

Independent Effects of Collocation Strength and Contextual Predictability on Eye

Movements in Reading

Hui Li^{1,2}

Kayleigh L. Warrington³

Ascensión Pagán⁴

Kevin B. Paterson⁵

Xialou Wang^{1,2,6}

1. School of International Studies, NingboTech University, China.

2. School of International Studies, Zhejiang University, China.

3. Department of Psychology, Nottingham Trent University, UK.

4. Department of Psychology, Northumbria University, UK.

5. Department of Neuroscience, Psychology and Behaviour, University of Leicester, UK.

6. School of Humanities and Communication Arts, Western Sydney University, Penrith, NSW, Australia.

Correspondence should be addressed to Xiaolu Wang, School of International Studies,

NingboTech University, No.1 Qianhu South Road, Ningbo, Zhejiang Province, China. Email:

lucywxl@zju.edu.cn or Kevin Paterson, Department of Neuroscience, Psychology and

Behaviour, University of Leicester, University Road, Leicester, UK, LE1 9HN. Email:

kbp3@le.ac.uk.

Abstract

Collocations are commonly co-occurring word pairs, such as “black coffee”. Previous research has demonstrated a processing advantage for collocations compared to novel phrases, suggesting that readers are sensitive to the frequency that words co-occur in phrases. However, a further question concerns whether this processing advantage for collocations occurs independently from effects of contextual predictability. We examined this issue in an eye movement experiment using adjective-noun pairs that are strong collocations (e.g., “black coffee”) or weak collocations (e.g., “bitter coffee”), based on co-occurrence statistics. These were presented in sentences where the shared concept they expressed (e.g., *coffee*) was predictable or unpredictable from the prior sentence context. We observed clear effects of collocation strength, with shorter reading times for strong compared to weak collocations. Moreover, these effects occurred independently of effects of contextual predictability. The findings therefore provide novel evidence that a processing advantage for collocations is not driven by contextual expectations.

Keywords: Eye Movements in Reading, Collocations, Contextual Predictability

Abstract = 150 words

Main text = 4037 words

People often use formulaic sequences (recurrent strings of words) in written or spoken communication. These include collocations, which are juxtapositions of two or more words, such as “black coffee” or “a quick shower”, that are often used together (Hill, 2000). These sequences are usually considered distinct from compound words (e.g., football, sunflower) or hyphenated compounds (e.g., machine-made), where the conjunction of two or more words is used to create a new or distinctive meaning. On one view, the frequent use of collocations results in these phrases effectively becoming lexicalized so that they are represented in the mental lexicon as a single block of language (e.g., Conkin & Schmitt, 2008, 2012; Siyanova-Chanturia et al., 2011; Underwood et al., 2004; Wray, 2002; Zang, 2019). On another view, the language processor keeps track of statistical information about word co-occurrences. This is thought to provide a means of exploiting redundancy in the linguistic input, so that familiar patterns of co-occurring words can be processed more quickly (Mc Donald & Shillcock, 2003a,b). More generally, the use of such formulaic language is considered to be a hallmark of linguistic proficiency, essential to the development of linguistic competence in L2 readers and speakers (e.g., Wray, 2000). Accordingly, research has investigated whether formulaic language is associated with specific processing advantages.

One approach has been to compare the processing of collocations relative to non-collocations. This has been investigated using an adaptation of the lexical decision task in which participants judge whether a target string is composed of real words or not (Durrant & Doherty, 2010; Ellis et al., 2009; Wolter & Yamashita, 2014). The typical finding is that collocations (e.g., “parish church”) are responded to quicker than novel phrases (“feature church”) and therefore recognized more easily. Other research using measures of eye movements has investigated whether this processing advantage for collocations is observed in reading. Eye movements are sensitive to factors affecting the recognition of words during reading, including the frequency of a word’s written usage and its predictability from the

prior linguistic context (Rayner, 1998, 2009). Reading times typically are shorter for words that have a higher frequency of usage or are more predictable from the context. This research has led to the development of sophisticated computational models of reading (e.g., the E-Z Reader model; Reichle et al., 1998, 2003). Crucially, such models incorporate the assumption that frequency of usage is computed across individual words and not phrases (see Cutter et al., 2014). Therefore, eye movement research showing a processing advantage for collocations (relative to matched non-collocative phrases) may influence the further specification of these eye movement models by demonstrating a need to consider frequency at a phrasal, as well as word, level (see Zang, 2019). Support for this view comes from eye movement studies showing that verb-noun collocations like “provide information” are read faster than matched non-collocations such as “compare information” (Vilkaite, 2016). Similarly, binomial phrases, which are collocations comprising words that appear in a set order (e.g., “bride and groom”), are read faster compared to the same phrases with the word order reversed (e.g., “groom and bride”; Siyanova-Chanturia et al., 2011; and for similar effects for Chinese idioms, see Yu et al., 2016). Such findings are important in suggesting a processing benefit for commonly used phrasal constructions.

