Abstract

Various studies have consistently shown that collocations are processed faster than matched control phrases, both in L1 and in L2. Most of these studies focused on adjacent collocations (e.g., *provide information*). However, research in corpus linguistics normally uses a span to identify collocations (e.g., ± 4 words), and these non-adjacent collocations (e.g., *provide <u>some of the</u> information*) occur very frequently in language. Nevertheless, how they are processed is less established. A recent study on reading non-adjacent collocations seems to suggest similar processing advantages as for adjacent collocations (Author 2016), but this study was limited to the performance of native speakers. The present study addresses the question of whether advanced non-native speakers also show processing advantages for non-adjacent collocations as native speakers do. Forty advanced non-native speakers of English read collocations in either adjacent or non-adjacent conditions, and their eye-movements were recorded. Mixed-effects analysis of their eye-movements was carried out. The results suggest that non-native speakers read adjacent collocations faster than non-formulaic controls, but this facilitation almost disappears for non-adjacent collocations.

Keywords: collocations, non-adjacent collocations, eye-tracking, mental lexicon

Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations?

Introduction

Formulaic language has been defined as recurring word combinations that are not constructed each time of use but rather have a stereotypical form, conventionalized meaning, and are highly familiar to a speech community (Van Lancker Sidtis 2012). It is usually understood as an opposite to novel language: “nonrecurrent combinations that do not show any significant degree of cohesion or fixedness” (Carrol, Conklin, and Gyllstad 2016: 405). One of the differences between formulaic language and novel language is that formulaic sequences are processed faster and/or more accurately than novel phrases both in receptive use (see Siyanova-Chanturia (2013) for an overview) and in production (e.g. Arnon and Priva 2013, 2014). This processing advantage has been replicated for various types of formulaic sequences: idioms (e.g., Tabossi, Fanari, and Wolf 2009; Vespignani, Canal, Molinaro, Fonda, and Cacciari 2009; Siyanova-Chanturia, Conklin, and Schmitt 2011; Rommers, Dijkstra, and Bastiaansen 2013); lexical bundles – sequences of three or more words recurring frequently in language (e.g., Schmitt, Grandage, and Adolphs 2004; Reali and Christiansen 2007; Nekrasova 2009; Arnon and Snider 2010; Tremblay, Derwing, Libben, and Westbury 2011); phrasal verbs (e.g., Kim and Kim 2012), and collocations (e.g., Durrant and Doherty 2010; Sonbul 2015). This facilitative effect seems to be due to the phrase frequency, familiarity, and predictability leading to some sort of entrenchment of these phrases in speaker’s memory and their faster activation (Siyanova-Chanturia and Martinez 2015).

Usage-based theories of language acquisition can easily account for the processing advantage of formulaic sequences. Kemmer and Barlow (2000) summarized the main features of usage-based approaches as the following: language is seen as a dynamic, ever changing system, where frequency plays a key role, and the linguistic representation is seen as emergent. Frequency leads to an entrenchment: each occurrence of any event leaves a trace in our brain which facilitates later occurrences of that event (Langacker 2000). Recent research seems to show that it is not only the frequency of individual words that matters when processing language, but also the phrasal frequency (e.g., Ellis, Frey, and Jelkanen 2009; Siyanova-Chanturia, Conklin, and van Heuven 2011; Wolter and Gyllstad 2013; Sonbul 2015). Linguistic competence emerges from a large number of utterances that we remember and from regularities that we extract from them (Ellis 2015). Importantly, when we learn a language as infants, we start learning from utterances (Tomasello 2005). Even if later on in our lives we generalize them into certain schemas, we continue to have multiple representations of certain units, including single words, abstract constructions, and specific utterances (see Langacker 2000, Tomasello 2005, also Wray's (2002) Heteromorphic Distributed Lexicon).

Formulaic language is a term used to refer to various types of sequences from frequent transparent word combinations to figurative idioms (see Wray (2002) for a discussion of various sequences classified as formulaic). One type of formulaic sequence is collocations. The definition of collocation varies considerably from any occurrence of words within a certain span (Sinclair 1991) to a sequence of words forming a unit of meaning (Barfield 2012). In this paper we adopt a frequency-based approach to collocations: "collocation typically denotes frequently repeated or statistically significant co-occurrences, whether or not there are any special semantic bonds between collocating items" (Moon 1998: 26), for example, *to take care, to provide information, environmental issues*.

Collocations (like other formulaic sequences) show a processing advantage. This processing advantage seems to be due to their frequency, but also their transitional probabilities, that is, the probability their constituent of words co-occurring together (McDonald and Shillcock 2003a, 2003b). This collocation facilitation effect can be explained by lexical priming theory (Hoey 2005, 2012) which holds that the words we encounter start to prime each other's use. Essentially, this theory is compatible with spreading activation models (Collins and Loftus 1975): when a word is activated this activation spreads to related words (in this case collocates but it could be semantically, phonologically related words, etc.) making their activation easier. Hence, if the association between two collocates is strong enough, one of them activates (or primes) the other. However, it has been noted that collocations are fluid and reflect tendencies of use instead of being fixed word sequences (Wray 2002). For example, the position of the collocates can vary (*provided information* and *some information provided*). Also, other words can intervene between the collocates (*provides* <u>*this kind of*</u> *information)*, or the collocates can be used in different morphological forms (*providing information, information was provided*). The following section will briefly describe studies of how collocational variation affects their processing, although it must be said that these effects are still relatively underexplored.

Processing of variation in collocations

There have been a few studies of the processing of collocation form variation. Bonk and Healy (2005), for example, looked at the positional variation of collocates. They used a primed word naming task with collocates presented to native speakers (NS) of English either forwards (i.e., in canonical order, e.g., *bend* – *rules*) or backwards (e.g., *rules* – *bend*). Both forwards and backwards conditions yielded facilitation effects, suggesting that positional variation does not disrupt processing.

Molinaro, Canal, Vespignani, Pesciarelli, and Cacciari (2013) looked at modified complex collocational prepositions read by NSs of Italian. They carried out a self-paced reading study and an ERP study to compare the core form of a collocational, or compound, preposition (e.g., *in the hands of*) to a modified form with an adjective inserted (e.g., *in the capable hands of*). The self-paced reading experiment showed that prepositions with an insertion were read slower, possibly because of more information to be integrated. However, the ERP results showed that the insertion did not disrupt the processing. The results of the ERP scans suggest that formulaic sequences allow some modification without losing their formulaic status, despite the fact that they have a preferred core structure.

More recently, Author (2016) carried out an eye-tracking study with NSs of English reading adjacent and non-adjacent collocations as compared to novel phrases (e.g., *provide information* and *compare information* versus *provide some of the information* and *compare some of the information*). When looking at the reading times of the final word in the collocation (*information*), she found that there was a facilitative effect for both adjacent and non-adjacent collocations, although the effect was somewhat smaller for the non-adjacent items. However, when looking at the reading time of the entire phrase (which was either a collocation with an insertion or a novel phrase with an insertion) she found a consistent facilitative effect for both adjacent and non-adjacent collocations.

This conclusion seems to be supported by research on statistical learning of non-adjacent dependencies as well. Studies show that humans are able to track and learn non-adjacent dependencies from a very early age (Heugten and Shi 2010). However, it has also been suggested that adjacent dependencies are easier to learn than non-adjacent ones (Gómez 2002; Newport and Aslin 2004). (Note that this finding was later questioned by Vuong, Meyer, and Christiansen (2011, 2016), who claimed that learning of adjacent and non-adjacent dependencies occurs simultaneously).

