From Words to Phrases: How phrase-level frequency affects natural reading in L1 and L2 speakers

Longjiao Sui* (School of Foreign Languages, Dalian Maritime University, Dalian, China)
Wanyin Li* (School of Psychology and Clinical Language Sciences, University of Reading; School of Psychology, University of Nottingham, UK)
Martin R. Vasilev (Department of Experimental Psychology, University College London, UK)
*joint first author

Dr Wanyin Li on the behalf of all authors,

her contact email: wanyin.li.words@gmail.com

Author Contributions

- **Longjiao Sui** collected, cleaned, and analysed the data, and contributed to writing and revising the manuscript.
- **Wanyin Li** conceptualised the study, pre-registered the research, and wrote and finalised the manuscript.
- **Martin R. Vasilev** contributed to the conceptualisation of the study and revised the manuscript.

Abstract

Language is creative yet systematic, featuring frequent multi-word sequences (MWS). Although readers are sensitive to MWS (phrase-level) frequency, controlled sentence experiments provide limited insight into how phrase-level and word-level frequency jointly shape natural reading. Using eye-tracking corpora of narrative reading, GECO (L1 speakers) and GECO-CN (L2 speakers), we analysed 1300 adjective–noun pairs to examine phrase-level and constituent-word frequency effects. Results revealed interactions between the two frequency measures, but with both showing distinct patterns across language proficiency groups. For L1 speakers, word-constituent frequency effects were strongest when phrase-level frequency was low, but weakened or became inhibitory when phrase-level frequency was high. However, for L2 speakers, word-constituent frequency effects increased with greater phrase-level frequency, and overall, both word- and phrase-level effects were stronger than in L1 speakers. These findings suggest that frequent MWS become entrenched in memory and that language proficiency modulates the reliance on phrasal versus lexical information during reading.

Keywords: *multi-word sequences; phrasal frequency; reading; eye-tracking; bilingualism*

Introduction

Human language is highly creative and productive, enabling us to generate an infinite number of novel constructions. Yet language is also systematic and characterised by many frequently co-occurring words, such as "third party", "provide information", or "in the middle of". These *multi-word sequences* (MWS; Sag et al., 2002), also referred to as *formulaic language* (Carrol & Conklin, 2020), *multi-constituent units* (Zang, 2019) or *multi-unit expressions* (Kallens & Christiansen, 2022), serve as fundamental building blocks of language (Arnon & Snider, 2010). Given their prevalence, it is not surprising that language users are highly sensitive to statistical patterns that govern them (Rebuschat & Williams, 2012; Romberg & Saffran, 2010), and the frequency with which certain words co-occur. For example, frequently co-occurring sequences (e.g., "third party", "provide information") are processed more quickly than less common counterparts (e.g., "legal party", "compare information"), even when individual word-level frequency is controlled for (e.g., Arnon & Snider, 2010; Arnon & Christiansen, 2017; Christiansen & Arnon, 2017). This processing advantage illustrates the MWS frequency effect, which is independent of word-level frequency and cannot be reduced to it.

However, much of the existing evidence on the MWS frequency effect, both phrase-level frequency (e.g., the frequency of "third party") and word-level frequency (e.g., the frequency of "third"), comes from highly controlled experimental settings that typically use limited stimulus sets (i.e., N < 160) with single sentences (H Li et al., 2021, 2022; Jiang & Siyanova–Chanturia, 2023; Sonbul, 2015; Vilkaitė & Schmitt, 2019; Vilkaitė, 2016), short discourses (Vilkaitė-Lozdienė, 2022) or judgment-based tasks without sentence contexts (e.g., Öksüz et al., 2021, 2024). As a result, it remains unclear how these frequency effects unfold during natural reading of

larger texts, where language users encounter a broader and more varied range of linguistic constructions embedded in extended, meaningful discourse rather than experimentally constructed sentences.

Importantly, reading continuous texts or narratives engages higher-order cognitive processes, such as sentence-level integration and discourse comprehension, which are generally absent when processing isolated sentences (e.g., Kaakinen et al., 2002; Kaakinen & Hyönä, 2005; Miellet et al., 2007; Miller et al., 2006). Using eye-tracking corpora with narrative reading provides a more ecologically valid method for investigating how phrase-level frequency effects interact with the cognitive demands of real-world language processing. Therefore, the present study analysed natural reading data from two large corpora- GECO (Cop et al., 2017) and GECO-CN (Sui et al., 2023), allowing us to compare how first-language (L1) and second-language (L2) speakers process MWS.

