

Cognitive Processing of Multiword Expressions in Native and Non-native Speakers of English: Evidence from Gaze Data

Victoria Yaneva, Shiva Taslimipoor, Omid Rohanian, and Le An Ha

Research Institute of Information and Language Processing
University of Wolverhampton, UK
{v.yaneva, shiva.taslimi, omid.rohanian, ha.l.a}@wlv.ac.uk

Abstract. Gaze data has been used to investigate the cognitive processing of certain types of formulaic language such as idioms and binominal phrases, however, very little is known about the online cognitive processing of multiword expressions. In this paper we use gaze features to compare the processing of verb - particle and verb - noun multiword expressions to control phrases of the same part-of-speech pattern. We also compare the gaze data for certain components of these expressions and the control phrases in order to find out whether these components are processed differently from the whole units. We provide results for both native and non-native speakers of English and we analyse the importance of the various gaze features for the purpose of this study. We discuss our findings in light of the E-Z model of reading.

1 Introduction

A large body of research supports the idea that words which often occur together (e.g. “fast food”, “distance learning”, “break the ice”) are stored in the mental lexicon as whole units as opposed to single words [20, 16, 3, 2]. It has been suggested that the long term memory stores these frequently occurring “formulaic” units in order to alleviate the burden that language comprehension poses on the short-term memory, a resource with a comparatively smaller span [3]. Such an approach to facilitating language processing could provide an explanation of the fact that a large proportion of the spoken and written language is formulaic: some corpus studies claim that between 52% and 58% of the language in the analysed corpora follows a formulaic pattern [7], and other studies claim that this figure is around 32% [9]. Furthermore, several eye tracking studies, as discussed in Section 2, provide evidence of processing advantage for idioms over non-idiomatic phrases, where the faster processing of the phrases is used as a proxy to measuring faster lexical access.

While research studies using different approaches support the idea that formulaic language is a specific form of a “shortcut” to language processing, many questions about the stages of processing of such phrases remain unanswered. For example, it is still uncertain whether non-native speakers with a high degree of

proficiency store formulaic sequences as units in their mental lexicon or not. A study by [20] reported no difference in the duration of gaze fixations between idiomatic and control phrases read by non-native speakers. In another study by [16] non-native speakers produced fewer fixations while reading idioms compared to control phrases, but there was no difference in the duration of these fixations. Another gap in the current research is that the few existing eye-tracking studies for formulaic language focused specifically on idioms and very little is known about the processing of other types of phraseological units. The selection of appropriate phrases as experimental stimuli is yet another under researched area. Some of the early eye tracking studies investigating formulaic units (e.g. [20]) did not control for the syntactic structure of the phrases used as stimuli, which may have biased some of their results. Last but not least, all eye-tracking studies investigating formulaic language, that we are aware of, used stimuli that were crafted for the purpose of the research as opposed to naturally occurring examples from corpora.

In this study, we use gaze data from native and non-native speakers of English to investigate whether multi-word expressions (MWEs) have a processing advantage over control phrases. As opposed to designing the target phrases ourselves, we extract these from the GECO corpus¹ [4], a large eye-tracking corpus containing gaze data from native and proficient non-native readers. The use of this data allows the comparison of naturally occurring phrases, as well as the use of more gaze features compared to previous research. We control for the syntactic structure of the investigated phrases by focusing on MWEs and control phrases following a verb and noun (V + N) and verb and particle (V + P) part-of-speech pattern, which do not contain other words in between.

While very informative, the use of eye tracking as an approach for investigating formulaic language is not straightforward. One of the reasons for this is that the very nature of formulaic units challenges the use of the word as a traditional unit of analysis in eye-tracking research [2]. For example, idioms have been described as “superlemmas”, entries that represent the phrase as a whole but are linked to each component word [18]. One way to overcome this problem is the so called *hybrid* approach where both the behaviour of the whole phrase and the behaviour of its component parts are analysed [2]. To account for this, as suggested by the existing literature discussed in Section 2, we compare both the whole units (MWEs versus control phrases), as well as the last word of the units (either noun versus noun or particle versus particle) in order to identify whether the native and non-native readers differ in the stage at which they recognise the unit (the so called “completion of familiarity check”). We discuss this process in light of the E-Z model of reading [13].

The annotation of the MWEs used in this study is available at:

<https://github.com/shivaat/mwe-geco>.

The rest of this paper is organised as follows. Section 2 presents the E-Z model of reading and related work on using gaze data for the investigation of

¹ The GECO corpus can be downloaded freely at:
<http://expsy.ugent.be/downloads/geco>.

formulaic language. Section 3 presents the data used in this study, while Section 4 presents the gaze features. The experiments are described in Section 6 and the main results are discussed in Section 7. Finally, Section 8 summarises the main findings of this study and introduces potential directions for future work.