Taken together these studies of collocation processing seem to suggest that frequent collocations retain their processing advantage even when some variation in form is introduced. However, these studies were all carried out with NSs. It does not necessarily follow that non-native speakers (NNSs) would show the same processing advantages. L2 speakers differ from L1 speakers in many aspects: L2 speakers usually have less exposure to their second language than L1 speakers, they already know formulaic sequences in their L1 and these sequences are not always equivalent to L2 formulaic sequences, and L2 speakers tend to be literate when they start learning their L2, so they pay more attention to individual words (Wray 2002). Despite these factors, there does seem to be a processing advantage in the L2 (at least for adjacent collocations), as we shall see in the following section.

Processing of collocations in L2

Recently, Hernández, Costa, and Arnon (2016) showed that L2 speakers are sensitive to phrasal frequency information and that this sensitivity holds across the frequency continuum, just as for NSs. Thus, it would seem that L2 speakers should not be that different from L1 speakers in the way they process collocations. Indeed, there have been numerous studies of collocation processing in an L2 that have shown processing advantages for collocations over matched novel phrases. Siyanova and Schmitt (2008), for example, were among the first scholars to look at the processing of collocations of different frequency, both in an L1 and L2. Their two studies showed that NNSs develop some sensitivity to native-like collocations and their frequency, but that they are less sensitive to collocation frequency than NSs.

Wolter and Gyllstad (2013) also investigated frequency effects, but with an acceptability judgement task. They found that their NNS participants processed more frequent collocations faster than infrequent ones, as did NSs. The researchers interpreted these findings as supporting usage-based models of language acquisition. Likewise, Sonbul (2015), using an eye-tracking methodology, compared reading times for high-frequency, mid-frequency, and low-frequency collocations, as well as for novel phrases. She showed that NNSs are sensitive to collocation frequency and read more frequent phrases faster.

Taken together, all these results show facilitative frequency effects for NNSs. However, it has to be taken into account that all these studies were carried out with relatively advanced learners (university students), who must have already had considerable exposure to English. It seems, though, that the frequency effect is not as consistent across all proficiency levels. For example, Siyanova-Chanturia, Conklin, and van Heuven (2011), in their binomial eye-tracking study, found that phrase frequency effects held for proficient, but not for less proficient, NNSs.

Thus, proficiency is an important consideration with respect to NNSs. As administering a proficiency test during a study is often not possible for practical reasons, researchers have looked at the effect of proficiency using various approximations, such as self-ratings (Chanturia, Conklin, and van Heuven, 2011) or vocabulary tests (Wolter and Gyllstad 2013; Sonbul 2015; Gyllstad and Wolter 2016). All these studies show a significant effect of language proficiency on processing speed. Hence it is important to control for the language proficiency of the participants.

Overall, it has been found that advanced NNSs show much the same processing advantages as NSs do. However, all of these studies focused on the processing of adjacent collocations only. Ellis (2006) suggested low salience of cues is one of the reasons L2 learners sometimes fail to intake all the relevant information (such as frequency, probability of occurrence.) available to them. In case of non-adjacent collocations, the association between the collocates seems to be less salient because the collocates do not directly follow each other. Hence NNSs might fail to show this facilitative effect for non-adjacent collocations. The present study therefore sets out to answer two main research questions:

1. Do collocation processing benefits for non-native speakers extend to non-adjacent collocations?
2. Does prior vocabulary knowledge moderate the facilitative effect of collocations for non-native speakers?

The aim of the second research question is to (at least partially) take participants' proficiency into account. Vocabulary knowledge seems to correlate with the four language skills: reading, writing, speaking and listening (Alderson, 2005; Milton, 2013). As such it is an indication of overall language proficiency.

Methodology

To answer the research questions, an eye-tracking study was carried out. The study used the same materials Author (2016). So the results of this study of L2 speakers are comparable with the results of L1 speakers in the previous study. The methodology is explained below.

Materials

Four sets of stimuli were created (see Figure 1). This design made it possible to compare adjacent collocations to adjacent control phrases and non-adjacent collocations to non-adjacent control phrases.

Forty verb-noun collocations were selected for the study (see Appendix for the complete list). All the collocations selected had to reach the threshold of 50 lemma occurrences in the BNC (Davies, 2004) and an MI score of at least three. All of these collocations are likely to be transparent for advanced learners. This is to say, the meaning of each one can be inferred if the meaning of their component parts is known.

[FIGURE 1 NEAR HERE]

Verb-noun collocations were chosen because they seem to be used in their non-adjacent forms very frequently. This is probably the case because a noun can be preceded by a determiner or a modifier. Also, verb-noun collocations were studied in Author (2016), which makes these studies directly comparable. While the insertions between the collocates can be of different lengths and different syntactic structures, in the present study the insertions always modified the noun and they were always three words long. A number of factors were considered when selecting the collocations and control items (see Table 1).

[TABLE 1 NEAR HERE]

Four sentences (one for each condition) were written with each of the collocations. The beginning of the sentence before the target phrase was kept identical in all four conditions and, whenever possible, the ending of the sentence was kept identical as well. When changes were necessary, they were minimal, and at least two words after the target noun remained the same in all four conditions. All words for collocations and controls were selected from the list of the first 2,000 most frequent lemmas in the BNC (Kilgarriff n.d.), and the words in the sentences were in the first 3,000 most frequent words list, in order to ensure the comprehension of the words used in the sentences. The BNC was used as a reference corpus as it is a corpus of British English and the L2 speakers who participated in the study lived in the UK at the time of testing.

A norming study was carried out to check for the naturalness of the sentences and the predictability of the final noun in all conditions. In the naturalness norming stage, the participants were asked to rate the sentences for their naturalness (1 being *very unnatural* and 5 being *very natural*). A rater always saw an experimental sentence in only one of the conditions. In the predictability part, the participants were presented with the target sentences up to the noun in the collocation and they were asked to guess the word to follow. They also saw only one of the four conditions for each sentence. The predictability score was calculated as a proportion of participants who guessed the noun correctly. The results of the noun predictability and sentence naturalness studies are presented in Table 2. The results of the norming studies were used in statistical analysis to control for naturalness and predictability.

In order to estimate participants' prior vocabulary knowledge, a part of the Vocabulary Levels Test (VLT) (Schmitt, Schmitt, and Clapham 2001) was administered. Because the participants were all relatively advanced L2 speakers living in the UK, the 3,000, 5,000, and 10,000 sections of the VLT were chosen for the study. The VLT was also used in Sonbul's (2015) study, which has shown a positive effect of vocabulary knowledge on collocation processing.

Participants

Forty participants (24 females and 16 males) took part of the study. They were all students of various disciplines at a British university (undergraduates = 16, postgraduates = 24). Their mean age was 26 (*SD* = 4.96) and they had spent on average 29 months in the UK before the time of the study (*SD* = 31.29). They started learning English when they were on average 8 (*SD* = 3.52) years old. They all had to meet the university's minimum language requirement, which was IELTS 6.0 or TOEFL IBT 79 or PTE Academic 55. They were asked to self-rate their proficiency in English from 1 to 5 and they rated themselves as follows: reading 4.30 (*SD* = 0.64), listening 4.05 (*SD* = 0.77), speaking 3.75 (*SD* = 0.86), and writing 4.08 (*SD* = 0.61). Their scores on the VLT were: 27.90 (*SD* = 2.45) on the 3K level, 25.20

(*SD* = 3.68) on the 5K level, and 13.52 (*SD* = 7.28) on the 10K level. These vocabulary scores correlated with the self-rated proficiency (r = .58, p = .00). The vocabulary scores were used in the analysis as an indicator of their proficiency because it is a more objective measure. The participants came from a variety of L1 backgrounds: Arabic (2), Bengali (1), Cantonese (3), Dutch (1), French (3), German (4), Greek (1), Hungarian (1), Italian (2), Lithuanian (1), Malay (1), Mandarin (5), Polish (4), Portuguese (6), Spanish (4), and Tegulu (1). Each of the participants was tested individually.