Given the wide variety of MWS types, this study focuses specifically on adjective-noun word pairs, including collocations and less common word pairs. From a corpus-based perspective, collocations are word pairs that co-occur more frequently than expected by chance (Biber et al., 1999). These collocations are typically semantically transparent and literal in meaning, for example, "provide information" or "third party". In the current study, we adopt this corpus-based, frequency-driven approach to collocation, focusing on transparent adjective–noun pairs only. Throughout the remainder of the paper, we use the term MWS to refer specifically to collocations examined here.

A growing body of research has demonstrated that collocations are processed more efficiently than novel word combinations, particularly when defined within the corpus-based and frequency-driven approach. This collocation processing advantage

has been robustly observed across a range of experimental paradigms, including acceptability judgment tasks (e.g., Öksüz et al., 2021, 2024; Vilkaitė-Lozdienė & Conklin, 2021; Wolter & Gyllstad, 2011; Wolter & Yamashita, 2018) and eye-tracking studies (e.g., Jiang & Siyanova–Chanturia, 2023; H Li et al., 2021; Sonbul, 2015). It is also consistently found in both L1 and L2 speakers (e.g., Jiang & Siyanova–Chanturia, 2023; W. Li et al., 2025; Sonbul, 2015), though L2 speakers have weaker representations of MWS (Vilkaitė & Schmitt, 2019; W. Li et al., 2025).

Accounts of the collocation processing advantage are provided by usage-based models of language (Goldberg, 2006; Tomasello, 2003) and single-system models of language processing (e.g., Christiansen & Chater, 2016), which posit that both recurrent MWS and individual words serve as fundamental building blocks of language. This view contrasts with the classic 'words-and-rules' perspective (Pinker, 1999), which assumes that words are stored individually and combined according to grammatical rules. In usage-based and single-system approaches, linguistic structure is not inherently specified but instead emerges from patterns of use: repeated exposure to frequent constructions leads to their entrenchment in memory, thereby facilitating retrieval and processing. Closely aligned with this view are statistical learning theories, which suggest that language users acquire knowledge by tracking distributional regularities in the input (Rebuschat & Williams, 2012; Romberg & Saffran, 2010; Saffran & Kirkham, 2018; Wang & Saffran, 2014). These theories emphasise that learners become sensitive to how often words co-occur and gradually develop expectations based on frequency patterns.

The collocation processing advantage has been demonstrated in several eye-tracking studies of reading. Importantly, this processing advantage appears to be robust to a range of linguistic manipulations, including word adjacency, contextual

predictability, morphological variation, and word order changes. For example, Vilkaitė (2016) examined verb–noun collocations in neutral contexts and found that L1 speakers processed both adjacent (e.g., "provide information") and non-adjacent (e.g., "provide some of the information") collocations faster than novel combinations across both early and late eye-tracking measures. This suggests that the collocational advantage is not dependent on surface word adjacency. Extending this line of research, Li et al. (2021) tested the independent effects of collocation strength and contextual predictability by embedding collocations in either neutral or contextually supportive sentences. Their findings showed that stronger collocations elicited fewer fixations than weaker ones, and collocation strength and contextual predictability independently facilitate processing across both early and late measures.

The effect has also been explored in a paragraph-reading study in Lithuanian that that focused on morphological variations (Vilkaitė-Lozdienė, 2022). The study found that verb–object collocations were read faster than novel combinations, particularly in later fixation measures (e.g., total reading time, total fixation count). While all morphologically varied collocations showed a processing advantage, the canonical form (infinitive + accusative) was processed most efficiently, followed by other forms, such as past tense. These results indicate that the collocation advantage persists across morphological variation, albeit with gradations based on form typicality. Similarly, W. Li et al. (2025) explored the role of word order by comparing canonical (e.g., "he provided information") and reversed collocations (e.g., "information was provided"). Despite the syntactic reversal, reversed collocations were still processed more quickly than their non-collocational counterparts in later eye-tracking measures, suggesting a robust processing benefit comes from the representation of collocational meanings. Taken together, these studies underscore the flexibility of the collocation

processing advantage across different types of linguistic variation.