2 Processing of Formulaic Language and The E-Z Model of Reading

Eye tracking is a process where an eye-tracking device measures the point of gaze of an eye (gaze fixation) or the motion of an eye (saccade) relative to the head and a computer screen [6]. Fixations are eye movements which stabilise the retina over a stationary object of interest, which, in the case of reading research, is the written text and its units (letters, words, phrases, etc). Gaze fixations and revisits (go-back fixations to a previously fixated object) have been widely used as measures of cognitive effort by taking into account their durations and the places in text where longer fixations occur [6].

The most prominent model of reading to date is arguably the E-Z Reader [12, 14, 13]. The model has several versions, which has introduced some changes in the use of terminology (e.g. earlier versions of the model use terms such as “familiarity check” and later versions of the model (e.g. E-Z Reader 7) refer to the same phenomenon as “first stage of lexical access” or “L1”). We have done our best to explain such variability in the use of the terminology.

The central assumptions of the model are that: (1) the signal to move the eyes to the right is given at the stage of word identification; and (2) attention is allocated from one word to the next in a strictly serial fashion [14]. The model thus has five stages: familiarity check, lexical access, early saccadic programming, late saccadic programming and saccadic movement. In later versions of the model the first two stages are referred to as “early visual processing” and “word identification”, which is divided into stages of early and late lexical access. In our review we focus on these two stages as they are most relevant to our research.

During early visual processing, visual features from the printed page are projected from the retina to the visual cortex so that the objects on the page (i.e., the individual words) can be identified [14]. Early visual processing is most rapid if the word is fixated near its center and thus the time needed to encode a word increases with word length. At this stage word-boundary information is obtained, which is needed for the programming of saccades to upcoming words.

The word identification process starts with the focus of attention to that word and is completed in two stages, reflecting early and late stages of lexical processing. The first stage, also referred to as “familiarity check”, corresponds to identifying the orthographic form of the word, where the phonological and semantic forms are not yet fully activated. This activation occurs during the second stage of lexical processing and thus corresponds to what is typically known as “lexical access”. The time required to complete the familiarity check is a linear function of the natural logarithm of the word’s frequency and its predictability

within a given sentence or phrase context. As most of the orthographic recognition has been completed during the first stage, the time required to complete the second stage of lexical processing is more influenced by a word's predictability [14]. Unlike with the first stage of lexical processing, word's predictability fully affects the second stage: "Words that can be predicted with complete certainty within a given sentence context will require no time in this second stage (i.e., if predictability = 1, then $t(L2) = 0$ msec)" [14].

This is particularly relevant to the investigation of the cognitive processing of MWEs and has been discussed in several studies focusing on idioms: "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" [15].

In another study on idiom processing ([20]) the findings are also discussed in light of the E-Z Reader. The reported results showed processing advantage for idioms for native speakers (e.g. "*honesty is the best policy*") compared to novel phrases (e.g. "*it seems that his policy of...*"). For non-native speakers this held true only for number of fixations but not for fixation durations (these two were the only gaze features used in the study). The authors argue that the processing advantage was observed because the familiarity check was completed earlier in the idiom condition due to the high predictability of the terminal word. However, there was no explicit comparison of the terminal words in the two conditions. This study was a seminal one for the field but has been criticised for a number of shortcomings such as using only one measure of processing time (fixation duration), embedding the target words into phrases with varying syntactic structure and varying length and not controlling whether the idioms used in the study were actually known to the non-native readers, with a post-test that strongly suggested half of them were not [15].

Another study by [16], also focusing on idioms, investigated their processing by native and non-native speakers in a biasing story context. The stimuli were three types of phrases, which were either idioms ("*at the end of the day*" - eventually), literal expressions ("*at the end of the day*" - in the evening) or matched novel phrases ("*at the end of the war*"). The authors investigated whether there was processing advantage for the entire unit, as well as its components before and after the recognition point (the point at which the expression becomes uniquely recognisable as idiomatic) [16]. This was done in order to test the *configuration hypothesis* [19], according to which individual words and their meanings are activated until the recognition point has been reached. Once it has been reached, the figurative meaning is accessed and the literal one is no longer viable. The findings of [16] supported the processing advantage of idioms for native speakers, which was found only in the late processing features (total reading time and fixation count). In contrast, the non-native speakers did not read the idioms faster than the novel phrases and required more time to retrieve figurative senses than literal ones, especially in words that were before the recognition point.

To the best of our knowledge, the only eye tracking study on formulaic language different from idioms is a study on binominal expressions (e.g. "*bride and*

groom”) [17], which aimed to find out whether readers were sensitive to the frequency with which phrases occur in language. The results indicated that first pass reading time, total reading time and fixation count were all affected by phrase length, proficiency, phrase type and phrasal frequency in both native and non-native readers.