Other research has examined whether readers are sensitive to variation in the frequency of usage of collocations. This commonly is computed using measures of phrasal frequency (Gries & Ellis, 2015) or mutual information (MI; Hunston, 2002). Phrasal frequency provides a raw count of how often words are used together in a phrase, while MI provides a conditionalized count (i.e., a ratio) of how often they are used together in a phrase rather than separately. Sonbul (2015) examined eye movements for sentences containing synonymous adjective-noun pairs. These included strong collocations, such as “fatal mistake”, that have both high phrasal frequency and MI, weaker collocations, like “awful mistake”, that have lower phrasal frequency and MI, and phrases like “extreme mistake”, with very low phrasal

frequency and MI. Reading times were shortest for strong collocations, longer for weak collocations, and longest for non-collocations; showing sensitivity to these frequency differences (i.e., differences in phrasal frequency and MI).

Another approach, by McDonald and Shillcock (2003a,b), examined effects of transitional probabilities. This refers to the statistical likelihood that one word follows another in text. However, by comparison with measures of phrasal frequency and MI, the calculation of transitional probabilities does not require that the words are adjacent in the text. McDonald and Shillcock found that reading times were shorter for verb-noun phrases like “accept defeat” that have high transitional probability compared to phrases like “accept losses” that have lower transitional probability. This led McDonald and Shillcock to propose that readers use transitional probabilities to exploit redundancy in the linguistic input to process text more rapidly. However, Frisson et al. (2005) suggested that transitional probabilities might constitute a specific measure of contextual predictability (i.e., the probability of words co-occurring in particular contexts). They tested this hypothesis by comparing eye movements for verb-noun phrases like those used by McDonald and Shillcock in sentences where these phrases were either predictable from the prior context or not. Contextual predictability rather than transitional probabilities influenced reading times, suggesting that transitional probabilities provide a measure of contextual constraint on word co-occurrence rather than a separate statistical measure.

This raises the possibility that collocation effects in other studies might also reflect contextual constraints on word co-occurrence. Accordingly, with the present experiment, we followed a similar approach to Frisson et al. (2005), by examining the processing of collocations in predictable versus neutral contexts. However, by contrast with Frisson et al., we employed adjective-noun pairs rather than verb-noun phrases and assessed collocation strength using both phrasal frequency and MI rather than a measure of transitive probability.

This enabled us to compare strong collocations like “black coffee”, that have high phrasal frequency and MI, with weaker collocations like “bitter coffee” that have lower phrasal frequency and MI. These were placed in sentences where the central concept (e.g., “coffee”) was predictable from the prior context or not. The key consideration was whether an effect of collocation strength would be observed independently of context. If so, we might infer that readers are sensitive to a phrase’s frequency of usage. By contrast, if effects of collocation strength are not observed independently of context, this might provide further evidence that word occurrence statistics provide a specific measure of contextual constraint.

Accordingly, to test these possibilities, we examined whether an interaction effect between collocation strength and contextual predictability was observed in measures of eye movements for the collocative phrase (e.g., “black coffee” versus “bitter coffee”) during sentence reading. We used standard statistical methods to test the null hypothesis that no such effect is observed, and Bayesian methods to assess the relative strength of evidence for models with and without an interaction effect. Our design purposively matched the adjectives in strong and weak collocations (e.g., “black” versus “bitter”) in terms of lexical frequency and letter length (and these words were also closely matched for syllable length). Additionally, to ensure that the observed effects in these analyses were not influenced by uncontrolled differences between these adjectives, we report additional analyses that assessed effects for only the collocation noun (e.g., “coffee”), which was identical across strong and weak collocation pairs (as suggested by Carrol and Conklin, 2014).

Method

Research Ethics. This study was approved by the research ethics committee in the School of Psychology at the University of Leicester and conducted in accordance with the British Psychological Society’s Code of Ethics and Conduct.