Eye-tracking method

Eye tracking is based on recording and analysing one's eye movements. We move our eyes to fixate objects of interest (Wade and Tatler 2011). Eye-tracking research makes an assumption that the time we spend looking at a word is the time we spend processing it. It is widely accepted that in reading, the duration of fixation on the word (the time that the eyes stay relatively still) is controlled by comprehension processes (Hyönä 2011). This conclusion is supported by many studies which show that word fixation time is influenced by a number of various factors including word length (Kliegl et al. 2004), age of acquisition (Juhasz and Rayner 2003), concreteness **(**Juhasz and Rayner 2003)**,** contextual predictability (Rayner 1998; Starr and Rayner 2001, Vainio, Hyönä, and Pajunen's 2009), orthographic neighbourhood size (Pollatsek, Perea, and Binder 1999), phonological neighbourhood density (Yates, Friend, and Ploetz 2008) and transitional probability **(**McDonald and Shillcock 2003a, 2003b). Hence, it seems that the characteristics of a word have an effect on the difficulty of that word and hence the time spent processing it.

There are various different measures that can be adopted for eye-movement analysis, and no single one of them is considered to be the best (Rayner 1998). Roberts and Siyanova-Chanturia (2013) even suggested that looking at different measures is one of the biggest advantages of the eye-tracking technique, as early measures (such as first fixation duration or gaze duration) give an indication of early stages of processing and lexical activation, while late measures (such as total reading time) can show semantic and syntactic processing. Therefore, in this study we chose a number of measures to analyse, focusing both on early measures: first fixation duration, gaze duration, first pass reading time; and on a late measure: total reading time.

Carrol and Conklin (2014) noted that most eye-movement research on reading has focused on individual words and that it is therefore challenging to directly apply the same measures for analysing multiword sequences, especially when choosing the area of interest to investigate. They suggested using a hybrid approach in which researchers analyse both the final word of the target phrase (which could potentially be a locus of facilitation) and the whole target phrase. In such a way, the whole phrase reading captures the effect even if some individual words are skipped and excluded from the individual word analysis. This approach is adopted in the present study: both the final word and the whole phrase reading times are analysed.

Procedure

Participants' eye movements were recorded using an *Eye-Link 1000 Plus* eye-tracker. The experiment began with a 9-point grid calibration. This was followed by five practice sentences. The participants were instructed to read the sentences for comprehension. Each trial started with a fixation point that appeared on the top left corner of the screen to check the calibration. Afterwards, each sentence was presented across one line in the middle of the screen. Experimental sentences were presented across four counterbalanced lists so that the participant always saw a collocation in only one of the four conditions. The experimental items were mixed with 40 filler items and 30 of those items were followed by Yes/No comprehension questions (e.g., *I know I will never forget that amazing summer we spent together in Italy. Did I enjoy my time in Italy?*). The threshold of adequate comprehension in the study was set at answering 80% of the questions correctly but the participants had no difficulty reaching it: on average they answered 93.60% of the questions correctly (*SD* = 6.37). They completed the reading task first, followed by a short language background questionnaire and the VLT.

Data analysis

The eye-tracking data was manually checked for each participant and a drift correction was manually performed when necessary. The recordings were cleaned using the standard procedure of the automatic cleaning function of the *Eyelink DataViewer*, leading to the loss of 1.61% of the data.

All the continuous reading measures were log-transformed. All the frequency measures used in the analysis were log-transformed and all the continuous predictors were centered. The four experimental conditions were inserted into mixed effects models as two two-level categorical variables - collocation status (collocation or not) and adjacency (adjacent phrase or not). These categorical variables were coded specifying custom contrasts of - 0.5 and 0.5. This coding was used to make sure that the models with interactions evaluate the main effects of collocation status and adjacency across all levels of factors.

The data was analysed using Linear Mixed Effects Models, package *lme4* version 1.1-7 (Bates, Maechler, Bolker, and Walker 2014). Linear effects models analysis is an extension of regression analysis. This analysis can account for both by-subject and by-item variation in the same model, instead of separately calculating by-subject and by-item *F* scores (Quené and van den Bergh 2008). Also, it allows to test effects and interactions of both discrete and continuous predictors in the same model (Quené and van den Bergh 2008; Cunnings 2012).

Separate models for each eye-tracking measure were fitted. Model fitting always started from a core model with collocation status, adjacency, and the interaction between those two variables predicting the outcome variable which was a different eye-tracking measure in each model. As the study also aimed to investigate the effect of vocabulary knowledge on collocation processing, VLT scores were also included in this core model in a three way interaction between the collocation status, adjacency, and vocabulary score[1].

Following the suggestions of Barr, Levy, Scheepers, and Tily (2013), the structure of random effects was kept maximal by including random intercepts by-subject and by-item and random slopes for main predictors – collocation status and adjacency – as well as the interaction between the two in the models. The random effect structures are reported for each model.

Starting from this core model, we added all the covariates[2] and applied a backwards step by step model selection. We considered these covariates: trial number, list of presentation, noun length, noun frequency, MI score of the phrases, verb length, verb frequency, phrase frequency, phrase length, insertion length, phrase naturalness, noun predictability, participant's age, gender, and L1. Descriptive statistics of the covariates are summarized in Table 2. Then we one by one removed all the effects that did not reach significance in the models ($t < 2$). We explicitly compared each new model to the previous one (based on AIC scores and Chi square analysis) in order to exclude only the effects that were not significant. The final selected models are reported for each measure. We also plotted the interactions for the ease of interpretation, using the *effects* package (Fox 2009) and analysed them statistically using the *Phia* package (Rosario-Martinez 2015). The *p* values

[1] The core model formula was: Reading time ~ Collocation status * Adjacency* VLT score + (1+ Collocation status * Adjacency | Subject) + (1+Collocation status * Adjacency | Item)

[2] Reading time ~ Collocation status * Adjacency* VLT score + List number + Trial number + Noun frequency + Noun length + Verb frequency + Verb length + Phrase frequency + Phrase length + MI + Noun predictability + Phrase naturalness + Insertion length + L1 + Age + Gender + (1+Collocation+Adjacency | Subject) + (1 + Collocation + Adjacency | Item)

were estimated using the *lmerTest* package (Kuznetsova, Brockhoff, and Bojesen Christensen 2015).

[TABLE 2 NEAR HERE]

Results

Firstly, the answers to the comprehension questions were checked for all the participants. The participants showed no difficulty: on average they answered 93.60% of the questions correctly. The results of the 3,000 level of the VLT also showed that the participants had sufficient knowledge of the 3,000 frequency band (M = 27.90 out of 30, SD = 2.44) and hence should not have had comprehension difficulties with the study materials.

Table 3 provides summary statistics for all eye-tracking measures, both for the final word and for the whole phrase.

[TABLE 3 NEAR HERE]

Final word reading

Table 4 summarizes the selected models for the final word reading time.

[TABLE 4 NEAR HERE]

Table 4 indicates that collocation status was a significant predictor for the reading time of the final word in collocation for all the three eye-tracking measures. However, while for the first fixation duration it came out as a significant main effect with no significant interactions, the other two models would suggest that there was a significant interaction between the collocation status and adjacency. These interactions are illustrated in Figure 2. They suggest that there was a facilitative effect for adjacent collocations, but not for non-adjacent collocations. The statistical analysis of the interactions showed that there was significant difference between collocation and control reading for the adjacent phrases, both for the gaze duration (χ^2 (1) = 12.66, $p < .01$) and for the total reading time (χ^2 (1) = 29.41, $p < .01$) measures. However, when looking at non-adjacent phrases, this difference was not [we need 'not' here, don't we?] significant, neither for gaze duration (χ^2 (1) = 0.27, $p = .61$), nor for total reading time (χ^2 (1) = 0.16, $p = .68$).

As for the vocabulary score, it came out as a significant predictor in both gaze duration and total reading time models, with participants who knew more vocabulary reading faster overall. However it did not significantly interact with the collocation status of the words.

