items, however, this means actively removing a substantial portion of the items that most clearly demonstrate the expected effect. For example, in Carrol and Conklin (submitted), native speaker showed a tendency to skip the final words of idioms around 35% of the time (e.g. *seat* was often skipped in *on the edge of your seat*) compared to less than 10% for control phrases (e.g. *on the edge of your chair*). Removal of skipped items would therefore impact the idioms much more than other items, leading to an imbalance in the data for any subsequent analyses. Crucially, this would also mean that the clearest examples of the idiom advantage would be discounted from any further durational analysis. One solution, therefore, is to consider both word level measures (skipping rates, then duration measures for non-skipped words) and phrase level measures (duration measures for all items), thereby capturing the full range of behaviour. So for an example like *on the edge of your seat*, analysis of the word level measures may be limited (if *seat* is skipped then no further durational analysis is possible), but the overall phrase level reading times would still be informative across a range of measures, allowing for direct comparison with reading times for non-idiom control phrases. Of course, a notable practical consideration is the increased analysis time that such an approach necessitates, especially if multiple eye-tracking measures are used, but it seems that such a trade-off may well be worthwhile as a way of accounting for formulaic processing in as much detail as possible. Certainly skipping rates should form part of any word level-analysis, hence a method that allows for their inclusion alongside other word and phrase level measures is essential.

The evidence discussed above is relatively clear in demonstrating formulaicity, i.e. there is a consistent advantage on a range of measures for idioms, and often the final words in particular, that can perhaps be best explained through their status as part of a formulaic unit. This is especially the case for short items (e.g. V-det-N idioms, binomials or simple two word combinations such

as collocations or spaced compounds), where any unequivocal recognition point is not reached until all words have been seen. This is not to say that a whole unit/direct retrieval explanation is prohibitively implicated, and several alternative explanations are plausible (notably a lexical priming mechanism, similar to that proposed by Hoey, 2005). A key question is therefore how we might best utilise eye-tracking to differentiate potential mechanisms of formulaic processing; clearly a fairly nuanced method of analysis is required if we are to distinguish whole form access from, for example, lexical priming or fast serial mapping of formulaic components (Wray, 2012).

An important conclusion is that those measurements that are typically used for single words (as delimited by orthographic considerations) may not necessarily scale up to formulaic units in a simple fashion. Additional variables that take into account the phrasal nature of such units (based on frequency and cohesion) might therefore be usefully included, as well as semantic considerations like transparency and decomposability (to what degree each of the component words contributes to the idiomatic meaning). To this end it seems logical to consider phrasal variables in the design or analysis of any eye-tracking investigation of formulaic language as a way of capturing this specifically phrase level behaviour.

What can syntactic and global discourse context tell us about formulaic language?

The syntactic structure in which a word appears has also been widely investigated in the eye-tracking literature. A basic assumption is that when reading, the natural approach is to produce a word-by-word analysis of the syntactic structure as each word is encountered (the incremental processing assumption highlighted in Pickering et al., 2004). Syntactic ambiguity, therefore, has been the focus of much research, but Staub and Rayner (2007) summarise that very few, if any, studies have demonstrated that such structural competition leads to any cost in terms of reading times. Note that this stands in clear contrast to studies of meaning ambiguity, where lexical competition shows an unequivocal cost in terms of longer fixation durations (as summarised by Clifton et al., 2007). Overall then, it seems that the mechanisms that contribute to sentence level reading behaviour are not the same (or at least not as straightforward) as those that control single word reading. The importance of this to formulaic language is paramount, since often a word-by-word analysis

is likely to provide an incorrect interpretation (e.g. for idioms such as *kick the bucket*). Arguably a word-by-word analysis of such items would present both a semantic and syntactic incongruity which would require reassessment to resolve.

At a global discourse level, there seems to be an effect primarily in later measures of the coherence or otherwise of the overall discourse context, for example, resolution of anaphoric reference or completion of complex inferences within a multi-sentence text (Garrod, O'Brien, Morris, & Rayner, 1990; Myers, Cook, Kambe, Mason, & O'Brien, 2000; O'Brien, Shank, Myers, & Rayner, 1988; Sturt, 2003). Some studies have looked at the global context more in terms of overall meaning, and the conclusion reached by, amongst others, Camblin, Gordon and Swaab (2007) is that global discourse context overrides any local, lexical effects when a rich enough context is provided. Thus, only when an absent or impoverished context is provided do lexical effects such as semantic relatedness emerge. In their study, Camblin et al. (2007) found that effects of disrupted global context were early to emerge and long lasting, as evidenced by significant effects in first pass reading time for a manipulation of the discourse context. When global discourse context was not influential (when it was impoverished or incongruous), low level semantic links showed an effect in terms of shorter reading times for semantically related words within a sentence.