The three eye tracking studies presented in this section give an important insight into the processing of formulaic language, however, they predominantly focused on idiomatic expressions with varying length (up to eight words) and used a very limited number of gaze features. Furthermore, the control phrases were designed in a way to be as similar as possible to the idiomatic ones but to have literal meaning, which resulted in having control phrases with lower frequency than the idiomatic ones [15]. In an attempt to overcome these shortcomings and to go beyond the state-of-the-art, we present an experiment where: i) we compare MWEs and control phrases extracted from a corpus as opposed to target phrases designed specifically for this study, ii) we control for their length and syntactic structure by focusing on two-word MWEs that are either verb - noun or verb - particle combinations, iii) investigate a larger number of gaze features.

The next sections present the data, gaze features and experimental analysis of our study.

3 Data

This section describes the corpus used in this study, as well as the procedure followed for the annotation of the MWEs.

The GECO corpus. The GECO corpus [4] used in this study is, to the best of our knowledge, the most recent eye tracking corpus for English, which: i) contains gaze data from a natural reading task (as opposed to e.g. single sentences), ii) is long enough to contain a sufficient number of MWEs, and iii) contains paired gaze data from native and non-native readers.

The text of the corpus is a novel by Agatha Christie entitled “The Mysterious Affair at Styles”, the English version of which contains 54,364 tokens and 5,012 unique types. The novel was selected based on the fact that its word frequency distribution was the most similar to the one in natural language use, as observed in the Subtlex database [4]. The novel was read by 14 English monolingual undergraduates from the University of Southampton and 19 Dutch (L1) - English (L2) bilingual students at Ghent University (intermediate and advanced). The two groups were matched on age and education level. The monolingual participants read only the English version of the novel, which amounted to a total of 5,031 sentences. The bilingual participants read chapters 1 - 7 in one language and 8 - 13 in the other in a counterbalanced order, thus reading 2,754 Dutch sentences and 2,449 English sentences. The eight bilingual participants who read the first part of the novel in English read 2,852 English sentences. Eye tracking data was collected for both the English version of the novel and its translation in

Dutch; however, in this study we only focus on the data about English. All participants completed a battery of language proficiency tests, as a result of which two bilinguals were classified as lower intermediate, ten were classified as upper intermediate and seven were scored as advanced L2 language users according to the LexTALE norms [11]. Full details about the method and procedure used for the development of the corpus could be found in [4].

Annotation of the MWEs and Control Phrases. MWEs have been investigated based on their many different characteristics such as fixedness [8], non compositionality [1], and semi-productivity [21]. We have used these properties as the main guidelines for annotating MWEs, specifically following the guidelines provided by the PARSEME project on identifying verbal MWEs.² In those guidelines, MWEs are defined as semantically non-compositional sequences of words with the following compulsory properties: i) their component words include a head word and at least one other syntactically related word; ii) they show some degree of orthographic, morphological, syntactic or semantic idiosyncrasy with respect to what is considered general grammar rules of a language (collocations are excluded); iii) at least two components of such a word sequence have to be lexicalised.

Two annotators with linguistic background labelled the GECO corpus for Verb + Noun and Verb + Particle constructions following the guidelines. We have considered cases where the components of an MWE can occur with at most three words in between. The kappa inter-annotator agreement is $k = 0.7864$. We have resolved the annotation differences by employing a third annotator to decide in cases of disagreement.

We extracted from the corpus all patterns of Verb + Noun and Verb + Prepositions (and Verb + a list of other particles such as *up*, *down*, *over*, *etc*) with no other words between the components. For the V + P pattern this resulted in a total of 1,220 Control phrases and 168 MWEs. Examples of V + P MWEs include “turned out” and “went on”, while “stayed at” and “went in” are examples of control phrases. For the V + N pattern there were 524 Control phrases and 36 MWEs, where “catching sight” and “brought home” are examples of MWEs and “heard footsteps” and “achieved triumphs” are examples of control phrases.

The annotated MWEs and Control phrases are available at:
<https://github.com/shivaat/mwe-geco>.

4 Gaze Features

The GECO corpus contains a number of early and late gaze features, out of which we selected six that have been widely used in general reading research. For comparison, previous studies investigating formulaic language through gaze data have used smaller number of features: two features (*fixation duration* and *fixation count*) in the case of [20] and three features (*first pass reading time*,

² <https://typo.uni-konstanz.de/PARSEME/images/shared-task/guidelines/PARSEME-ST-annotation-guidelines-v6.pdf>

total reading time and *fixation count*) in the case of [16]. All gaze features were averaged over 14 readers for the monolingual data and 19 readers for the bilingual data and are reported in milliseconds or in counts. The features used in this study are as follows.

Word First Fixation Duration is the duration of the first fixation that was on the current word. It is an early measure of word processing and is informative of lexical access and early syntactic processing, as well as oculomotor processes and visual properties of the region [5]. *Word Fixation Count* is the total number of fixations falling on the current word. *First Run Fixation Count* is the number of all fixations in a trial falling in the first run of the word. *Word Go Past Time* is the summation of all fixation durations from when the current word is first fixated until the eyes enter a word with a higher word identification number. *Word Total Reading Time* is the summation of all fixation durations on the current word, including refixations of the region after it was left. *Word Skipping Probability* is the probability that a word may be skipped, i.e. not fixated. A word is considered skipped if no fixation occurred in first-pass reading.