[FIGURE 2 NEAR HERE]

Whole phrase reading

The second area of interest in the study was the whole phrase: a collocation with or without an insertion.

[TABLE 5 NEAR HERE]

As Table 5 shows, for first pass reading time, collocation status almost reached significance as a main effect, but none of the interactions with collocation status were significant. For total reading time, though, there was a significant interaction between collocation status and adjacency, as well as a three-way interaction between the collocation

Formatted: Highlight

status, adjacency, and VLT score. To analyse this interaction further, two separate models (one for the adjacent phrases and one for the non-adjacent ones) were fitted. They are presented in Table 6, which shows that for the adjacent phrases, there was both a significant effect of collocations and an interaction between collocation status and VLT score. This interaction is explored further in Figure 3. The figure shows that for the participants with the lowest VLT scores, the difference between reading collocations and controls was rather small, but with an increase in their vocabulary scores the facilitative effect of collocation status also increased.

However, for the non-adjacent collocations, there was no significant effect of collocation status, although the coefficient approached significance (*p* = .07). VLT score came out as a significant predictor with participants with larger scores reading faster overall, but there was no interaction for collocation status and vocabulary knowledge.

[TABLE 6 NEAR HERE]

[FIGURE 3 NEAR HERE]

Table 7 presents the 95% confidence intervals for coefficients of collocation status and adjacency and the interaction between the two in all the models presented. The relative ‘narrowness’ of the intervals suggest we can have confidence in the above results. [I don’t know what to say either, but maybe something like this? I have added a comment in your Comments document to refer to this.]

[TABLE 7 NEAR HERE]

Commented [LV1]: Honestly, I do not know what to add here…. I have never seen CIs reported for mixed effects models

Formatted: Highlight

Formatted: Highlight

Discussion

Collocation effect

If we look at the adjacent collocations first, they consistently show a facilitative effect. They were read faster than the control phrases in all but one model, and even for first pass reading time, the coefficient for collocation status was virtually significant (p = .051). This result in the present study is in line with the previous findings of processing advantages for collocations for advanced L2 speakers (Siyanova and Schmitt 2008; Wolter and Gyllstad 2011, 2013; Sonbul 2015). The cause of this advantage may be the frequency and mutual expectancy of the items. It seems that when reading adjacent collocations, advanced L2 speakers behave like L1 speakers.

With respect to non-adjacent collocations, the models for gaze duration and total reading time point in the same direction: collocations were read faster when they were adjacent but this effect disappeared when collocations were presented with words intervening in between the collocates. This finding would suggest that for L2 speakers, there is a facilitative effect for adjacent, but not for non-adjacent collocations. As there was no significant interaction for collocation status and adjacency for the first fixation duration and for the first pass reading time, it would seem that when looking at the early measures of collocation processing, collocations were fixated shorter across all conditions. That is, regarding initial lexical access to the word there was some facilitation for all collocations over the control phrases. In the late measures, though, there was no clear facilitation for non-adjacent items anymore. Therefore, the results do not allow to draw the conclusion that there was no effect whatsoever for the non-adjacent collocations: early measures showed an effect for all collocations and in the total reading time of the whole phrase model collocation status also almost reached significance (p = .07) for the non-adjacent collocations. Rather it would seem that there is some trend towards processing non-adjacent collocations faster, but the effect is very small and does not reach significance.

Effect of prior vocabulary knowledge

The second research question addressed the effect of prior vocabulary knowledge on collocation processing. The VLT score came out as a significant predictor in most of the models suggesting that vocabulary knowledge affects language processing in L2. The coefficient for the VLT score as a main effect was always negative, showing that overall reading speed increased with an increase in vocabulary scores. This finding is not surprising, as vocabulary knowledge correlates with overall language proficiency and fluency and with various language skills (Meara 1996; Meara and Jones 1988; Alderson 2005; Milton 2013). The participants who were more proficient were able to access and process words faster.

The VLT score was also a part of a significant three-way interaction for the total reading time of the whole phrase. Analysed further, this interaction showed that for the adjacent collocations, the facilitative effect depended on learner's vocabulary score: for the participants with the lowest vocabulary scores, there was almost no facilitation, but it increased with an increase in the vocabulary scores. This finding could be interpreted in two ways. On the one hand, as already discussed, vocabulary knowledge correlates with language proficiency: more proficient learners are expected to have more collocational knowledge as well. On the other hand, as collocational knowledge can be conceptualized as a part of word knowledge (Nation 2013), an increase in vocabulary knowledge might more directly lead to increasing knowledge of collocational patterns. Nation (2013) conceptualized knowledge of the word as consisting of knowledge of its *form*, *meaning*, and *use*, with knowledge of collocations being included in the realm of *use*. Qian (2002) reported correlations between vocabulary size and depth (as measured by collocational knowledge together with knowledge of synonyms and polysemy). Hence, a general increase in vocabulary knowledge seems to lead to stronger knowledge of collocational patterns in language and to more native-like processing.

Sonbul (2015) also found an effect of proficiency (as measured by a vocabulary score) on collocation processing speed, using the same vocabulary test. In her study, there was an interaction between collocation frequency and the VLT score, albeit only for the off-line rating task. In the present study this interaction manifested itself in the online eye-tracking data as well. Taken together these two studies would suggest processing advantage for collocations is moderated by one's language proficiency.

Other factors

Other factors that came out as significant predictors in various models were noun length and phrase length. This was predictable, considering the suggestions in the literature that a length of the word is a much stronger predictor of reading time than its frequency or predictability (Kliegl, Grabner, Rolfs, and Engbert 2004; Baayen 2008).

Predictability of the noun never came out as a significant predictor in the models. While it was not a significant predictor in most of models for NSs either (Author 2016), it was significant for the first fixation duration. However, we have to consider that predictability norms were established with NSs of English and even for them the predictability of the words was very low.

It seems important to briefly discuss the effect of collocation frequency which was shown to be important for processing collocations in an L2 in other studies (e.g., Wolter and Gyllstad 2013; Sonbul 2015). Phrasal frequency was not a significant predictor in any models in the present study. The reason why phrase frequency was not a significant predictor might be the fact that both phrase frequency and MI score was essentially accounted for by the collocation status (i.e., collocations were more frequent and had higher MI scores than novel phrases). Hence that fact that collocation status was a significant predictor corroborates the findings of the previous studies that frequent collocations are read faster than matched novel phrases.

Finally, a large variation between different individual readers has been reported for NSs (Rayner 1998). It could be even larger with NNSs (especially in this study due to their different L1s, different length of time spent in the UK, etc.). To include this variation in the model, random effects always included by-subject random intercepts and by-subject slopes for collocation status and adjacency. Also, we tried adding L1 as a potential covariate to all of the models. However, it was never a significant predictor of reading speed. As participants spoke many different L1s, the differences between individuals speaking different L1s might have already been captured in by-subject random effects

Comparison with the native speaker data

As the study directly follows the design of Author's (2016) study, we now compare the results of the two. Table 8 summarizes the significance of collocation status and adjacency as well as the interaction between those two variables in all the models for both NSs and NNSs.

[TABLE 8 NEAR HERE]

Looking at whole phrase reading, there was a clear difference between the NSs, who consistently read phrases containing collocations faster, and NNSs, to who the facilitation for non-adjacent collocations did not reach significance in both phrase reading models. The final word total reading time model largely confirmed this, as NNSs only showed facilitation for adjacent collocations. This finding suggests that NSs enjoy a facilitative effect for collocations even when collocations are non-adjacent (especially when looking at phrase reading times), but this effect almost disappears for NNSs.

Commented [LV2]: The reviewer writes there is a "basic mistake" in this sentence. I am afraid I cannot find any. An article before *facilitative*?