Participants. Thirty-two young adults (20 females) aged 18-21 years ($M = 19$ years, SD

= 1.11) from the University of Leicester participated in the experiment. All were native English speakers with no history of dyslexia, and normal or corrected vision, determined using a Bailey-Lovie eye chart (Bailey & Lovie, 1976). To our knowledge, this is the first eye movement study of contextual predictability effects on the processing of strong and weak collocations, which limits the potential for conducting a meaningful *a priori* power analysis to guide sample size decisions. Moreover, the study by Frisson et al. (2005), which is closest in terms of design, reported null effects with respect to the interaction between contextual predictability and the transitive probability of words in a phrase, and so effect sizes from this study would not be helpful for estimating the likely power of our experiment. Accordingly, we used software created by Westfall (<http://jakewestfall.org/>) to estimate the smallest effect size that our design could detect for the interaction. This was in the region of Cohen's $d = .38$ to $.42$, corresponding to a small- to medium-sized effect.

Materials and Design. Stimuli were 48 pairs of adjective-noun collocations from the National British Corpus (Burnage & Dunlop, 1992). Each pair comprised the same noun combined with a different adjective (e.g., “black coffee”, “bitter coffee”). The adjectives in each pair were closely matched for letter length and lexical frequency (see Table 1). We ensured that each pair did not differ in length by more than one letter, that adjective pairs were of similar length across the stimulus set ($t(94) = .73, p = .47$), and did not differ significantly in lexical frequency ($t(94) = .81, p = .42$; using the CELEX database, Baayen et al., 1995). Adjective pairs also did not differ in emotional valence ($t(92) = .103, p = .30$; as determined using norms for stimuli obtainable from Warriner et al., 2013). We also examined the number of syllables in the adjectives; as syllable length, as well as letter length, has been shown to influence eye movements in reading (Ashby & Rayner, 2004). This analysis showed no significant difference in the number of syllables in the adjectives for strong versus weak collocations ($t(94) = .71, p = .48$; see Table 1 for means).

We assessed the association between each adjective and noun combination using two sets of co-occurrence statistics, applying these separately following Sonbul (2015). Phrasal frequencies from the British National Corpus (Burnage & Dunlop, 1992) provided a raw index of how often each combination is used as a phrase. Mutual Information (MI) scores, also obtained from the National British Corpus, provided a conditionalized measure (i.e., a ratio) of the frequency that words are used together relative to used separately (Hunston, 2002). All phrases had an MI above 3. This indicates that the word-pair is three times more likely to occur together in a phrase as separately in text. This value ($MI = 3$) is a conventional cut-off for when a phrase should be regarded as a collocation (see Hunston, 2002). Accordingly, all the phrasal stimuli in the present experiment were collocations. However, stimuli were purposively selected so that one phrase in each stimulus pair had both a higher phrasal frequency and a higher MI than the other (following Sonbul, 2015). Based on these scores, we categorized the higher-scoring phrase as a strong collocation and the other as a weak collocation. An independent-samples t -test confirmed that, across the stimulus set, strong and weak collocations differed significantly in both phrasal frequency ($t(94) = 4.69, p < .001$) and MI ($t(94) = 13.88, p < .001$).

Table 1

To examine effects of contextual predictability, we created 48 pairs of sentence frames, constructed using a range of syntactic structures, into which the strong and weak collocations could be inserted interchangeably (see Figure 1). Sentences were 9 to 20 words long ($M = 14.7, SD = 2.48$), including the collocation, which always appeared near the middle of the sentence, and sentences were presented as a single line of text. Sentences were selected to provide a neutral context or one that strongly predicted the target concept (e.g., *coffee*).

Figure 1

A modified cloze procedure was used to assess predictability. Twenty participants

provided written continuations for sentences truncated immediately before the collocation.

We considered a collocation to be predictable if a continuation included it or words related to its noun. For instance, if the expected collocation was “black coffee” or “bitter coffee”, both these specific phrases or continuations related to the concept of “coffee” (e.g., “cup of coffee”, “espresso”) were taken to demonstrate predictability. Continuations for the selected items contained the target phrase or a related phrase more often in predictable than neutral contexts (77% vs. 2%, $t(94) = 33.41$, $p < .001$).

Another 20 participants assessed the sentences for naturalness (using a 5-point Likert scale where 1 = very unnatural and 5 = very natural; see Table 1). A two-way ANOVA confirmed no difference in naturalness ratings across neutral and predictable contexts, $F(1, 19) = 1.949$, $p = .179$, $\eta^2 = .093$, or between sentences containing strong and weak collocations, $F(1, 19) = .300$, $p = .590$, $\eta^2 = .016$, with no interaction, $F(1, 19) = .511$, $p = .483$, $\eta^2 = .026$. Strong and weak collocations therefore appeared equally acceptable (and so not anomalous) in neutral and predictable contexts.