Late measures such as *Word Total Reading Time* or *Word Fixation Count* account for late syntactic processing, textual integration processes, lexical and syntactic/semantic processing and disambiguation in general [5].

The distribution of each gaze feature for the V + N and V + P MWEs and Control phrases for both native (L1) and non-native (L2) speakers are presented in Appendix A.

5 Experimental Design

In the experiments presented below we set out to find out whether there is a processing advantage for MWEs compared to control phrases (Experiment 1) and whether the completion of the familiarity check happens earlier for MWEs than for control phrases (Experiment 2). We also discuss the differences between early and late gaze features for the two groups of phrases in order to draw conclusions about the early and late stages of processing of the phrases.

In Experiment 1 we test whether: i) *native* speakers process whole MWEs differently from whole control phrases and whether ii) *non-native* speakers process whole MWEs differently from whole control phrases. In order to account for the processing of the entire unit we add the values of the gaze features of its component parts together (either V + P or V + N, respectively) for each participant group. We do this separately for each gaze feature. For example, if in the phrase “shook hands” the verb “shook” has a *First Fixation Duration* of 149 msec and the noun “hands” has a *First Fixation Duration* of 189 msec, then the entire phrase has a *First Fixation Duration* of 338 msec. We compare the whole units within groups in order to find out whether there is processing advantage for MWEs over control phrases, as suggested by the related work.

In Experiment 2 we focus only on the last part of the MWEs and the Control phrases, namely the nouns and particles. We compare their processing within groups based on the six gaze features in order to find out whether nouns and

particles which belong to MWEs are processed faster and skipped more often compared to nouns and particles from the control phrases. As suggested by the literature discussed in previous sections and the E-Z model in particular, such faster processing of the last word of the MWE would be an indication of an earlier completion of the familiarity check stage of the phrase processing.

The results of our experiments are presented in the next section.

6 Results

This section presents the statistical analysis results for entire MWEs versus Control phrases, as well as those for the last words of the MWEs and the last words of the Control phrases.

6.1 Experiment 1: MWEs versus Control Phrases

This section presents a comparison between a number of gaze features for V + N and V + P MWEs and Control phrases. The gaze feature values for the whole phrases are computed by adding the values of each individual word in the phrase. The Control phrases also follow the V + N and V + P patterns, but were not classified by the annotators as MWEs based on the annotation guidelines.

Monolingual data A Shapiro-Wilk test revealed that the data for the majority of the gaze features for V + N and V + P phrases was non-normally distributed; hence, we used a Wilcoxon Signed Rank test in order to compare the two types of phrases (MWEs versus Control phrases).

For V + N combinations, the analysis revealed that there were statistically significant differences for all gaze metrics except for First Fixation Duration ($Z = -1.854$, $p = 0.064$) and Skipping Probability ($Z = -1.738$, $p = 0.082$). The results for the rest of the features were as follows (two-tailed): Fixation Count ($Z = -3.166$, $p = 0.002$); First Run Fixation Count ($Z = -2.499$, $p = 0.012$); Go Past Time ($Z = -3.441$, $p = 0.001$), and Total Reading Time ($Z = -2.906$, $p = 0.004$). Table 1 shows the Median values of each gaze feature for both types of phrases.

These results indicate that the processing of V + N MWEs indeed takes less time and less fixations compared to the processing of control V + N phrases.

The case was similar for V + P combinations, where the Wilcoxon Signed Rank test revealed statistically significant differences for all features (two-tailed), except Skipping Probability ($Z = -0.587$, $p = 0.557$). The results for the rest of the features were as follows (two-tailed): Fixation Count ($Z = -3.646$, $p = 0.000$); First Run Fixation Count ($Z = -2.934$, $p = 0.003$); First Fixation Duration ($Z = -2.871$, $p = 0.004$); Go Past Time ($Z = -2.640$, $p = 0.008$), and Total Reading Time ($Z = -3.996$, $p = 0.000$). The Median values are presented in Table 1. Interestingly, the median value for First Fixation Duration is higher for the V + P MWEs (259.89) than the V + P control phrases (258.32), which is the only feature indicating that the MWEs are more challenging to process than the control phrases.

Table 1: Median values for MWEs and Control phrases for **monolingual** participants (added feature values for both words in a phrase) Note that Skipping Probability > 1 because it represents the added skipping probability for each word in the phrase.

	V+N MWEs	V+N Control	V +P MWEs	V+P Control
Fixation Count	1.67	1.93	1.43	1.5
First Run Fixation Count	1.48	1.62	1.29	1.3
First Fixation Duration	296.54	310.57	259.89	258.32
Go Past Time	377.18	482.82	354.03	367.34
Total Reading Time	345.86	402.2	299.39	323.86
Skipping Probability	1	0.79	1.07	1.07

Bilingual data A Shapiro-Wilk test signified a non-normal distribution of the data so, similar to the case with the monolingual participants, a Wilcoxon Signed Rank test was used to compare the V +P and V + N MWEs and control phrases for bilingual participants.