The facilitative effect for adjacent collocations for both NSs and NNSs might be due to collocational frequency (frequent patterns become entrenched in memory and can be activated faster (Barlow and Kemmer 2000; Tomasello 2005; Bybee 2006)) and also to the mutual expectancy between the elements of the collocation (in line with the studies of the effect of transitional probabilities for NSs McDonald and Shillcock 2003a, 2003b). Language processing is sensitive to statistical properties, such as frequency information and transitional probabilities (see Ellis 2012, for an overview) and both native and advanced NNSs seem to be capable of cumulating such statistical information about the language.

The difference of processing non-adjacent collocations can have different potential explanations. Firstly, NNSs might be slower and less fluent in integrating inserted modifiers into the context when reading the collocations in non-adjacent conditions. Therefore, they might lose the collocation effect, as they process language more slowly overall. In lexical priming theory (Hoey 2005, 2012) once the first word in a collocation is encountered, it primes the second word and hence facilitation occurs for adjacent collocations. But if the collocation is non-adjacent, this priming effect might have dissipated for the L2 speakers by the time the second element of the collocation is encountered. The same process could be explained by a spreading activation model in language processing (Collins and Loftus 1975). The second element of the collocation is initially activated. However, if there is additional information to integrate (as in nonadjacent collocations), the facilitative effect for the second element of the collocation either remains very small (as in first-pass reading time model) or disappears.

Another possibility (not necessarily mutually exclusive with the first one) is that NNSs might have not accumulated enough exposure to language to strengthen the links between the collocates enough. As all the collocations selected for the study are more frequently used in their non-adjacent forms rather than their adjacent forms (based on BNC frequency information), NSs potentially learn them as non-adjacent dependencies from the very beginning. However, non-adjacent dependencies seem to be more difficult to acquire (Gómez 2002; Newport and Aslin 2004). Therefore, NNSs might initially only attend to adjacent collocations when accumulating frequency information and transitional probabilities. Because of that, NNSs might not cumulate enough information about the links between the verb and the noun. If this is the case, NNSs might process non-adjacent collocations in the same manner as NSs, but in order to achieve a native-like facilitative effect, they would need more exposure to target items than they actually have.

The bottom line is that NSs enjoy the processing benefits of collocations, regardless of whether they are adjacent or not. Advanced NNSs also benefit from processing collocations faster than novel language, but this benefit is almost solely restricted to collocations which are adjacent. The fact that NNSs do not experience facilitation for nonadjacent collocations is important, because most frequency/corpus-based identification criteria for collocations typically include a span, often of ±4 words. That is, a word can be up to four words away from the node word and it would still be identified as a collocate (e.g., *spend* a lot of *time*). It makes good sense for researchers to apply this criterion with respect to NSs because NSs process these separated words in an expedited manner, which suggests some form of 'psycholinguistic reality' even for non-adjacent collocations. However, such corpus extraction techniques may well produce word combinations which NNSs do not process as being linked. This suggests caution in linking corpus output to human processing: recurrent patterns that exist in corpora are not necessarily processed as collocations by NNSs speakers.

Limitations and directions for future research

There were inevitably some limitations of the study and there are numerous ways that the findings could be extended. First of all, the NNSs in the present study were relatively advanced users of English studying in an English-medium university. It is still unclear how sensitive to collocations lower proficiency L2 speakers would be. It might well be that they would not show any processing advantage even for adjacent collocations. It would be interesting to see how much exposure to L2 is needed until a facilitative effect occurs.

Moreover, as the experiment needed to follow a tightly controlled design, the number of inserted words was kept the same in all sentences for the non-adjacent condition (i.e., three). It might well be that the facilitative effect would be significant for non-adjacent collocations if there was only one or two words intervening between the collocates. It remains an empirical question, but the fact that the insertion length came out as a significant predictor in the models would suggest that facilitation effect depends on how far from each other the collocates are or what the syntactic structure of the insertion is.

One other issue that could be taken into consideration in future studies is congruency of collocations with one's L1. Congruency is understood as "presence or absence of [a] literal L1 equivalent" (Peters 2016: 114). It seems that congruent collocations are easier to learn and potentially easier to process for L2 speakers (Gyllstad and Wolter 2016). With a mixed L1 group as in this study, congruency ~~is~~ was impossible to control for. Wolter and Gyllstad (2011) suggest~~ed~~, though, that if a collocation is known, its congruency might not matter that much and the participants can process a known collocation faster no matter what the congruency is. Therefore, in the present study we tried to maximize the chances of the items being known to the participants as a partial solution to the problem of congruency.

We did not collect any familiarity ratings. Familiarity could be an important aspect to control for, as Gyllstad (2016) suggested that familiarity is a significant predictor of phrasal processing speed across all types of phrases: free combinations, collocations, and idioms. However, as the collocations used in this study were not figurative and consisted of high frequency words, we felt that participants would indicate being familiar with the items simply because they could understand them, no matter if they had seen them before or not.

Furthermore, the study analysed verb-noun collocations only. These collocations tend to be used in their non-adjacent form very frequently in language, as they allow a modifier to precede the noun. However, the results cannot be directly extended to all types of lexical collocations without further research. Also, these findings cannot be generalized to non-transparent collocations (or idioms) as semantic transparency plays a role in processing (Gyllstad and Wolter 2016); they have an additional figurative meaning and might be more salient in language flow.

To sum up, the study looked at processing non-adjacent collocations, and showed that NNSs show clear facilitation for adjacent collocations and almost no facilitation for the same items when they are presented in their non-adjacent form. This result would suggest that it is difficult to achieve native-like processing facilitation for non-adjacent collocations, even at advanced levels of proficiency.

References

Author. 2016.

Alderson, J. C. 2005. *Diagnosing Foreign Language Proficiency: The Interface between Learning and Assessment*. London ; New York: Continuum.

Arnon, I. and **U. C. Priva**. 2013. More than words: The effect of multi-word frequency and constituency on phonetic duration. *Language and Speech*, *56*(3), 349–371.

Arnon, I. and **U. C. Priva**. 2014. Time and again: The changing effect of word and multiword frequency on phonetic duration for highly frequent sequences. *The Mental Lexicon*, *9*(3), 377–400.

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

Baayen, R. H. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge, UK ; New York: Cambridge University Press.

Barfield, A. 2012. Lexical collocations. In C. A. Chapelle (Ed.), *The Encyclopedia of Applied Linguistics*. Blackwell Publishing Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0690/abstract

Barlow, M. and **S. Kemmer** (eds.). 2000. *Usage-based models of language*. The Center for the Study of Language and Information Publications.

Barr, D. J., R. Levy, C. Scheepers and **H. J. Tily.** 2013. 'Random effects structure for confirmatory hypothesis testing: Keep it maximal,' *Journal of Memory and Language* 68/3: 255–278.

Bates, D., M. Maechler, B. Bolker, and **S. Walker.** 2014. *lme4: Linear mixed-effects models using Eigen and S4_. R package (Version 1.1-7).* Retrieved from http://CRAN.R-project.org/package=lme4>

Bonk, W. and **A. F. Healy.** 2005. 'Priming effects without semantic or associative links through collocation,' presented at the *46th Annual Meeting of the Psychonomic Society*, Toronto, Canada.

Carrol, G. and **K. Conklin.** 2014. 'Eye-tracking multi-word units: some methodological questions,' *Journal of Eye Movement Research* 7(5)/5: 1-11.

Carrol, G., K. Conklin and **H. Gyllstad**. 2016. Found in translation. *Studies in Second Language Acquisition*, *FirstView*, 1–41.

Collins, A. M. and **E. F. Loftus**. 1975. A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Cunnings, I. 2012. An overview of mixed-effects statistical models for second language researchers. *Second Language Research*, *28*(3), 369–382.

Davies, M. 2004. *BYU-BNC*. (Based on the British National Corpus from Oxford University Press). Retrieved from http://corpus.byu.edu/bnc/

Durrant, P. and **A. Doherty**. 2010. 'Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming,' *Corpus Linguistics and Linguistic Theory* 6/2: 125–155.