One advantage of eye-tracking is that we can easily insert words into a variety of wider contexts to compare reading patterns. Semantic predictability of specific words as a result of preceding context has been shown to be a strong determinant of reading times (Ehrlich & Rayner, 1981; Frisson, Rayner, & Pickering, 2005; Rayner et al., 2004), with words that are strongly predicted or highly constrained showing considerably shorter reading times as well as a higher likelihood of being skipped. Conversely, words that are semantically anomalous (and by definition therefore have low predictability) show inflated reading times (Murray & Rowan, 1998; Rayner et al., 2004). The predictability of formulaic units is, however, not entirely a function of the preceding discourse context: many studies of idioms presented in isolation have shown that the minimal lexical context provided leads to faster processing of the final word compared to a control (e.g. Carrol & Conklin, 2014, where seeing the isolated prime phrase *on the edge of your…* led

to faster lexical decisions for *seat* than for a control word like *plate*). Underwood et al., (2004) showed that terminal words of formulaic sequences were read more quickly and with fewer fixations than the same words used in non-formulaic contexts, so it is clear that idioms (and specifically the highly predicted final words) are undoubtedly read more quickly and fixated less often than either control phrases or the same words used in non-formulaic contexts. Crucially, this is not driven by global discourse context in the way that semantic expectancy would be.

It seems that context, whether syntactically defined or whether it is provided by a more global discourse mechanism, shows effects that usually emerge in later eye-tracking measures. What is important when dealing with formulaic language is that we have to balance the local, lexical context provided by a very specific combination of words and the global discourse context that might lead a reader to expect a semantically congruent lexical item, whether this is a single word or a formulaic unit. In this sense, using the hybrid models of idiom representation as our guide might represent the best approach, where *the whole is greater than the sum of the parts*. Taking a holistic view of the phrase allows us to examine its behaviour as a whole, while analysis of the individual words (and in particular those that occur later in the sequence) might reveal more about precisely what is being activated; "hybrid" is therefore an appropriate label for such analysis, since it actively combines the most useful elements of two different approaches. In some ways this echoes the overall conclusion reached by Staub and Rayner (2007) that models of naturalistic reading do a good job of accounting for the many lexical factors (length, frequency, predictability, etc.) that affect eye movements, but that higher level factors are to some degree under-explored. They suggest that the lexical factors should be considered as the "primary engine" (2007, p. 336), and that higher level structural or discourse considerations will typically exert a later influence, for example in re-reading behaviour or total reading times when additional attention is required to make sense of a problematic text. (It is noteworthy, however, that results from Camblin et al. (2007) outlined previously argue in the opposite direction, suggesting that global features will very often override any lexical level effects.) Again, the conclusion is that using only single words as our base units of analysis in eye-tracking is likely to pose problems and will not necessarily tell us much about how formulaic language is parsed and processed in real time.

To summarise the issue thus far, eye-tracking as a way of investigating the form of idioms and other multi-word units is not necessarily a straightforward process: there is something of a paradox inherent in the analysis of "whole units" through segmentation into component words, while to treat them only as single units is to eliminate the fine grained detail that eye-tracking can provide (and to ignore much of the evidence demonstrating the internal constituency of such units, c.f. Konopka & Bock, 2009; Sprenger et al., 2006). The multi-word space that idioms take up means that the traditional early measures become less reliable on a whole phrase level; at the same time, only utilising later measures would obscure the involvement of the automatic, intralexical processes that are also of interest.