For V + N combinations the test revealed similar results to those with the monolingual participants: MWEs are processed more efficiently than control phrases. The only exceptions were again First Fixation Duration, which did not yield a significant result ($Z = - 0.707$, $p = 0.48$) and Skipping Probability ($Z = - 0.188$, $p = 0.851$). The results for the rest of the features were as follows (two-tailed): Fixation Count ($Z = - 2.821$, $p = 0.005$); First Run Fixation Count ($Z = - 2.636$, $p = 0.008$); Go Past Time ($Z = - 3.158$, $p = 0.002$), and Total Reading Time ($Z = - 4.509$, $p = 0.000$). The Median values are presented in Table 2.

For V + P combinations there was no difference on the First Fixation Duration measure ($Z = - 1.414$, $p = 0.158$) and Skipping Probability ($Z = - 0.023$, $p = 0.981$). It was interesting to note that the Go Past Time measure was significantly lower for the Control phrases ($Z = - 2.645$, $p = 0.008$). The rest of the features differed significantly between MWEs and Control phrases in favour of MWEs being easier to process: Fixation Count ($Z = - 3.314$, $p = 0.001$); First Run Fixation Count ($Z = - 2.881$, $p = 0.004$); Total Reading Time ($Z = - 3.381$, $p = 0.001$). The Median values are presented in Table 2.

Conclusions for MWEs versus Control Phrases The comparison between V + N and V + P MWEs and control phrases using both monolingual and bilingual gaze data reveals that, much in line with previous research on other formulaic sequences (e.g. idioms), the MWEs are generally processed with lower numbers of fixations and lower reading times. Another finding, which relates to previous research is that early gaze measures and First Fixation Duration in particular, were not discriminative between formulaic and non-formulaic phrases. While Skipping Probability is not a discriminative feature, V + P combinations are skipped a lot more often than V + N combinations and the value of the Median for Skipping Probability is higher in cases where the phrase is a MWE.

Table 2: Median values for MWEs and Control phrases for **bilingual** participants (added feature values for both words in a phrase) Note that Skipping Probability > 1 because it represents the added skipping probability for each word in the phrase.

	V+N MWEs	V+N Control	V +P MWEs	V+P Control
Fixation Count	2.28	2.33	1.8	1.89
First Run Fixation Count	1.74	1.89	1.56	1.56
First Fixation Duration	349.67	358.7	309.56	303.11
Go Past Time	497.56	595.47	468.07	451.61
Total Reading Time	484.11	248.28	400.6	404.1
Skipping Probability	0.67	0.67	1	0.9

In terms of differences between monolingual and bilingual participants, we notice that while bilingual participants also process MWEs more efficiently, they have Go Past measures lower for control phrases for the V + P pattern, meaning that they spent longer integrating the meaning of the MWEs with information previously read.

6.2 Experiment 2: Last word of the MWEs versus last word of the Control phrases

In this section we present a comparison between the last words of each unit (nouns from MWEs versus nouns from control phrases and particles from MWEs versus particles from Control phrases). This comparison is motivated by findings from previous research, which state that due to predictability effects, the last words in formulaic sequences are processed through fewer and shorter fixations and are skipped more often than those in non-formulaic ones.

Monolingual data A Wilcoxon Signed Rank test was used to first compare the nouns (N) which were part of MWEs and those which were part of Control phrases, where a processing advantage was revealed for the former.

Again, there were statistically significant results for all features except First Fixation Duration ($Z = -1.147$, $p = 0.251$). Interestingly, in this case Skipping Probability was a discriminative feature, showing that nouns in MWEs were skipped significantly more often than those in Control phrases ($Z = -2.187$, $p = 0.029$). The results for the rest of the features were as follows (two-tailed): Fixation Count ($Z = -2.950$, $p = 0.003$); First Run Fixation Count ($Z = -2.123$, $p = 0.034$); Go Past Time ($Z = -3.441$, $p = 0.001$), and Total Reading Time ($Z = -2.388$, $p = 0.017$). The Median values are presented in Table 3.

Interestingly, for particles (P), the results show that none of the features are discriminative enough between particles which are part of MWEs and those from the Control phrases. A possible explanation for this is that both are skipped often and processed using very few and very short fixations. The results are Fixation Count ($Z = -1.343$, $p = 0.179$); First Run Fixation Count ($Z = -1.255$, $p =$

Table 3: Median values for Ns and Ps for **monolingual** participants

	N MWEs	N Control	P MWEs	P Control
Fixation Count	0.86	0.93	0.57	0.5
First Run Fixation Count	0.78	0.85	0.5	0.46
First Fixation Duration	164.35	161.14	100.61	91
Go Past Time	197.64	246.6	129.17	123.89
Total Reading Time	179.82	198.92	110.36	100.84
Skipping Probability	0.43	0.36	0.64	0.69

0.209); First Fixation Duration ($Z = -1.132$, $p = 0.257$); Go Past Time ($Z = -0.120$, $p = 0.904$), and Total Reading Time ($Z = -1.105$, $p = 0.269$); and Skipping Probability ($Z = -0.231$, $p = 0.818$). The Median values are presented in Table 3.