Ellis, N. C. 2006. Language acquisition as rational contingency learning. *Applied Linguistics*, *27*(1), 1–24.

Ellis, N. C. 2012. Formulaic Language and Second Language Acquisition: Zipf and the Phrasal Teddy Bear. *Annual Review of Applied Linguistics*, *32*, 17–44.

Ellis, N. C. 2015. Implicit AND explicit language learning: Their dynamic interface and complexity. In P. Rebuschat (Ed.), *Implicit and Explicit Learning of Languages* (pp. 1–24). Amsterdam: John Benjamins Publishing Company.

Ellis, N. C., E. Frey and **I. Jelkanen.** 2009. The psycholinguistic reality of collocation and semantic prosody. In U. Römer and R. Schulze (Eds.), *Exploring the Lexis-grammar Interface* (pp. 89–114). John Benjamins Publishing.

Fox, J. 2009. 'Effect displays in R for Generalised Linear Models,' *Journal of Statistical Software* 8/15: 1–27.

Gómez, R. L. 2002. 'Variability and detection of invariant structure,' *Psychological Science* 13/5: 431–436.

Gyllstad, H. 2016. Comparing L1 and L2 Phraseological Processing: Free Combinations, Collocations and Idioms. Presented at the AAAL, Orlando, USA. Retrieved from https://lucris.lub.lu.se/ws/files/13091847/Poster_AAAL_Henrik_Gyllstad_Final.pdf

Gyllstad, H. and **B. Wolter**. 2016. Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning*, 66(2), 296–3.

Hernández, M., A. Costa and **I. Arnon**. 2016. More than words: multiword frequency effects in non-native speakers. *Language, Cognition and Neuroscience*, *31*(6), 785–800.

Heugten, M. van and **R. Shi.** 2010. 'Infants' sensitivity to non-adjacent dependencies across phonological phrase boundaries,' *The Journal of the Acoustical Society of America* 128/5: 223–228.

Hoey, M. 2005. *Lexical priming: A new theory of words and language*. New York: Routledge.

Hoey, M. 2012. Lexical Priming. In *The Encyclopaedia of Applied Linguistics*. Blackwell Publishing Ltd. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/9781405198431.wbeal0694/abstract

Hyönä, J. 2011. 'Foveal and parafoveal processing during reading' in Liversedge, S. P., I. D. Gilchrist, and S. Everling (eds.), *The Oxford Handbook of Eye Movements* (pp. 819–838). Oxford University Press.

Juhasz, B. J. and **K. Rayner.** 2003. 'Investigating the effects of a set of intercorrelated variables on eye fixation durations in reading,' *Journal of Experimental Psychology. Learning, Memory and Cognition* 29/6: 1312–1318.

Kemmer, S. and **M. Barlow**. 2000. 'Introduction: A Usage-Based Conception of Language' in Barlow, M. and S. Kemmer (eds.), *Usage Based Models of Language* (pp. vii–xxviii). The Center for the Study of Language and Information Publications.

Kilgarriff, A. n.d.. *BNC database and word frequency lists.* Retrieved September 17, 2014, from http://www.kilgarriff.co.uk/bnc-readme.html#lemmatised

Kim, S. H. and **J. H. Kim.** 2012. 'Frequency effects in L2 multiword unit processing: Evidence from self-paced reading,' *TESOL Quarterly* 46/4: 831–841.

Kiss, G. R., C. **Milroy,** and **J. Piper.** 1973. 'An associative thesaurus of English and its computer analysis' in Aitken, A. J., R. W. Bailey, and N. Hamilton-Smith (eds.), *The Computer and Literary Studies*. University Press.

Kliegl, R., E. Grabner, M. Rolfs, and **R. Engbert**. 2004. 'Length, frequency, and predictability effects of words on eye movements in reading,' *European Journal of Cognitive Psychology* 16/1-2: 262–284.

Kuznetsova, A., P. B.Brockhoff, and **B. Bojesen Christensen.** 2015. *lmerTest: Tests in Linear mixed effects models* (Version 2.0-29). Retrieved from http://CRAN.R-project.org/package=lme4>

Langacker, R. W. 2000. A Dynamic Usage-Based Model. In M. Barlow and S. Kemmer (Eds.), *Usage-Based Models of Language* (pp. 1–64). Stanford, California: The Center for the Study of Language and Information Publications.

McDonald, S. A. and **R. C. Shillcock.** 2003a. 'Eye movements reveal the on-line computation of lexical probabilities during reading,' *Psychological Science* 14/6: 648–652.

McDonald, S. A. and **R. C. Shillcock.** 2003b. 'Low-level predictive inference in reading: the influence of transitional probabilities on eye movements,' *Vision Research* 43/16: 1735–1751.

Meara, P. 1996. The dimensions of lexical competence. In *Performance and competence in second language acquisition* (Vol. 35, pp. 33–52). Cambridge University Press

Meara, P. and **G. Jones**. 1988. Vocabulary size as a placement indicator. In P. Grunwell (ed.), *Applied Linguistics in Society* (pp. 80-87). London: CILT.

Milton, J. 2013. 'Measuring the contribution of vocabulary knowledge to proficiency in the four skills' in Bardel C., C. Lindqvist, and B. Laufer (eds.), *L2 vocabulary acquisition, knowledge and use* (pp. 57–78). European Second Languaeg Association.

Molinaro, N., P. Canal, F. Vespignani, F. Pesciarelli, and **C. Cacciari.** 2013. 'Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions,' *Language and Cognitive Processes* 28/6: 1–27.

Moon, R. 1998. *Fixed Expressions and Idioms in English: A Corpus-based Approach*. Oxford: Clarendon Press.

Nation, I. S. P. 2013. *Learning Vocabulary in Another Language* (2nd ed.). Cambridge university press.

Nekrasova, T. M. 2009. 'English L1 and L2 speakers' knowledge of lexical bundles,' *Language Learning* 59/3: 647–686.

Nelson, D. L., C. L.McEvoy, and **T. A**. **Schreiber.** 1998. *The University of South Florida word association, rhyme, and word fragment norms.* Retrieved from http://w3.usf.edu/FreeAssociation/.

Newport, E. L. and **R. N. Aslin.** 2004. 'Learning at a distance I. Statistical learning of non-adjacent dependencies,' *Cognitive Psychology* 48/2: 127–162.

Peters, E. 2016. The learning burden of collocations: The role of interlexical and intralexical factors. *Language Teaching Research*, *20*(1) 113–138.

Pollatsek, A., M. Perea and **K.S. Binder**. 1999. The effects of" neighborhood size" in reading and lexical decision. Journal of Experimental Psychology: Human Perception and Performance, 25(4), 1142.

Qian, D. D. 2002. Investigating the Relationship Between Vocabulary Knowledge and Academic Reading Performance: An Assessment Perspective. *Language Learning*, *52*(3), 513–536.

Quené, H. and **H. van den Bergh**. 2008. Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425.

Rayner, K. 1998. 'Eye movements in reading and information processing: 20 years of research,' *Psychological Bulletin* 124/3: 372–422.

Reali, F. and **M. H**. **Christiansen.** 2007. 'Word chunk frequencies affect the processing of pronominal object-relative clauses,' *The Quarterly Journal of Experimental Psychology* 60/2: 161–170.

Roberts, L. and **A**. **Siyanova-Chanturia.** 2013. 'Using eye-tracking to investigate topics in L2 acquisition and L2 processing,' *Studies in Second Language Acquisition* 35/2: 213–235.

Rommers, J., T. Dijkstra, and **M. Bastiaansen**. 2013. 'Context-dependent semantic processing in the human brain: Evidence from idiom comprehension,' *Journal of Cognitive Neuroscience* 25/5: 762–776.