Phrasal meaning and formulaic language

We have so far considered processing primarily in terms of form, but a second aspect of formulaic language particular to idioms is their meaning (e.g. "die" for *kick the bucket*). Thus, we can also ask to what degree a figurative meaning is activated (as opposed to incremental activation of the literal meanings of component words) and how might eye-tracking be used to explore this? In this regard it seems logical that later measures, broadly reflecting meaning integration, should be more important, i.e. the pattern of overall reading times alongside regression paths/refixation times should be most important in establishing how well any given sequence has been understood within a sentence. Especially in the case of idioms, which presumably require their own semantic entry (Wray, 2012), a clear pattern should emerge for those items that are understood easily within a given context and those which are not (less transparent, less well known idioms). In this sense, effects should be comparable to those seen for single words. Results summarised by Clifton et al. (2007) regarding lexical ambiguity show that if disambiguating information encountered after an ambiguous word demonstrates that a subordinate meaning was intended, the result is significant disruption to reading (in the form of longer fixations and regressions) as a reflection of the reanalysis that is required. Similarly, Rayner et al. (2004) showed early effects for words that were semantically anomalous, but for words that were merely implausible the effects only emerged in later measures. If formulaic sequences are therefore treated as single units, we would expect similar patterns to emerge.

One study to look at this is Siyanova-Chanturia, Conklin and Schmitt (2011), who compared the reading times of figurative and literal uses of ditropic idioms (idioms that can plausibly have a literal and a figurative meaning, such as *at the end of the day*). They found that for native speakers there were no differences on any measures for the two meanings: both were read more quickly than a control phrase (*at the end of the war*) but neither was fixated fewer times or read more quickly in early or late measures than the other. Non-native speakers, on the other hand, showed a clear advantage for literal uses. Importantly, this was observed only in the later measures (total reading time and number of fixations), with first pass reading time showing no difference between a figurative use, a literal use or a control phrase. It seems clear here that the overall reading time, including the amount of time spent in revisiting material, is a fairly robust measure of how easily an idiom has been understood in the wider context, with more problematic (less compositional) material requiring greater consideration and cognitive effort. In support of this, Carrol, Conklin and Gyllstad (in preparation) examined bilingual idiom processing using eye-tracking of simple sentences containing either English idioms or translations of Swedish idioms. The intention was to explore how Swedish-English bilingual speakers read the translated forms, but of particular interest here is the behaviour of the English native speaker controls used in the study. When reading translations of Swedish idioms (e.g. *to make a painting*, meaning "to make a mistake"), English native speakers showed no difference compared to controls for early measures such as first pass reading time, either for the whole phrase (*make a painting* vs. *sell a painting*) or for the final word (*painting* in the two conditions), however the whole phrase showed significantly longer total reading times and total number of fixations in the idiom condition than the control condition. This is an entirely logical finding, since the idiom should be nonsensical to English speakers while the control was wholly compositional, but it is noteworthy that this was reflected only in the later measures, with early measures showing no effects of the unknown phrase (lexical access/recognition of the component words did not seem to be problematic, but integration of the intended phrasal meaning was).

Overall, these results seem to support the view of formulaic sequences as whole units (or at least as individual choices/meaning units), since the effects seen for both unknown idioms and implausible single words are comparable. The analysis of whole phrase reading in terms of meaning integration certainly seems to be more suited to late measures, and analysis of regions before and after the idiom might also be a useful way to approach this. For example, as well as the total time spent reading an idiom itself, how much do readers then need to return to the prior context in an attempt to integrate the meaning, or how much time is spent reading a following disambiguating region in the case of literally plausible items? Titone and Connine (1999) analysed idioms and the following disambiguating region and found that results differed according to whether the idiom was more or less decomposable: when literal and figurative analysis of the idiom overlap, meaning integration is facilitated, whereas when the results of literal and figurative analysis differ (for non-decomposable idioms) this process is more difficult, and costs are seen both in terms of idiom reading times and increased reading times for following regions. Ciéslicka, Heredia and Olivares (2014) examined idiom processing in English-Spanish and Spanish-English bilinguals. Their results showed that idioms and post-idiom regions were affected by language dominance and contextual support. Total reading times for both idiom and post-idiom regions were shorter for English dominant participants and when context supported figurative meanings, and re-reading patterns for the idioms also demonstrated this effect. Overall, this study suggests that salience and context – key factors in allowing a reader to integrate the intended figurative meaning – are modulated by language dominance, and the effects were seen chiefly in late measures. This supports a view whereby formulaic sequences can be largely equated with single words, at least in terms of how they are understood in any given context. It therefore seems logical that, just as for single words, late measures like total reading time, total number of fixations and regression patterns should be the chief way of examining the dimension of meaning.