Importantly, the magnitude of the collocation processing advantage is influenced by both phrase-level frequency and the frequency of individual constituents (e.g., Öksüz et al., 2021, 2024). When collocations are highly frequent, language users are more sensitive to phrase-level frequency, but less to word-level constituent frequency, suggesting that collocations are lexicalised. In contrast, when collocation frequency is lower, language users exhibit greater reliance on the frequency of individual words, indicating a shift toward more compositional, word-based processing. These findings highlight a gradient of lexicalisation in collocation processing: as exposure to a collocation increases, it becomes more entrenched in memory and more likely to be processed as a single unit. However, lexicalisation does not imply the absence of compositional analysis. Even highly frequent collocations do not entirely bypass word-level processing; rather, readers continue to show sensitivity to the frequency of individual words within the collocation. This suggests that collocation processing involves a dynamic interaction between holistic and compositional mechanisms, shaped by experience and frequency of use.

While such findings highlight the importance of frequency at both the collocational and constituent levels, our understanding of how these factors unfold across time in natural reading remains limited. Judgment tasks cannot capture the fine-grained distinction between early processing stages, where individual word frequency is expected to emerge earlier, and later stages, where the phrasal frequency effect is more likely to emerge and strengthen as semantic integration unfolds.

Furthermore, the MWS processing advantage is also moderated by language experience and proficiency. While L2 speakers demonstrate an MWS advantage (W Li et al. 2025; Sonbul, 2015; Vilkaitė & Schmitt, 2019), their representations of MWS

are generally weaker than those of L1 speakers (Vilkaitė & Schmitt, 2019; W. Li et al., 2025). According to the lexical entrenchment hypothesis (Diependaele et al., 2013), greater exposure to a language strengthens lexical representations, lowering activation thresholds and facilitating faster processing. In contrast, limited exposure in L2 leads to weaker representations, slower access, and increased reliance on word-by-word integration of MWS Consequently, the lexicalisation of MWS in L2 speakers appears more attenuated, particularly when constituent frequency is considered. For instance, Ishida (2024) and Wolter and Yamashita (2018) reported stronger effects of phrase-level frequency and constituent frequency in L2 compared to L1 speakers. Similarly, Shi et al. (2023) found that L2 readers exhibited longer reading times for figurative MWS compared to literal ones, a pattern that was not observed in L1 speakers. Furthermore, eye-tracking evidence from Vilkaitė and Schmitt (2019) further showed that L2 speakers, unlike L1 speakers, did not exhibit faster reading for MWS when the constituents were non-adjacent. This suggests that the activation of the second word after encountering the first word of an MWS is weaker in L2 speakers than in L1 speakers.

The Present Study

To investigate how native English speakers (L1) and Chinese L2 speakers of English process MWS during the natural reading of English, we analysed eye-tracking data from two established reading corpora: GECO (for native English speakers) and GECO-CN (for native Chinese speakers of L2 English). Specifically, we examined the influence of word-level frequency (i.e., the frequency of individual words within an MWS) and phrase-level frequency (i.e., the frequency of the entire MWS) on eye-movement patterns during reading a novel. Our study also aimed to determine

whether language proficiency moderates these effects.

Research Questions and Hypotheses

1. Do word-level frequency and phrase-level frequency influence the reading of MWS

We predicted that individual words with higher frequency would be associated with shorter fixation durations and fewer fixations when embedded within an MWS. This prediction is grounded in robust findings of word frequency effects in single-word reading (e.g., Clifton et al., 2016; Inhoff & Rayner, 1986; Rayner, 1998; Rayner & Duffy, 1986; Schilling et al., 1998) and extended to MWS contexts following Sonbul (2015). We further hypothesised that word frequency effects would emerge in early eye-tracking measures, e.g., first fixation duration and first-pass reading time and persist into later measures. In addition, we hypothesised that MWS with higher phrase-level frequency would be read more easily, as reflected in shorter fixation durations and fewer fixations on the MWS (e.g., Jiang & Siyanova–Chanturia, 2023; H. Li et al., 2021, 2022; W. Li et al., 2025). We further expected that this MWS frequency effect would be more prominent in later measures of reading (e.g., total reading time), consistent with the view that facilitation from MWS frequency may reflect semantic-level integration (W. Li et al., 2025).