Stimuli were divided into two lists. Each included half the predictable sentence frames and half the neutral sentence frames. One member of each collocation pair appeared in a neutral frame and the other in a predictable frame for one list, with the opposite allocation of collocations to frames for the other list. This ensured each participant viewed a collocation only once but an equal number of strong and weak collocations in neutral and predictable frames. Strong and weak collocations were viewed equally often in these frames across the experiment. Stimuli were intermixed with 50 filler sentences in each list, which began with 8 practice sentences. Each participant read 154 sentences.

Apparatus and Procedure. An EyeLink 1000 eye-tracker (SR Research inc.) recorded right-eye gaze location every millisecond during binocular reading. Sentences were displayed in 20-point Courier New font as black-on-grey text on a 24-inch high-resolution (1920×

1080) Benq TRT monitor with a 144 Hz refresh rate. At 80 cm viewing distance, each letter subtended approximately 0.3° and so was of normal size for reading (Rayner & Pollatsek, 1989).

Participants took part individually and were instructed to read normally and for comprehension. A chin and forehead rest was used to minimize head movements. A three-point horizontal calibration procedure was used to calibrate the eye-tracker to the participant's eye movements (ensuring $< 0.35^\circ$ spatial error). Calibration accuracy was checked prior to each trial and the eye-tracker recalibrated as necessary to maintain this high spatial accuracy. At the start of each trial, a fixation cross appeared on the left side of the screen. Once the participant fixated this location for 200 ms, a sentence was presented with its first letter replacing the cross. On finishing reading a sentence, the participant pressed a response button and the sentence disappeared, replaced by a yes/no comprehension question on 25% of trials. This was answered by pressing one of two buttons. The experiment lasted approximately 40 minutes for each participant.

Results

Accuracy answering comprehension questions averaged 95% ($> 80\%$ for all participants), and did not differ across conditions ($ps > .1$). Participants therefore had no difficulty comprehending the sentences. Prior to data analysis, short fixations (< 40 ms) were combined with nearby fixations, after which fixations under 80 ms and over 1000 ms were deleted (affecting 5.4% of fixations), following standard procedures. Fixations more than 2.5 SD from the mean per condition for each participant were also removed as outliers (affecting 3% of data).

The remaining data were analyzed using R (R Core Team, 2019) and the `glmer` function, gamma family and identity link in the `lme4` package (Bates et al., 2012), following Lo & Andrews (2015). For all analyses, participants and stimuli were specified as crossed

random effects, with collocation strength and contextual predictability specified as fixed factors. Contrasts comparing levels of the fixed factors were implemented using the “contr.sdif” function in the MASS package (Venables & Ripley, 2002). A full random effects model was used where possible (Barr et al., 2013). If this failed to converge, we increased iterations using the “bobyqa” optimizer (Powell, 2009), before trimming the random structure until it converged, first for random effects for stimuli, then participants. For all analyses, t/z values greater than 1.96 were considered statistically significant (see, e.g., Baayen, 2008).

Eye movement measures are reported for the specific regions of text comprising the collocation or only its noun (see Carrol & Conklin, 2014). This helped ensure that the observed effects were not influenced by uncontrolled differences between adjectives in the collocation, by assessing if the same pattern of effects was observed for a region of text (i.e., the collocation noun) that was identical across strong and weak collocations pairs.

We included eye movement measures sensitive to first-pass processing (processing within a region prior to a saccade to its right or a regression to its left) as well as measures of later processing. Measures for the collocation comprised: first-pass reading time (FPRT, sum of all first-pass fixations in a region), regression-path duration (RPD, sum of all fixations from the first fixation in a region until a fixation to its right, so including fixations following a regression; Liversedge et al., 1998); total reading time (TRT, sum of all fixations within a region) and regressions in (RI, probability of a regression back to a region). Additional measures for the noun comprised: word-skipping (SKIP, probability of not fixating a word during first-pass); first-fixation duration (FFD, length of the first first-pass fixation on a word); single-fixation duration (SFD, length of the first-pass fixation for words receiving only one first-pass fixation); and gaze duration (GD; sum of all first-pass fixations on a word). Note that collocations were skipping infrequently, so this is not reported.