There is also a need to accommodate those idiom theories that posit automatic activation of the literal meanings of component words as an obligatory part of idiom comprehension (c.f. Cieslicka & Heredia, 2011; Holsinger & Kaiser, 2010; Sprenger et al., 2006; Titone & Connine, 1999). One clue to resolving this may come from the literature on compounds (both spaced, as in Cutter et al. (2014) discussed earlier, and non-spaced, e.g. *newspaper*). Ample evidence suggests that English compounds are decomposed (Andrews, Miller & Rayner, 2004), and this is true whether they are semantically

transparent or otherwise (Pollatsek & Hyönä, 2005; Juhasz, 2007). It is important, therefore, to also consider aspects such as compositionality and transparency (traditional metrics in idiom research) and their potential influence on eye movements when deciding on the best approach for the analysis of formulaic units. In this regard, it should also be noted that the discussion so far has focused largely on idioms, but there are many other types of formulaic language. There are good reasons for taking idioms as our prototype: they are unquestionably the most studied of all formulaic types, and they arguably best represent formulaic language as a wider field, given that they "vary along all linguistic dimensions relevant to MWEs [multi-word expressions] generally, including familiarity, literal plausibility, semantic decomposability, and other linguistic attributes" (Titone et al., 2015, p. 173). However, it is equally important to consider how other types of formulaic language might best be analysed, especially items such as collocations (*abject poverty*) and binomials (*king and queen*) which are formulaic only by virtue of frequency and conventionality rather than because they represent a "single meaning" in any way. Again, a hybrid approach might represent the most flexible solution, but careful consideration of the many intra-lexical factors that have been identified in previous studies is equally important.

Conclusions

This short paper has aimed to highlight what we see as a gap in the application of eye-tracking to natural reading behaviour. Our "traditional" measures of eye-tracking relate broadly to single words, and more recently this has been applied to sentence-level syntactic processing and discourse-level understanding/integration, but formulaic sequences have become an important consideration in modern linguistics and must be accommodated in any theoretical approach to language and reading. The key issue is how we might distinguish between the determinants of processing for individual lexical items, such as predictability from context or single word frequency, and a more complex representation of MWUs (which undoubtedly includes predictability but which may well reflect a more nuanced level of cohesion within the mental lexicon). In other words, how do we identify the "added extra" advantage that formulaic sequences seem to have over matched, non-formulaic language, and how do we distinguish this from other language processing mechanisms that might be at play?

It is therefore an open question as to how best we might reconcile these lines of investigation. Eye-tracking has the considerable advantage of presenting the text all at once in a naturalistic way, so it is of great value to the investigation of formulaic language as it can be presented in highly natural contexts. Our methods of interpreting the data, however, must be refined if we want to say more about the nature of this important linguistic phenomenon. Clifton et al. (2007) make a clear distinction between those lexical factors that are best reflected in early measures and the higher level influences that may require a broader set of measurements. Given that formulaic sequences seem to fall to some extent between these two stools, it seems necessary to reconsider our approach to their analysis. A fruitful method might be to borrow the "hybrid" model adopted in the idiom literature and consider formulaic sequences as simultaneously compositional strings and whole units, thereby gaining the maximum benefit of analysis of each word and an overall consideration of the phrase. Crucially, however, formulaic units are neither one thing nor the other: they are not simply combinations of individual words and they are not immutable, unanalysed wholes, so our analysis must bear this in mind and be tailored accordingly.

This discussion has shown that a traditional approach to eye-tracking that takes the single word as its basic unit of analysis is problematic when we consider the range of linguistic units that are inherently multi-word in their construction. The flexibility of eye-tracking and the range of measures available mean that the tools are already in place to tackle this issue, but clearly determining how to apply these measures represents one of the next challenges in the application of this methodology to the study of the "word".

References

Altarriba, J., Kroll, J., Sholl, A., & Rayner, K. (1996). The influence of lexical and conceptual constraints on reading mixed-language sentences: Evidence from eye fixations and naming times. *Memory and Cognition, 24*(4), 477-492.

Andrews, S., Miller, B., & Rayner, K. (2004). Eye movements and morphological segmentation of compound words: There is a mouse in mousetrap. *European Journal of Cognitive Psychology, 16*, 285-311. doi: 10.1080/09541440340000123

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Beninca, P., & Semenza, C. (2012). Is "hit and run" a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology, 3*(11). doi: 10.3389/fpsyg.2012.00011

Ashby, J., Rayner, K., & Clifton, C. (2005). Eye movements of highly skilled and average readers: Differential effects of frequency and predictability. *Quarterly Journal of Experimental Psychology, 58A*, 1065-1086.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27*, 668–683.