2. Do word-level frequency and phrase-level frequency interact in shaping reading behaviour?

We predicted an interaction between word-level and phrase-level frequency. Specifically, we expected word-constituent frequency effects to be stronger when phrase-level frequency is low, reflecting a greater reliance on bottom-up, word-level processing. In contrast, when phrase-level frequency is high, we expected reduced word-constituent frequency effects, indicating more holistic, phrase-level processing

(Öksüz et al., 2021, 2024).

3. Are word-level and phrase-level frequency effects stronger for L2 speakers than for L1 speakers?

Drawing on findings from Wolter and Yamashita (2018), we predicted that L2 speakers would show stronger word-constituent frequency effects than L1 speakers. Then, based on findings from Ishida (2024) and theoretical accounts suggesting that bilinguals rely more heavily on distributional information (Diependaele et al., 2013), we predicted that phrase-level frequency would have a stronger influence on reading behaviour in L2 speakers than in L1 speakers.

4. Does L2 proficiency enhance sensitivity to MWS frequency during reading?

Similarly, we hypothesised that L2 language users with lower English proficiency would show stronger phrase-level frequency effects than ones with lower proficiency (Diependaele et al., 2013). We expected this to be reflected in greater reductions in fixation durations and fixation counts for high-frequency MWS among more proficient bilingual readers.

Method

Transparency and Openness

This study and its analysis were preregistered at https://doi.org/10.17605/OSF.IO/RH8T5. We report all data exclusions, all manipulations, and all measures entered into our analysis. All data and analysis code are available at https://osf.io/ksn8m/overview.

Participant

The participants and materials of this study come from the GECO (Cop et al., 2017) and GECO-CN (Sui et al., 2023) corpora. The GECO corpus is an eye-tracking dataset comprising 19 Dutch–English bilinguals (mean *age*: 21.2 years; *SD* = 2.2

years) and 14 British English monolinguals (mean *age*: 21.8 years; *SD* = 5.6 years)who read the full novel *The Mysterious Affair at Styles* by Agatha Christie (1920) while their eye movements were recorded. This study only included the monolingual data from the GECO corpus. Reading took place over four self-paced sessions, each covering a fixed set of chapters. Monolingual participants read the entire book in English. GECO-CN (Sui et al., 2023) extends the original GECO corpus by including data from 30 Chinese–English bilinguals (mean *age*: 25.3; *SD* = 2.6). All participants were undergraduate, master's or Ph.D. students. This dataset follows the same experimental design and uses the same novel as in the original GECO study. Participants read one half of the novel in Chinese (simplified characters) and the other half in English. The language order was counterbalanced: half read Chapters 1–7 in Chinese and 8–13 in English, and the other half in the reversed order.

To evaluate whether our study design has sufficient sensitivity to detect the expected effects, we conducted power simulations using the simr package (1.0.8 version; Green & MacLeod, 2016) in R. Effect size estimates were derived from the regression coefficients reported in Öksüz et al. (2021). Simulated datasets were structured to reflect the GECO corpus: 14 L1 participants contributing 679 observations each, and 30 L2 participants contributing 347 observations each. Power simulations indicated the following sample sizes are required to achieve 80% power: 15 participants for detecting the main effect of phrase-level frequency, 10 participants for the main effect of adjectives frequency, 10 participants for the main effect of noun frequency, 10 participants for the main effect of group, 40 participants for the interaction between noun frequency and phrase-level frequency, and 40 participants for the interaction between adjectives frequency and phrase-level frequency. For the interaction between phrase-level frequency and group, a sample of 45 participants

yields approximately 60% power. Overall, these simulations suggest that the study is generally sufficiently powered to detect the predicted main effects and the key frequency-by-frequency interactions. The only potential limitation is the reduced sensitivity for the phrase-level frequency × group interaction (≈60% power with 45 participants).