Rosario-Martinez, H. 2015. *phia: Post-hoc interaction analysis. R package* (Version 0.2-0). Retrieved from http://CRAN.R-project.org/package=phia

Schmitt, N., S. Grandage, and **S**. **Adolphs.** 2004. 'Are corpus-driven recurrent clusters psycholinguistically valid? ' in Schmitt, N. (ed.), *Formulaic Sequences*. John Benjamins Publishing.

Schmitt, N., D. Schmitt, and **C. Clapham.** 2001. 'Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test,' *Language Testing* 18/1: 55–88.

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

Siyanova, A. and **N. Schmitt.** 2008. 'L2 learner production and processing of collocation: A multi-study perspective,' *Canadian Modern Language Review/La Revue Canadienne Des Langues Vivantes* 64/3, 429–458.

Siyanova-Chanturia, A. 2013. '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.

Siyanova-Chanturia, A., K. Conklin, and **N. Schmitt.** 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.

Siyanova-Chanturia, A., K. Conklin, and **W. J. van Heuven**. 2011. '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: 776–784.

Siyanova-Chanturia, A., and **R. Martinez**. 2015. 'The idiom principle revisited,' *Applied Linguistics* 36/5: 549–569.

Sonbul, S. 2015. 'Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing,' *Bilingualism: Language and Cognition* 18/3: 419–437.

Starr, M. S. and **K. Rayner**. 2001. 'Eye movements during reading: some current controversies,' *Trends in Cognitive Sciences* 5/4: 156–163.

Tabossi, P., R. Fanari, and **K. Wolf.** 2009. 'Why are idioms recognized fast? ' *Memory and Cognition* 37/4: 529–540.

Tomasello, M. 2005. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, Mass.: Harvard University Press.

Tremblay, A., B. Derwing, G. Libben, and **C. Westbury**. 2011. 'Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks,' *Language Learning* 61/2: 569–613.

Vainio, S., J. Hyönä, and **A. Pajunen**. 2009. 'Lexical predictability exerts robust effects on fixation duration, but not on initial landing position during reading,' *Experimental Psychology* 56/1: 66–74.

Vespignani, F., P. Canal, N. Molinaro, S. Fonda, and **C. Cacciari**. 2009. 'Predictive mechanisms in idiom comprehension,' *Journal of Cognitive Neuroscience* 22/8, 1682–1700.

Vuong, L., A. S. Meyer and **M. H. Christiansen**. 2011. 'Simultaneous online tracking of adjacent and non-adjacent dependencies in statistical learning' in Carlson C., C. Hölscher, and T. Shipley (eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*. Cognitive Science Society.

Vuong, L. C., A. S. Meyer and **M. H. Christiansen**. 2016. ‘Concurrent Statistical Learning of Adjacent and Nonadjacent Dependencies’. *Language Learning*, 66/1, 8–30.

Wade, N. J. and **B. W. Tatler**. 2011. 'Origins and applications of eye movement research' in Liversedge S. P., I. D. Gilchrist, and S. Everling (eds.), *The Oxford Handbook of Eye Movements* (pp. 17–45). Oxford University Press.

Wolter, B. and **H. Gyllstad**. 2011. 'Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge,' *Applied Linguistics* 32/4: 430–449.

Wolter, B. and **H. Gyllstad.** 2013. 'Frequency of input and L2 collocational processing,' *Studies in Second Language Acquisition* 35/3: 451–482.

Wray, A. 2002. *Formulaic Language and the Lexicon*. New York: Cambridge University Press.

Yamashita, J. and **N. Jiang**. 2010. 'L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations,' *TESOL Quarterly* 44/4: 647–668.

Yates, Mark, John Friend, and **Danielle M. Ploetz**. 2008. ‘The effect of phonological neighborhood density on eye movements during reading,’ *Cognition* 107/2: 685–692.

Table 1

Criteria for selecting target items

	Item	Criteria	Procedure
Word pair frequency	Collocations	Lemma frequency ≥ 50	Frequencies obtained from the BNC using BYU interface (Davies 2004-).
	Controls	Frequency checked to be lower than the collocation frequency for each individual item	
MI score	Collocations	MI ≥ 3	MI estimated from the BNC using BYU corpus interface (Davies 2004-)
	Controls	MI < 1	
Association	Collocations	None of the word pairs listed as associates of each other, either forwards, nor backwards	Two association databases checked for association information (Kiss, Milroy, and Piper 1973; Nelson, McEvoy, and Schreiber 1998).
	Controls		
Verbs	Collocations	All verbs in 2,000 most frequent lemma list 4–9 letters long, controls matched to collocations	Kilgarrif's (n.d.) list used as the list of the most frequent lemmas in the BNC.
	Controls		
Intervening words	Collocations	Always 3 words	BYU interface (Davies 2004) used to check for the MI scores
	Controls	No strong collocates of the noun (MI < 2)	

Table 2

Descriptive statistics of the predictors in the models

Predictor	**Mean**	***SD***	**Range**
Noun length	7.30	1.23	4–12
Verb length (collocations)	6.00	1.33	4–9
Verb length (controls)	6.00	1.23	4–9
Noun frequency	15,864.65	10,574.50	4,398–50,109
Verb frequency (collocations)	18,219.10	11,861.77	5,222–49,571
Verb frequency (controls)	15,864.65	10,574.50	4,398–50,109
MI (collocations)	4.11	0.97	3.02–6.28
MI (controls)	-0.94	1.11	-3.21–0.58
Phrase frequency (collocations)	366.80	486.14	59–2,653
Phrase frequency (controls)	16.20	17.14	0–66
Insertion length	6.94	7.40	0–21
Naturalness [a] (adjacent collocations)	4.46	0.34	3.4–5
Naturalness (adjacent controls)	4.10	0.48	3.2–5
Naturalness (non-adjacent collocations)	4.38	0.46	3.4–5
Naturalness (non-adjacent controls)	4.17	0.47	3.2–5
Predictability[b] (adjacent collocations)	0.04	0.08	0–0.2
Predictability (adjacent controls)	0	0	–
Predictability (non-adjacent collocations)	0.07	0.13	0–0.4
Predictability (non-adjacent controls)	0.01	0.04	0–0.2
Vocabulary test [c]	66.63	11.45	50–88

Note: All lengths are measured in letters. Frequencies refer to a raw number of occurrences in the BNC.

[a] Measured on the Likert scale from 1 to 5, where 5 means very proficient

[b] Measured as a proportion of participants who guessed the noun in the collocations correctly

[c] Cumulative score, max = 90

Table 3

Summary statistics

	Measure (in ms)	Adjacent collocations	Controls	Non-adjacent collocations	Non-adjacent controls
Final word	First fixation duration	245.52 (90.64)	263.77 (97.73)	244.03 (88.14)	251.80 (95.70)
	Gaze duration	303.87 (151.60)	334.57 (162.33)	310.20 (159.93)	309.81 (141.20)
	Total reading Time	415.18 (301.37)	519.93 (382.07)	437.95 (327.73)	440.93 (319.20)
Whole phrase	First pass reading time	544.61 (255.34)	580.26 (299.89)	1037.19 (505.23)	1067.60 (473.69)
	Total reading time	802.76 (506.79)	970.76 (641.86)	1598.42 (909.28)	1667.26 (996.97)

Note: Standard deviations are provided in parentheses.