Camblin, C. C., Gordon, P. C., & Swaab, T. Y. (2007). The interplay of discourse congruence and lexical association during sentence processing: Evidence from ERPs and eye tracking. *Journal of Memory and Language, 56*(1): 103-128.

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition, 17*(4), 784-797. doi: 10.1017/S1366728913000795

Carrol, G., & Conklin, K. (submitted). Cross language priming extends to formulaic units: Evidence from eye-tracking suggests that this idea "has legs".

Carrol, G., & Conklin, K. (in preparation). Lexical Priming as a mechanism of formulaic language.

Carrol, G., Conklin, K., & Gyllstad, H. (in preparation). Found in translation: The influence of L1 on the processing of idioms in L2.

Ciéslicka, A. B., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain and Language, 116*, 136-150. doi: 10.1016/j.bandl.2010.09.007

Ciéslicka, A. B., Heredia, R. R. & Olivares, M. (2014). It's All in the Eyes: How Language Dominance, Salience, and Context Affect Eye Movements During Idiomatic Language Processing. In M. Pawlak & L. Aronin (Eds.), *Essential Topics in Applied Linguistics and Multilingualism* (pp. 21-41). doi: 10.1007/978-3-319-01414-2_2

Clifton, C., Staub, A., & Rayner, K. (2007). Eye Movements in Reading Words and Sentence. In R. van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye Movements: A Window on Mind and Brain* (pp. 341-372). Oxford: Elsevier.

Conklin, K., & Schmitt, N. (2008). Formulaic Sequences: Are They Processed More Quickly than Nonformulaic Language by Native and Nonnative Speakers? *Applied Linguistics, 29*, 72-89. doi: 10.1093/applin/amm022

Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory and Cognition, 40*(6), 1778-1786. doi: 10.1037/xlm0000013

Duffy, S., Morris, R., & Rayner, K. (1988). Lexical ambiguity and fixation times in reading. *Journal of Memory and Language, 27*, 429-446. doi: 10.1016/0749-596X(88)90066-6

Ehrlich, S., & Rayner, K. (1981). Contextual effects on word perception and eye movements during reading. *Journal of Verbal Learning and Verbal Behavior, 20*, 641-655. doi: 10.1016/S0022-5371(81)90220-6

Engbert, R., Nuthmann, A., Richter, E. M., & Kliegl, R. (2005). SWIFT: a dynamical model of saccade generation during reading. *Psychological Review, 112*(4), 777-813. doi: 10.1037/0033-295X.112.4.777

Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31*(5), 862-877.

Garrod, S., O'Brien, E., Morris, R., & Rayner, K. (1990). Elaborative inferencing as an active or passive process. *Journal of Experimental Psychology: Learning, Memory and Cognition, 16*, 250-257.

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

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London: Routledge.

Holsinger, E., & Kaiser, E. (2010). *Effects of Context on Processing (non)-Compositional Expressions*. Paper presented at the 84th Annual Meeting of the Linguistics Society of America Baltimore, MD.

Inhoff, A. (1984). Two Stages of Word Processing during Eye Fixations in the Reading of Prose. *Journal of Verbal Learning and Verbal Behavior, 23*, 612-624. doi: 10.1016/S0022-5371(84)90382-7

Inhoff, A., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics, 40*, 431-439. doi: 10.3758/BF03208203

Juhasz, B. J. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. Van Gompel, M. Fischer, W. Murray, & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 373-389). Amsterdam: Elsevier.

Juhasz, B. J., Pollatsek, A., Hyönä, J., Drieghe, D., & Rayner, K. (2009). Parafoveal processing within and between words. *The Quarterly Journal of Experimental Psychology, 62*, 1356 –1376.

Juhasz, B. J., Starr, M., Inhoff, A., & Placke, L. (2003). The effects of morphology on the processing of compound words: Evidence from naming, lexical decisions and eye fixations. *British Journal of Psychology, 94*, 223-244.

Kliegl, R., Risse, S., & Laubrock, J. (2007). Preview benefit and parafoveal-on-foveal effects from word n + 2. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 1250–1255.

Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognitive Psychology, 58*(1), 68-101. doi: 10.1016/j.cogpsych.2008.05.002

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

Murray, W., & Rowan, M. (1998). Early, mandatory pragmatic processing. *Journal of Psycholinguistic Research, 27*, 1-22. doi: 10.1023/A:1023233406227

Myers, J., Cook, A., Kambe, G., Mason, R., & O'Brien, E. (2000). Semantic and episodic effects on bridging inferences. *Discourse Processes, 29*, 179-200. doi: 10.1207/S15326950dp2903_1

O'Brien, E., Shank, D., Myers, J., & Rayner, K. (1988). Elaborative inferences during reading: do they occur on-line? *Journal of Experimental Psychology: Learning, Memory and Cognition, 14*, 410-420.

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, 52A*, 717-737.

Pickering, M. J., Frisson, S., McElree, B., & Traxler, M. (2004). Eye movements and semantic composition. In M. Carreriras, & C. Clifton (Eds.), *On-line Study of Sentence Comprehension: Eyetracking, ERPs and Beyond* (pp. 33-50). New York: Psychology Press.

Pollatsek, A., & Hyona, J. (2005). The role of semantic transparency in the processing of Finnish compound words. *Language and Cognitive Processes, 20*, 261-290.

Pollatsek, A., Hyona, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance, 26*, 820-833.

Radach, R., Inhoff, A. W., Glover, L., & Vorstius, C. (2013). Contextual constraint and n + 2 preview effects in reading. *The Quarterly Journal of Experimental Psychology, 66*(3), 619-633. doi: 10.1080/17470218.2012.761256

Rayner, K., & Duffy, S. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition, 14*, 191-201. doi: 10.3758/BF03197692

Rayner, K., Warren, T., Juhasz, B. J., & Liversedge, S. P. (2004). The effect of plausibility on eye movements in reading. *Journal of Experimental Psychology: Learning, Memory and Cognition, 30*(6), 1290-1301.

Rayner, K., & Well, A. (1996). Effects of contextual constraint on eye movements in reading: A further examination. *Psychonomic Bulletin & Review, 3*, 504-509. doi: 10.3758/BF03214555

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

Reichle, E. D., Liversedge, S. P., Pollatsek, A., & Rayner, K. (2009). Encoding multiple words simultaneously in reading is implausible. *Trends in Cognitive Science, 13*(3), 115-119. doi: 10.1016/j.tics.2008.12.002

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: evidence from idiom comprehension. *Journal of Cognitive Neuroscience, 25*(5), 762-776. doi: 10.1162/jocn_a_00337

Sereno, S., O'Donnell, P., & Rayner, K. (2006). Eye movements and lexical ambiguity resolution: Investigating the subordinate bias effect. *Journal of Experimental Psychology: Human Perception and Performance, 32*(2), 335-350.

Sinclair, J. M. (1991). *Corpus, concordance, collocation*. Oxford: Oxford University Press.

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

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: Learning, Memory and Cognition, 37*(3), 776-784. doi: 10.1037/a0022531

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

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In M. G. Gaskell (Ed.), *The Oxford Handbook of Psycholinguistics* (pp. 327-342). Oxford: Oxford University Press.

Sturt, P. (2003). The time course of the application of binding constraints in reference resolution. *Journal of Memory and Language, 48*, 542-562.

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

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

Titone, D., Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting Bilingual and Monolingual Idiom Processing. In R. Heredia, & A. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 171-207). Cambridge: Cambridge University Press.

Titone, D., & Connine, C. (1999). On the compositional and noncompositional nature of idiomatic expressions. *Journal of Pragmatics, 31*, 1655-1674. doi: 10.1016/S0378-2166(99)00008-9

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 Publishing Company.

Van Lancker Sidtis, D. (2012). Two-track mind: Formulaic and novel language support a dual-process model. In M. Faust (Ed.), *The Handbook of the Neuropsychology of Language* (pp. 342–367). Chichester: Wiley-Blackwell.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge: Cambridge University Press.

Wray, A. (2012). What Do We (Think We) Know About Formulaic Language? An Evaluation of the Current State of Play. *Annual Review of Applied Linguistics, 32*, 231-254. doi: 10.1017/S026719051200013X

Wray, A. (2014). Why Are We So Sure We Know What a Word Is? In J. Taylor (Ed.), *The Oxford Handbook of the Word*. doi: 10.1093/oxfordhb/9780199641604.013.032

Wray, A., & Perkins, M. (2000). The functions of formulaic language: an integrated model. *Language and Communication, 20*(1), 1-28. doi: 10.1016/S0271-5309(99)00015-4