Material

The novel *The Mysterious Affair at Styles* by Agatha Christie (1920) was selected based on several criteria: its public domain status (enabling unrestricted use for research), suitable length, accessible vocabulary, and availability in multiple languages to support cross-linguistic comparisons (see Cop et al., 2017 for details). The English version of the novel contains 5242 sentences, 56,841 words and 5,363 unique word types. The average number of words per sentence is 10. 84 ($SD = 8.27$), and the average word length is 4.31 ($SD = 2.41$).

From the novel, we extracted adjacent adjective–noun pairs, including repeated ones, as our target MWS (1332 adjacent adjective–noun pairs). To ensure consistency and avoid confounds, we excluded MWSs that contain capitalised words or punctuation (e.g., hyphens, commas) between the two words, as well as phrases that span across two lines or appear at the beginning or end of a line. Due to differences in line length between the two corpora, 638 phrases were excluded from the bilinguals and 653 from the monolinguals, leaving 694 and 679 phrases for analysis, respectively (see Table 1). The raw frequencies of entire phrases, as well as of their component adjectives and nouns, were extracted from the British English Corpus (counts based on 100 million words; Davies, 2004; see Fig. 1).

Table 1.
Descriptive Statistics of Materials

	L2 speakers			L1 speakers		
	M	**SD**	**Range**	**M**	**SD**	**Range**
Phrase-level Frequency	2.57	2.41	[0-8.66]	2.52	2.38	[0-8.66]
Adjective Frequency	8.73	2.10	[0-11.93]	8.73	2.11	[0-11.93]
Noun Frequency	8.66	1.84	[1.61-12.42]	8.65	1.85	[1.61-12.42]
Phrase-level Length	12.94	3.28	[7-25]	12.94	3.27	[7-25]
Adjective Length	6.12	2.15	[3-14]	6.12	2.16	[3-14]
Noun Length	5.82	2.18	[3-13]	5.81	2.17	[3-13]
MI Scores	2.09	2.10	[0-9.54]	2.02	2.08	[0-9.54]

*The frequencies of nouns, adjectives, and MWS were logarithmically transformed. *Mutual Information (MI) score, a common measure of exclusivity of words*

Figure 1.

Frequency Distributions of Nouns, Adjectives, and MWS between Language Groups
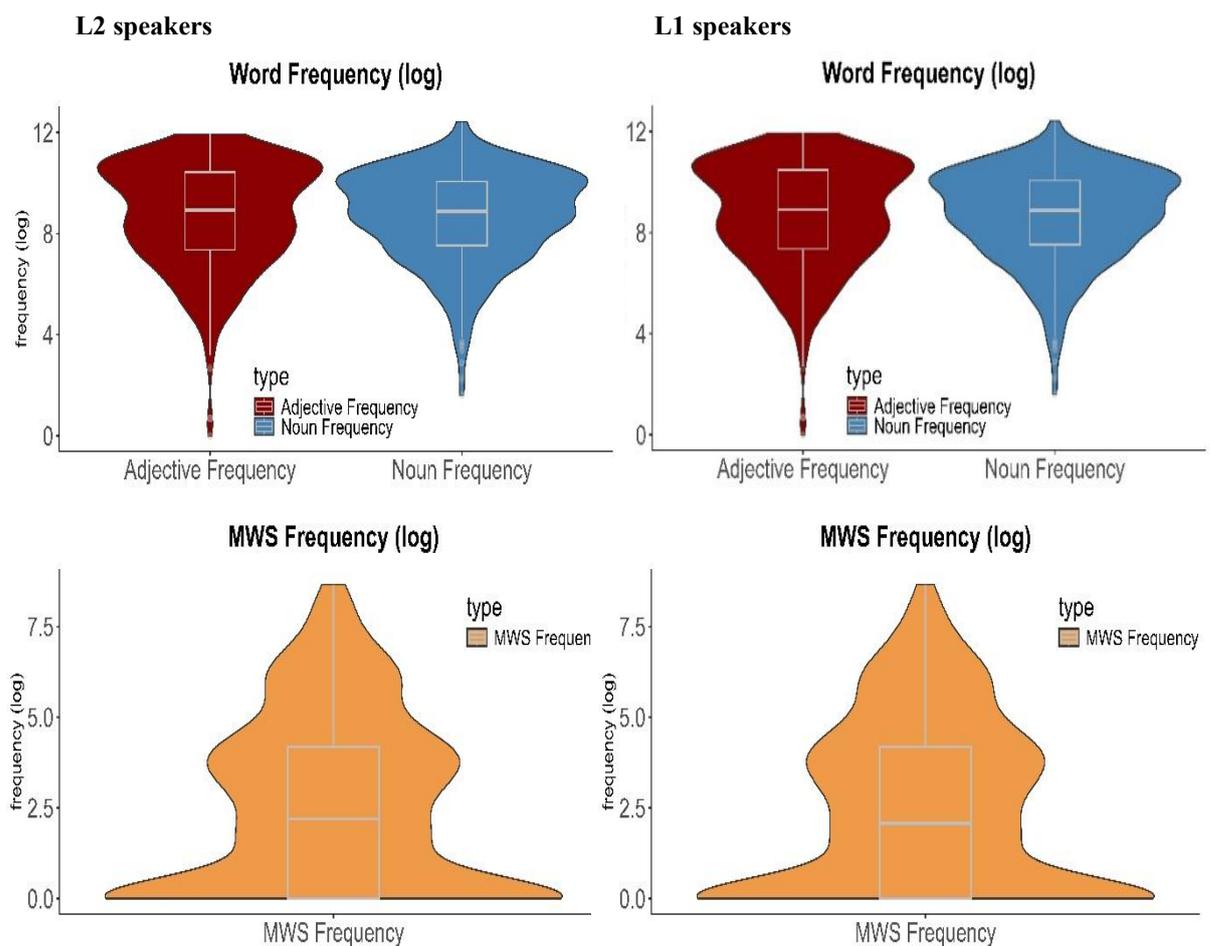
Statistical Analysis