Bilingual data The comparison between the last words of the MWEs and the Control phrases was repeated using gaze data from bilingual readers in order to identify potential differences in their processing.

Contrary to what we found for monolingual participants, the bilingual ones read the nouns in MWEs and Control phrases with a similar number of fixations and fixation durations. This reveals that the predictability effect was significantly lower for readers to whom English was not a first language. The only statistically different comparison was found for Go Past Time ($Z = -2.781$, $p = 0.005$). The results for the rest of the features were: Fixation Count ($Z = -1.248$, $p = 0.212$); First Run Fixation Count ($Z = -0.314$, $p = 0.754$); First Fixation Duration ($Z = -0.063$, $p = 0.95$); Total Reading Time ($Z = -1.618$, $p = 0.106$); and Skipping Probability ($Z = -0.492$, $p = 0.623$). The Median values are presented in Table 4.

Table 4: Median values for Ns and Ps for **bilingual** participants

	N MWEs	N Control	P MWEs	P Control
Fixation Count	1.3	1.11	0.7	0.67
First Run Fixation Count	0.9	0.9	0.6	0.56
First Fixation Duration	179.73	179.58	122.56	112.11
Go Past Time	246.33	287.44	167.56	157
Total Reading Time	257.67	248.26	146.95	135.6
Skipping Probability	0.3	0.3	0.6	0.6

For Particles the results were similar to those for the monolingual data and to the results for Nouns for the bilingual participants: there were no significant differences found. This suggests that particles which were parts of MWEs were

processed similarly to particles which were from control phrases: Fixation Count ($Z = -1.474$, $p = 0.14$); First Run Fixation Count ($Z = -0.661$, $p = 0.509$); First Fixation Duration ($Z = -0.628$, $p = 0.53$); Go Past Time ($Z = -1.666$, $p = 0.096$); Total Reading Time ($Z = -1.231$, $p = 0.218$); and Skipping Probability ($Z = -1.088$, $p = 0.277$). The Median values are presented in Table 4.

Conclusions for Last word of the MWEs versus last word of the Control phrases The results from the statistical analysis indicated that native speakers show a processing advantage for nouns which are a part of MWEs compared to nouns in Control phrases. This was not the case, however, with non-native speakers, who did not find these nouns more predictable than the nouns in the Control phrases. With regards to particles, both monolingual and bilingual gaze data revealed that these are often skipped and processed equally efficiently when they are a part of MWEs and when they are not.

7 Discussion

The analysis of the data presented in the previous section revealed important findings about the processing of MWEs by native and non-native speakers. We discuss these findings in the context of previous eye-tracking research for idioms and in light of the E-Z model of reading presented in Section 2.

Processing Advantage for MWEs. In line with previous research on idioms (e.g. [16, 15]), MWEs were processed with fewer fixations and shorter fixations compared to Control phrases for native readers. In contrast with idiom research though, where non-native readers did not read idioms more efficiently compared to novel phrases, our findings suggest that in the case of V + N and V + P MWEs such processing advantage exists for non-native readers, too. The effect was significant for both V + N and V + P phrases. The majority of the features (except First Fixation Duration) showed that there was a clear processing advantage for the formulaic phrases as whole units for both groups of readers. As shown by the median values, MWEs of the type V + P had a higher skipping probability than those of the V + N type, especially when they were part of a MWEs. The latter was not unexpected given the effect of word length on parafoveal word recognition (particles are usually shorter than nouns) and the overall processing advantage for MWEs.

Completion of the Familiarity Check. An interesting finding emerged when analysing the last word of the MWEs of the V + N type. As expected, native speakers showed a processing advantage for nouns which were a part of MWEs compared to nouns in Control phrases and skipped them more often, but this was not the case with non-native readers. One possible explanation for this difference is that since native speakers have a higher exposure to the English language, they used the first word of the MWE as a disambiguating region and were better able to predict the second part of the MWEs due to familiarity with

the phrases. The E-Z model of reading explains this as a completion of the familiarity check earlier in the case of MWEs compared to Control phrases. While the non-native readers did show a processing advantage for MWEs over Control phrases, the first word of the MWE did not act as a disambiguating region for them and thus the familiarity check was not completed earlier than with the Control phrases. It is then possible that the processing advantage for MWEs for non-native readers occurred due to higher familiarity with the overall phrase in a way that allows both words to be processed using fewer and shorter fixations. Since no disambiguation and facilitated semantic access occurred after reading the first word, it could be suggested that the MWEs were not yet entrenched as whole units in the mental lexicon of the non-native participants. Further analysis is thus necessary in order to explain the source of the processing advantage for the whole N + P phrases over Control phrases for non-native readers and to further investigate at what stage these phrases begin being stored as “wholes”.