Collocation Effects. Mean eye movements for the collocation are shown in Table 2 and

statistical effects reported in Table 3. All measures, except for regressions-out (i.e., regressions from the collocation), showed a significant effect of collocation strength, with shorter reading times and fewer regressions-in (i.e., regressions back to the collocation) for strong compared to weak collocations. In addition, all measures showed an effect of contextual predictability, with shorter reading times, and both fewer regressions-out and fewer regressions-in (i.e., both from the collocation and also back to the collocation), for collocations in predictable compared to neutral contexts. No significant interactions were observed in eye movement measures (all $t/z < 1.30$).

Tables 2-5

Collocation Noun Effects. Mean eye movement measures for the noun are shown in Table 4 and the corresponding statistical effects reported in Table 5. All reading time measures showed an effect of collocation strength, with shorter reading times for nouns in strong than weak collocations. In addition, a main effect of contextual predictability was observed in all measures. This was due primarily to increased word-skipping, shorter reading times and fewer regressions-in (i.e., regressions back to the noun) in predictable compared to neutral contexts. However, we also observed a small increase in regressions-out (i.e., regressions from the noun) in predictable compared to neutral contexts. This appears to reflect a higher probability of a regressive eye movement to check the contextual fit of the collocative noun when the prior context was more constraining. Crucially, no significant interactions were observed in eye movement measures (all $t/z < 1.90$).

Bayes Factor Analyses. The lack of a significant interaction effect in the above analyses cannot be interpreted as the absence of an interaction. Accordingly, we used Bayes factors (Kass & Raftery, 1995) to assess the strength of evidence for models including an interaction effect against alternative models without an interaction effect. These were performed using the `lmBF` function from the `BayesFactor` package (version 0.9.12-2; Rouder et al., 2012) in R.

Bayes factors for the glmer models reported here are not currently implemented within this package, so models were first refit using the lmer function from the lme4 package (Bates et al., 2012). This produced the same pattern of statistical results as the glmer models. Analyses were restricted to continuous eye movement measures. Marginal likelihood was obtained using Monte Carlo sampling, with iterations set at 100,000, and the scaling factor for g-priors set to 0.5. Participants and stimuli were specified as random variables. Model comparisons (models with versus models without an interaction effect) were made using standard interpretation categories (Vandekerckhove et al., 2014; derived from Jeffreys, 1961). Bayes factors (BFs) > 3 were taken to provide weak to moderate support for models with an interaction effect, and BFs > 10 to provide strong support for such models, whereas BFs < 1 provided evidence in favor of a model without an interaction effect. In all measures, the results provided support for models without an interaction effect (BFs < 0.22). Thus, these additional analyses provide compelling positive evidence that effects of collocation strength were independent of context.

Discussion

The present findings provide valuable evidence that eye movements are sensitive to the frequency of usage of collocations. In particular, we observed shorter reading times for frequently used “strong” collocations compared to less frequently used “weak” collocations. These effects emerged early in the eye movement record, in measures of first-pass processing, indicating that collocation frequency influenced an early stage of phrasal processing. This is consistent with previous demonstrations of a processing advantage for more frequently used collocative phrases (Sonbul, 2015; Vilkaite, 2016).

We also observed clear effects of contextual predictability, in line with previous research (see Rayner, 2009). As with the collocation effect, this effect of contextual predictability emerged early in the eye movement record, in first-pass reading times for the

collocation phrase, and both word-skipping rates and early measures of fixational processing (i.e., first-fixation durations) for the collocative noun. The timing of this effect is important, as it indicates that contextual influences on processing were experienced at broadly the same timing as the collocation effect. Crucially, however, there was no interaction between contextual predictability and collocation strength (with Bayes Factors strongly favoring models with no interaction effects over models with interaction effects). Our findings therefore suggest that collocation strength was processed independently of the contextual predictability of that phrase. This contrasts with previous research showing that a processing advantage for frequently used verb-phrases (as defined using transitional probabilities, i.e., the statistical co-occurrence of words) could be explained in terms of contextual predictability (Frisson et al., 2005). These previous findings led to the proposal that apparent processing benefits for frequently co-occurring words might reflect a form of contextual constraint rather than a separate statistical measure.

The present findings show this is not the case for collocations, as defined using a combination of phrasal frequency and MI scores. In particular, our findings showed that effects of collocation strength, defined in terms of the frequency of usage of words as a phrase, are observed independently of effects contextual predictability. One possibility, as outlined in the Introduction, is that words that are used together frequently in phrase might effectively become lexicalized so that they are represented as a single unit of language in the mental lexicon (e.g., Conkin & Schmitt, 2008, 2012; Siyanova-Chanturia et al., 2011; Underwood et al., 2004; Wray, 2002; Zang, 2019). The findings from the present experiment are not directly informative about whether collocations or other types of formulaic language become lexicalized. However, we consider that the present findings contribute to the debate concerning this issue by demonstrating that the frequency of usage of such phrases can influence eye movements during reading, and that this effect cannot be simply explained in

terms of a specific form of contextual constraint.