We followed the participant exclusion criteria outlined in Cop et al. (2017) and

Sui et al. (2023). Specifically, one L1 speaker was excluded for having attended only one of four sessions, resulting in insufficient data (<30%). Consequently, the final analysis included 14 L1 speakers and 30 L2 speakers. Furthermore, fixations shorter than 100 ms are not considered reflective of meaningful language processing (Sereno & Rayner, 2003) and were removed from the analysis. In each reading time measure, fixation durations more than 2.5 standard deviations above each participant's mean were excluded (excluded 2.24%–2.58% of outliers for the entire collocation region, 2.34%–2.76% for the adjective region, and 2.26%–2.93% for the noun region).

All analyses were performed in R version 4.4.3 (R Core Team, 2025) with packages lme4 version 1.1-37 (Bates et al., 2015), lmerTest version 3.1-3 (Kuznetsova et al., 2017), emmeans 1.11.1 (Lenth, 2018) and ggplot2 3.5.2 (Wickham, 2016). Data were analysed with linear mixed-effect models (for fixation duration measures) and generalised mixed-effect models (for fixation counts and skipping probability). While previous research on the MWS advantage has typically analysed two reading regions: the entire MWS and its final constituent, the present study also considered the first constituent region to achieve a more comprehensive account of MWS processing (e.g., Kuperman et al., 2009). To test whether L1 and L2 speakers process MWS differently, we added this variable as a grouping factor.

To capture both earlier and later processing, several reading measures were used. Earlier processing measures included first fixation duration (FFD), the durations spent on a word during its initial fixation; first-run reading time (FRRT), the summed durations of all fixations in the first pass of the interest area (i.e., gaze duration). Later processing measures comprised regression path reading time (RPRT), the sum of all fixation durations from entering an interest area until exiting it to the right; selective regression path reading time (SRPRT), which is the sum of fixation durations from

entering an interest area before exiting it to the right, excluding fixations on previous regions; total reading time (TRT), the sum of all fixations on a word, including during regressions, and total fixation count (TF), the total number of fixations in the interest area. Skipping probability was also included, which reflects whether a word is skipped during the first pass reading. Following previous recommendations (Brysbaert & Stevens, 2018), we fitted a log-Gaussian distribution to the reading durations, a Poisson distribution to the fixation counts, and a Binomial distribution to the skipping probability.