Unsurprisingly, the processing of the last word in the V + P combinations was no different for MWEs and Control phrases for both native and non-native speakers. The most likely explanation for this is that particles are processed parafoveally in both types of phrases and are thus skipped at a very high rate, as shown by the median values in Tables 3 and 4.

Stages of Processing. The discriminative power of the different gaze features used in this study is informative of the stages of processing where the differences between MWEs and Control phrases occur. Early gaze features such as First Fixation Duration were not discriminative between the two types of phrases, which means that the two types of phrases (MWEs and Control) evoked similar orthographic recognition mechanisms. Thus, the differences in the processing of the MWEs occurred at the later stages of lexical and syntactic processing, as evidenced by late features such as Fixation Count, First Run Fixation Count, Go Past Time and Total Reading Time (if the familiarity check was completed earlier and the last word was skipped more often, then the differences in the late gaze features are expected). Even though previous research on formulaic language using gaze data did not utilise as many gaze features and focused on longer phrases, similar findings have been reported with regards to early features: “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” [15]. The present study shows that this is also the case for shorter MWEs consisting of two words.

Finally, analysis of the skipping probability revealed that even though this feature was not discriminative for the whole phrases, the Median values show that V + P combinations are skipped more often than V + N combinations and this is especially the case when the phrase is a MWE. Again, this can be explained with the parafoveal processing of the particles.

Limitations. The main limitation of this study is related to the small number of participants whose gaze data was analysed. We argue that this is a characteris-

tic of most eye-tracking studies and especially the ones which use eye-tracking corpora. By comparison, the Dundee corpus [10], which is the only other eye-tracking corpus for English of a similar size that we are aware of, contains gaze data from only 10 people. Due to the small number of participants, we were not able to conduct between-group comparisons as individual differences in such small samples could not have been properly accounted for. One possible way to solve this in our future work would be to build individual models for each participant.

Another limitation is the fact that the length in characters of the compared examples was not controlled for, however, this was not feasible given that the aim of this study was to compare data from naturally occurring examples from corpora as opposed to carefully designed laboratory stimuli. The division of the examples into V + N and V + P partially accounts for this as we compare phrases with the same part-of-speech patterns.

8 Conclusions and Future Work

We presented an initial study towards investigating the processing of MWEs versus control phrases in native and non-native speakers of English using gaze data from a large corpus. We focused on MWEs following verb + noun and verb + particle part-of-speech patterns and discussed our findings in the context of previous research on idioms and the E-Z model of reading. Our results indicate that both native and non-native speakers show a statistically significant processing advantage for MWEs compared to Control phrases as measured by a number of gaze features. However, the analysis of the last word from each phrase showed that while native readers complete the familiarity check of the phrase after reading the first word, non-native readers do not exhibit a similar pattern and more research is needed in order to investigate the source of the processing advantage for MWEs in non-native readers. Feature analysis revealed that the differences in the processing of the MWEs and the Control phrases occur at the later stages of syntactic, semantic and lexical processing, as opposed to the stages of early lexical processing.

Future work includes analysis of the relationships between the gaze data and several factors which may interact with it such as frequency, length, familiarity, predictability, etc. A suitable approach to investigating this would be mixed-effects modelling. Other avenues for future research include the investigation of other types of MWEs including ones which contain other words in between, as examples of these were fairly frequent within the GECO corpus.

References

1. Baldwin, T., Kim, S.N.: Multiword expressions. In: Handbook of Natural Language Processing, second edition., pp. 267–292. CRC Press (2010)
2. Carrol, G., Conklin, K.: Eye-tracking multi-word units: some methodological questions. *Journal of Eye Movement Research* 7(5) (2015)