Such findings are highly relevant to the future development of computational models of eye movement control in reading. As we noted in the Introduction, a core assumption of current models (e.g., the E-Z Reader model; Reichle et al., 1998, 2003) is that lexical frequency is computed only across words and not phrases. Our findings sit alongside evidence from other studies showing that eye movements in reading are sensitive to the frequency of usage of various multi-constituent linguistic units, including idioms, spaced compounds, and collocations (e.g., Conkin & Schmitt, 2008, 2012; Siyanova-Chanturia et al., 2011; Wray, 2002; Zang, 2019). These findings imply that current models of eye movement control may need to be modified to include mechanisms that are sensitive to both the frequency of usage of multi-constituent units, as well as individual words, if they are to fully account for effects of lexical frequency in reading.

Acknowledgement

The research was funded by a Major Project of National Social Science Foundation grant (14ZDB155) and a Humanities and Social Science Foundation grant from the Education Ministry of the People's Republic of China (No. 19YJC740027).

Hui Li is first author, and Xiaolu Wang and Kevin Paterson are joint corresponding authors. All authors contributed to the experimental design. Hui Li, Kayleigh Warrington, Ascension Pagan and Kevin Paterson designed the materials, Hui Li collected the data, Hui Li, Kayleigh Warrington and Ascension Pagan analysed the data, Hui Li and Kevin Paterson wrote the manuscript. Kayleigh Warrington, Ascension Pagan and Xiaolu Wang gave critical comments.

Stimuli, data files and R scripts used for analyses are available via the University of Leicester Figshare site: <https://figshare.com/s/6a977198684e9a10fe76>

References

- Ashby, J., & Rayner, K. (2004). Representing syllable information during silent reading: Evidence from eye movements, *Language and Cognitive Processes*, 19, 391-426, DOI: 10.1080/01690960344000233
- Baayen, R.H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press; Cambridge, UK.
- Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). The CELEX Lexical Database (Release 2) [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania.
- Bailey, I.L, Lovie, J.E. (1976). New design principles for visual acuity letter charts. *American Journal of Optometry and Physiological Optics*, 53, 740-745.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68, 255-278.
- Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42.
- Burnage, G., & Dunlop, D. (1992). Encoding the British National Corpus. In J. M. Aarts, P. de Haan, & N. Oostdijk (Eds.), *English language corpora: Design, analysis, exploitation* (pp. 79-95). Amsterdam: Rodopi.
- Carrol, G., & Conklin, K. (2014). Eye-tracking multi-word units: Some methodological questions. *Journal of Eye Movement Research*, 7, 1-11.
- Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, 29, 72-89.
- Conklin, K., & Schmitt, N. (2012). The processing of formulaic language. *Annual Review of*

Applied Linguistics, 32, 45-61.

- Cutter, M. G., Drieghe, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 40, 1778-1786. doi: 10.1037/xlm0000013
- Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, 6, 125-155. <http://doi.org/10.1515/cllt.2010.006>
- Ellis, N., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody. In U. Romer & R. Schulze (Eds.), *Exploring the lexis-grammar interface*. Amsterdam: John Benjamins.
- Frisson, S., Rayner, K., & Pickering, M. J. (2005). Effects of contextual predictability and transitional probability on eye movements during reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 862-877. <http://dx.doi.org/10.1037/0278-7393.31.5.862>
- Gries, S.T., & Ellis, N.C. (2015). Statistical measures for usage-based linguistics. *Language Learning*, 65, 228-255.
- Hill, J. (2000). *Revising priorities: From grammatical failure to collocational success*. In M. Lewis (Ed.), *Teaching collocation* (pp. 47–69). Hove, England: Language Teaching Publications.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press. DOI: <https://doi.org/10.1017/CBO9781139524773>
- Jeffreys, H. (1961). *Theory of probability (3rd ed.)*. Oxford, UK: Oxford University Press.
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, 90, 773-795. <https://doi.org/10.1080/01621459.1995.10476572>
- Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998) Eye movements and measures