Table 4

Summary of the selected models for the final word reading measures

	First fixation duration				Gaze duration				Total reading time			
Fixed effects	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
Intercept	2.373	0.010	243.57	.000***	2.452	0.013	191.80	.000***	2.587	0.022	116.18	.000***
Collocation	0.023	0.007	3.15	.002**	0.026	0.009	2.81	.008**	0.047	0.011	4.29	.000***
Adjacency	-0.075	0.021	-3.56	.001***	-0.067	0.026	-2.55	.012*	-0.021	0.014	-1.53	.134
VLT	-0.001	0.001	-0.65	.519	-0.003	0.001	-3.09	.004**	-0.005	0.002	-2.63	.012*
Insertion length	0.005	0.001	3.19	.002**	0.004	0.002	2.17	.033*				
Age	0.005	0.002	2.79	.008**	0.007	0.002	2.70	.010*	0.012	0.004	3.02	.005**
Verb frequency	-0.025	0.012	-2.12	.035*								
Noun length					0.013	0.002	5.35	.000***	0.016	0.004	3.99	.000***
Collocation * Adjacency	-0.022	0.018	-1.23	.225	-0.040	0.017	-2.27	*	-0.080	0.024	-3.36	.001**
Collocation * VLT	0.001	0.001	1.07	.288	0.000	0.001	0.33	.746	0.001	0.001	1.32	.189
Adjacency * VLT	0.000	0.001	0.09	.933	0.000	0.001	0.56	.582	0.000	0.001	0.71	.484
Collocation * Adjacency * VLT	-0.002	0.001	-1.06	.296	-0.000	0.001	-0.08	.935	-0.002	0.002	-0.95	.344
Random effects	**Variance**				**Variance**				**Variance**			
Item	0.000				0.006				0.017			
Collocation | Item	0.000				0.000				0.000			
Adjacency | Item	0.000				0.001				0.001			
Collocation * Adjacency | Item	0.002				0.000				0.003			
Subject	0.003				0.000				0.002			
Collocation | Subject	0.000				0.000				0.000			
Adjacency | Subject	0.000				0.000				0.002			

Collocation * Adjacency | Subject	0.003	0.001	0.002
Residual	0.018	0.027	0.044

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 5

Summary of the selected models for the whole phrase reading measures

	First pass reading time [a]				Total reading time			
	β	***SE***	***t***	***p***	***B***	***SE***	***t***	***p***
Intercept	2.827	0.015	192.21	.000***	3.013	0.023	133.24	.000***
Collocation	0.025	0.012	2.04	.051	0.049	0.010	4.88	.000***
Adjacency	0.059	0.038	1.57	.121	0.059	0.036	1.63	.108
VLT score	-0.004	0.001	-2.89	.006**	-0.005	0.002	-2.61	.013*
Noun length	-0.017	0.004	-4.24	.000***	-0.014	0.004	-3.15	.003**
Phrase length	0.015	0.003	5.81	.000***	0.016	0.002	6.64	.000***
Age	0.008	0.003	2.91	.006**	0.011	0.004	2.53	.016*
Collocation * Adjacency	0.001	0.022	0.03	.974	-0.061	0.021	-2.91	.006**
Collocation * VLT	0.000	0.001	0.15	.886	0.001	0.001	1.46	.145
Adjacency * VLT	0.002	0.001	1.99	.048*	0.001	0.001	0.68	.502
Collocation * Adjacency * VLT	-0.001	0.002	-0.86	.389	-0.004	0.002	-2.28	.026*
Random effects	**Variance**				**Variance**			
Item	0.007				0.018			
Collocation | Item	0.000				0.000			
Adjacency | Item	0.000				0.000			
Collocation * Adjacency | Item					0.003			
Subject	0.000				0.001			
Collocation | Subject	0.001				0.001			
Adjacency | Subject	0.002				0.003			
Collocation * Adjacency | Subject					0.003			
Residual	0.049				0.031			

*** $p < .001$, ** $p < .01$, * $p < .05$

[a] The model contains no interaction in the random effects, because the model with an interaction failed to converge. However, as the interaction of the fixed effects was clearly non-significant, having it in random effects might have overcomplicated the model

Table 6

Models for the adjacent and non-adjacent phrases

	Total reading time							
	Adjacent phrases				**Non-adjacent phrases**			
	β	***SE***	***t***	***p***	***β***	***SE***	***t***	***p***
Intercept	2.872	0.023	124.11	.000***	3.032	0.030	101.02	.000***
Collocation	0.079	0.016	5.09	.000***	0.021	0.011	1.85	.073
VLT	-0.005	0.002	-2.71	.010*	-0.005	0.002	-2.38	.023*
Age	0.011	0.004	2.38	.023*	0.010	0.005	2.31	.027*
Collocation * VLT	0.003	0.001	2.37	.019*	-0.001	0.001	-0.76	.453
Verb length	0.015	0.007	2.09	.043*				
Verb frequency					-0.046	0.023	-2.03	.044*
Insertion length					0.018	0.003	6.68	.000***
Random effects	**Variance**				**Variance**			
Subject (intercept)	0.018				0.019			
Collocation | Subject	0.001				0.000			
Item (intercept)	0.002				0.003			
Collocation | Item	0.001				0.000			
Residual	0.040				0.022			

*** $p < .001$, ** $p < .01$, * $p < .05$

Table 7

95% confidence intervals for the significant effects of collocation status and adjacency

	Final word reading			**Whole phrase reading**	
	First fixation	Gaze duration	Total reading time	First pass	Total reading time
Collocation	0.009, 0.037	0.008, 0.045	0.025, 0.068	0.001, 0.049	0.029, 0.068
Adjacency	-0.117, -0.034	-0.119, -0.016	-	-	-
Collocation* Adjacency	-	-0.074, -0.005	-0.127, -0.034	-	-0.103, -0.020

Formatted Table

Table 8

Comparison between the results of native speakers (Author 2016) and non-native speakers (the present study)

	Final word reading models					
	First fixation		Gaze duration		Total reading time	
Effect	**NS**	**NNS**	**NS**	**NNS**	**NS**	**NNS**
Adjacent collocations	–	+	?	+	+	+
Non-adjacent collocations	–	+	–	–	?	–

	Whole phrase reading models			
	First pass		Total reading time	
Effect	**NS**	**NNS**	**NS**	**NNS**
Adjacent collocations	+	?	+	+
Non-adjacent collocations	+	?	+	?

Note: Plus signs (+) indicate a significant facilitative effect; minus signs (–) show no significant effect; question marks (?) indicate that the effect was significant at $p < .1$ level.

Figure 1

The categories of stimuli and examples

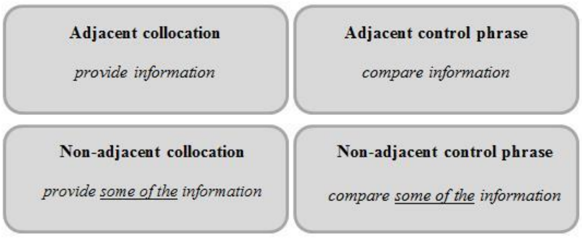

Figure 2

Interaction between collocation status and adjacency in the gaze duration and the final word total reading time models

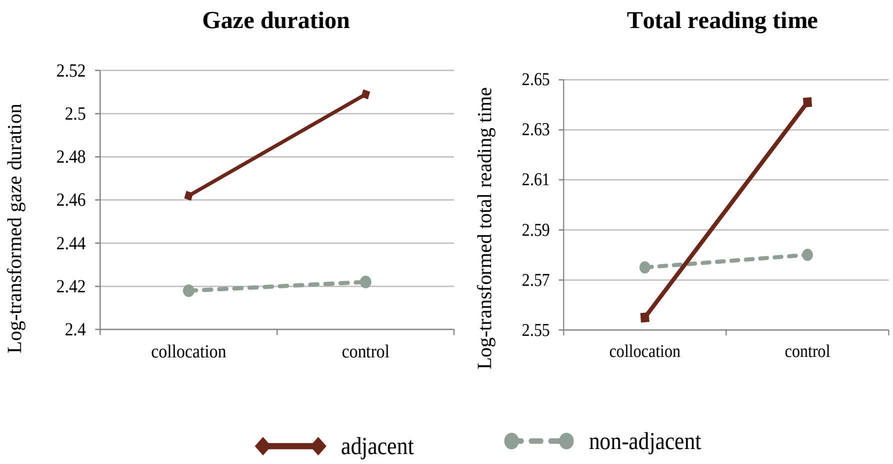

Figure 3

Interaction between VLT score and collocation status in total reading time