3. Conklin, K., Schmitt, N.: The processing of formulaic language. *Annual Review of Applied Linguistics* 32, 45–61 (2012)
4. Cop, U., Dirix, N., Drieghe, D., Duyck, W.: Presenting geco: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior Research Methods* 49(2), 602–615 (2017), <http://dx.doi.org/10.3758/s13428-016-0734-0>
5. Demberg, V., Keller, F.: Data from eye-tracking corpora as evidence for theories of syntactic processing complexity. *Cognition* 109(2), 193–210 (2008)
6. Duchowski, A.: *Eye Tracking Methodology: Theory and Practice*. Springer, second edn. (feb 2009)
7. Erman, B., Warren, B.: The idiom principle and the open choice principle. *Text-Interdisciplinary Journal for the Study of Discourse* 20(1), 29–62 (2000)
8. Fazly, A., Stevenson, S.: A distributional account of the semantics of multiword expressions. *Italian Journal of Linguistics* 1(20), 157–179 (2008)
9. Foster, P.: Rules and routines: A consideration of their role in the task-based language production of native and non-native speakers. *Researching pedagogic tasks: Second language learning, teaching, and testing* pp. 75–93 (2001)
10. Kennedy, A., Pynte, J., Murray, W.S., Paul, S.A.: Frequency and predictability effects in the dundee corpus: An eye movement analysis. *The Quarterly Journal of Experimental Psychology* 66(3), 601–618 (2013)
11. Lemhöfer, K., Broersma, M.: Introducing lextale: A quick and valid lexical test for advanced learners of english. *Behavior Research Methods* 44(2), 325–343 (2012)
12. Rayner, K., Duffy, S.A.: Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition* 14(3), 191–201 (1986)
13. Rayner, K., Pollatsek, A., Ashby, J., Clifton Jr, C.: *Psychology of reading*. Psychology Press (2012)
14. Reichle, E.D., Rayner, K., Pollatsek, A.: The ez reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences* 26(4), 445–476 (2003)
15. Siyanova-Chanturia, A.: Eye-tracking and erps in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon* 8(2), 245–268 (2013)
16. Siyanova-Chanturia, A., Conklin, K., Schmitt, N.: 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 (2011)
17. Siyanova-Chanturia, A., Conklin, K., Van Heuven, W.J.: Seeing a phrase time and again matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37(3), 776 (2011)
18. Sprenger, S.A., Levelt, W.J., Kempen, G.: Lexical access during the production of idiomatic phrases. *Journal of memory and language* 54(2), 161–184 (2006)
19. Tabossi, P., Zardon, F.: The activation of idiomatic meaning in spoken language comprehension. *Idioms: Processing, structure, and interpretation* pp. 145–162 (1993)
20. Underwood, G., Schmitt, N., Galpin, A.: The eyes have it. *Formulaic sequences: Acquisition, processing, and use* 9, 153 (2004)
21. Villavicencio, A.: Verb-particle constructions and lexical resources. In: *Proceedings of the ACL 2003 Workshop on Multiword Expressions: Analysis, Acquisition and Treatment - Volume 18*. pp. 57–64. MWE '03, Association for Computational Linguistics, Stroudsburg, PA, USA (2003)

A Appendix: Distribution of Gaze Features

Tables 5 - 8 present the distribution of each gaze feature for the V + N and V + P MWEs and Control phrases for both native (L1) and non-native (L2) speakers.

Table 5: Eye-tracking data statistics for V+N MWE

Gaze Features	Mean		SD		Variance	
	L1	L2	L1	L2	L1	L2
First Run Fixation Count	1.49	1.77	0.26	0.30	0.07	0.091
Word First Fixation Duration	299.8	351.1	58.14	69.2	3380	4784
Word Go Past Time	433.5	543.4	137.6	206.7	18942	42733
Skipping Probability	0.9	0.67	0.22	0.27	0.05	0.07
Word Fixation Count	1.73	2.3	0.41	0.6	0.17	0.37
Total Reading Time	365.16	501.54	96.81	155.4	9372	24137

Table 6: Eye-tracking data statistics for V+P MWE

Gaze Features	Mean		SD		Variance	
	L1	L2	L1	L2	L1	L2
First Run Fixation Count	1.33	1.57	0.32	0.38	0.1	0.15
Word first Fixation Duration	262.9	314.9	64.3	76.5	4140	5851
Word Go Past Time	393.7	507.9	159.6	228.1	25468	52053
Skipping Probability	1.04	0.92	0.23	0.3	0.05	0.08
Word Fixation Count	1.55	1.93	0.5	0.6	0.22	0.41
Total Reading Time	325.28	426.15	110.9	143.53	12296	20600

Table 7: Eye-tracking data statistics for V+N Control Phrases

Gaze Features	Mean		SD		Variance	
	L1	L2	L1	L2	L1	L2
First Run Fixation Count	1.67	1.94	0.43	0.5	0.18	0.26
Word first Fixation Duration	314.9	358.1	71.81	81.8	5157	6691
Word Go Past Time	520.4	657.6	221.5	328.4	49085	107875
Skipping Probability	0.81	0.67	0.31	0.33	0.09	0.11
Word Fixation Count	2.03	2.45	0.69	0.83	0.47	0.69
Total Reading Time	429.17	259.88	159.73	126	25513	15877

Table 8: Eye-tracking data statistics for V+P Control Phrases

Gaze Features	Mean		SD		Variance	
	L1	L2	L1	L2	L1	L2
First Run Fixation Count	1.33	1.57	0.33	0.39	0.11	0.16
Word first Fixation Duration	258.5	305.8	63.7	77.2	4052	5967
Word Go Past Time	389.3	491	156.9	213.9	24631	45773
Skipping Probability	1.06	0.92	0.26	0.29	0.07	0.09
Word Fixation Count	1.57	1.93	0.48	0.63	0.22	0.4
Total Reading Time	329.88	425.26	107.88	152.77	11637	23338