- of reading time. In: G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55-75). BV, Netherlands: Elsevier.
- Lo, S., & Andrew, S. (2015). To transform or not to transform: using generalized linear mixed models to analyze reaction time data. *Frontiers in Psychology*, 6, 1171.
- McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science*, 14, 648-652.
- McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research*, 43, 1735-1751.
- M. J. D. Powell (2009), "The BOBYQA algorithm for bound constrained optimization without derivatives", Report No. DAMTP 2009/NA06, Centre for Mathematical Sciences, University of Cambridge, UK. http://www.damtp.cam.ac.uk/user/na/NA_papers/NA2009_06.pdf.
- R Core Team (2019). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. <<http://www.R-project.org>>.
- Rayner, K. (1998). Eye Movements in Reading and Information Processing 20 Years of Research. *Psychological Bulletin*, 124, 372-422.
- Rayner K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, 62, 1457-1506. doi: <https://doi.org/10.1080/17470210902816461>.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ; Prentice Hall.
- Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review*, 105, 125-157.
- Reichle, E. D., Rayner, K., Pollatsek, A. (2003). The E-Z Reader model of eye-movement

- control in reading: Comparisons to other models. *Behavioral and Brain Sciences*, 26, 445-476.
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, 56, 356-374.
<https://doi.org/10.1016/j.jmp.2012.08.001>
- Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011). Seeing a phrase “time and again” matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 776-784.
- Sonbul, S. (2015). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition*, 18, 419-437. <http://dx.doi.org/10.1017/S1366728914000674>
- Underwood, G., N. Schmitt & A. Galpin. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (ed.), *Formulaic sequences* (pp.153-172). Amsterdam, the Netherlands: John Benjamins.
- Vandekerckhove, J., Matzke, D., & Wagenmakers, E.-J. (2014). Model comparison and the principle of parsimony. In J. R. Busemeyer, Z. Townsend, J. Wang, & A. Eidels (Eds.), *Oxford handbook of computational and mathematical psychology* (pp. 125-144). Oxford, UK: Oxford University Press.
- Venables, W. N., & Ripley, B. D. (2002). *Modern applied statistics with S (4th ed.)*. New York, USA: Springer.
- Vilkaite, L. (2016). Are non-adjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition* 42, 1632-1642.
- Yu, L., Cutter, M.G., Yan, G., Bai, X., Fu, Y., Drieghe, D., & Liversedge, S.P. (2016). Word n+2 preview effects in three-character Chinese idioms and phrases. *Language*,

Cognition and Neuroscience, 31, 1130-1149.

Warriner, A. B., Kuperman, V. & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods*, 45, 1191-1207.

<https://doi.org/10.3758/s13428-012-0314-x>

Wolter, B., & Yamashita, J. (2014). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, 36, 1193-1221.

Wray, A. (2000). Formulaic sequences in second language teaching: Principle and practice, *Applied Linguistics*, 21, 463-489.

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

Zang, C. (2019). New perspectives on serialism and parallelism in oculomotor control during reading: The multi-constituent unit hypothesis. *Vision*, 3, 1-13.

Figure Legend

Figure 1. An example stimulus. Collocations are shown underlined with the alternative weak and strong collocations separated using a slash. Note that sentence stimuli were shown normally and including either the strong or weak collocation in the experiment.

Figure 1

Predictive context

To wake up, Julie went to the cafeteria and ordered a black coffee / bitter coffee and a sandwich.

Neutral context

My friend and I decided to make some black coffee / bitter coffee this weekend.

Table 1. Summary of stimulus characteristics

Stimulus Characteristic	Strong Collocation	Weak Collocation
Adjective length (letters)	5.9 (.2)	6.1 (.3)
Adjective length (syllables)	1.9 (.1)	2.1 (.1)
Adjective frequency (counts/million)	4.5 (.1)	4.7 (.1)
Frequency of usage (counts/million)	296.7 (59.8)	16.2 (2.9)*
MI score	8.6 (.3)	4.4 (.2)*
Naturalness (Predictable Context)	4.2 (.4)	4.1 (.5)
Naturalness (Neutral Context)	4.1 (.4)	4.1 (.6)

Note. The Standard Error of the Mean, calculated across both participant and item variance, is shown in parentheses. Asterisks indicate statistically significant effects, $p < .05$.

Table 2. Eye movements for the collocation

	Predictable Context		Neutral Context	
	Strong	Weak	Strong	Weak
	Collocation	Collocation	Collocation	Collocation
FPRT (ms)	379 (6)	398 (7)	391 (6)	407 (7)
RO (%)	7 (1)	13 (1)	14 (1)	16 (1)
RPD (ms)	427 (8)	466 (9)	477 (10)	504 (10)
TRT (ms)	473 (9)	539 (11)	536 (11)	602 (12)
RI (%)	19 (1)	24 (2)	28 (2)	31 (2)

Note. The Standard Error of the Mean is shown in parentheses. FPRT = first-pass reading time, RO = regressions-out, RPD = regression path duration, TRT = total reading time, RI = regressions-in.

Table 3. Summary statistics for the collocation phrase

Factor	Statistic	FPRT	RPD	TRT	RI	RO
Intercept (global mean)	β	407.42	487.46	560.65	-1.4	-2.82
	SE	8.49	7.03	7.85	0.21	0.31
	t/z	47.99	69.3	71.42	-6.79	-9.15
Context (predictable-neutral)	β	-10.53	-26.6	-43.18	-0.49	-0.54
	SE	4.68	4.8	6.37	0.09	0.13
	t/z	-2.25*	-5.54*	-6.78*	-5.21*	-4.09*
Collocation (weak-strong)	β	19.36	35.17	70.6	0.28	0.55
	SE	5.8	9.26	7.46	0.09	0.13
	t/z	3.34*	3.8*	9.47*	2.99*	4.19*
Context x Collocation	β	-1.45	4.43	-7.78	0.01	0.34
	SE	7.15	7.83	7.24	0.19	0.26

t/z	-0.20	0.56	-1.07	0.03	1.28
-----	-------	------	-------	------	------

Note. Asterisks indicate statistically significant effects, $p < .05$. FPRT = first-pass reading time, RPD = regression path duration, TRT = total reading time, RO = regressions-out, RI = regressions-in.

Table 4. Eye movement measures for the collocation noun

	Predictable Context		Neutral Context	
	Strong	Weak	Strong	Weak
	Collocation	Collocation	Collocation	Collocation
SKIP (%)	21 (1)	18 (1)	17 (1)	16 (1)
FFD (ms)	213 (2)	224 (3)	221 (3)	226 (3)
SFD (ms)	212 (3)	227 (4)	222 (3)	234 (4)
GD (ms)	228 (3)	241 (4)	241 (4)	241 (3)
RO (%)	6 (1)	11 (1)	8 (1)	11 (1)
RPD (ms)	256 (6)	290 (8)	276 (6)	293 (7)
RI (%)	17 (2)	19 (2)	22 (2)	23 (2)
TRT (ms)	270 (5)	300 (6)	301 (6)	333 (7)

Note. The Standard Error of the Mean, calculated across both participant and item variance, is shown in parentheses. All values are in ms unless otherwise stated. SKIP = word-skipping rate, FFD = first-fixation duration, SFD = single-fixation duration, GD = gaze duration, RO = regressions-out, RPD = regression-path duration, RI = regressions-in, TRT= total reading time.

Table 5. Summary statistics for the collocation noun

Factor	Statistic	SKIP	FFD	SFD	GD	RPD	TRT	RI	RO
Intercept (global mean)	β	-2.59	220.61	225.32	237.87	282.93	302.3	-1.61	-2.59
	<i>SE</i>	0.39	6.83	7.26	6.44	8.47	8.79	0.17	0.38
	<i>t/z</i>	-6.71	32.3	31.03	36.84	33.41	34.38	-9.74	-6.71
Context (predictable-neutral)	β	0.37	-4.97	-8.32	-7.52	-11.33	-23.51	-0.31	0.37
	<i>SE</i>	0.11	2.19	2.66	2.69	3.95	4.18	0.11	0.11
	<i>t/z</i>	3.33*	-2.26*	-3.12*	-2.79*	-2.87*	-5.63*	-2.81*	3.33*
Collocation (weak-strong)	β	-0.16	9.99	15.5	7.95	27.68	32.08	0.14	-0.16
	<i>SE</i>	0.11	4.74	5.96	5.51	8.84	8.39	0.11	0.11
	<i>t/z</i>	-1.48	2.11*	2.6*	1.44	3.13*	3.82*	1.3	-1.48
Context x Collocation	β	-0.19	4.38	-0.12	9.48	10.31	-6.31	0.01	-0.19
	<i>SE</i>	0.22	4.22	5.01	4.99	6.75	7.57	0.22	0.22

<i>t/z</i>	-0.85	-0.02	-0.02	1.89	1.53	-0.83	0.03	-0.85
------------	-------	-------	-------	------	------	-------	------	-------

Note. Asterisks indicate statistically significant effects, $p < .05$. SKIP = word-skipping rate, FFD = first-fixation duration, SFD = single-fixation

duration, GD = gaze duration, RO = regressions-out, RPD = regression-path duration, RI = regressions-in, TRT= total reading time.