In each model, the main predictors were: adjective frequency, noun frequency, phrase-level frequency, group (L1 vs L2 speakers) and covariates were: language proficiency (measured by LexTALE score), noun repetition (the sequential numbering of repeated words throughout 1-2 or 3-4 sessions, including repetitions that were not selected), adjective repetition, phrase-level repetition, and the phrase-level length.. Following Öksüz et al. (2024), we used form frequency for all frequency measures. All continuous factors were scaled and centred, with frequency predictors first log-transformed. The categorical factor, Group was summed coded (L1 speakers = -1, L2 speakers = 1). The correlation between Group and language proficiency was moderate ($r < 0.6$), therefore, we included both variables in each model.

In all models, the maximal random structure model included the interactions of word- and phrase-level frequency with participants, along with random intercepts for each word token (Barr et al., 2013). If convergence did not occur, "bobyqa" or "optimx" optimiser (Powell, 2009) was used to increase iterations, before simplifying random effects by removing insignificant interactions and then variables in the order of least variance explained until the model converged (Veldre & Andrews, 2014). Fixed effects were subsequently simplified by comparing reduced models with the

full model using the *anova()* function; non-significant interactions and main effects were removed when their exclusion did not result in model convergence failure. Based on these tests, reduced models were reached to fit the data best for each dependent measure. In all models, p < 0.05 are considered significant. Additionally, the Variance Inflation Factors (VIFs) were calculated with the *vif()* function from the car package (Fox et al., 2023) to check for collinearity. All VIFs were around 2, indicating a very low degree of correlation among predictor variables.

Results

The full results for the adjective–noun combination region are reported in Table 2. To provide a more comprehensive account of MWS processing, we conducted additional analyses of the adjective region that were not part of the pre-registered plan. The results of Adjective and Noun region are reported in the Appendix.

MWS Region

A robust main effect of group was observed across all measures: L1 speakers read entire adjective–noun phrases more quickly, made fewer and shorter fixations, and skipped phrases more often than L2 speakers, as shown in Table 2. The interaction between group and phrase-level frequency was observed: in FFD, L2 speakers showed a facilitative frequency effect, while L1 speakers exhibited an inhibitory effect.

Table 2

Analysis of Eye-tracking Measures for the Entire Phrase

		Estimate	Std.	t value	Pr(>|t|)	VIF
FIRST FIXATION	**(Intercept)**	**5.42**	**0.02**	**269.11**	**< 0.001*****	
	Group L2	**0.06**	**0.02**	**3.13**	**0.003****	**1.25**
	Adjective Frequency	**-0.02**	**0.00**	**-5.84**	**< 0.001*****	**1.22**
	Noun Frequency	0.00	0.00	-0.77	0.445	1.56
	Phrase-level Frequency	0.00	0.00	-0.19	0.852	1.67
	Group L2:Adjective Frequency	-0.01	0.00	-1.66	0.103	1.45

		Estimate	Std. Error	t value	p	VIF
	Group L2:Noun Frequency	0.00	0.00	0.52	0.603	1.45
	Group L2:Phrase-level Frequency	**-0.01**	**0.00**	**-2.34**	**0.019***	**1.76**
	Noun Frequency:Phrase-level Frequency	-0.01	0.00	-1.57	0.118	1.42
	Group L2:Noun Frequency:Phrase-level Frequency	0.00	0.00	1.66	0.096.	1.40
FIRST RUN READING TIMES	**(Intercept)**	**6.01**	**0.03**	**203.64**	**< 0.001***	
	Group L2	**0.24**	**0.03**	**7.10**	**< 0.001***	**1.81**
	Adjective Frequency	**-0.06**	**0.01**	**-6.66**	**< 0.001***	**1.54**
	Noun Frequency	**-0.04**	**0.01**	**-4.32**	**< 0.001***	**2.27**
	Phrase-level Frequency	0.00	0.01	-0.23	0.821	1.91
	Language Proficiency	**-0.07**	**0.03**	**-2.49**	**0.017***	**1.44**
	Phrasal Length	**0.08**	**0.01**	**12.83**	**< 0.001***	**1.25**
	Phrasal Repetition	**-0.01**	**0.01**	**-2.73**	**0.006****	**1.15**
	Noun Repetition	**-0.02**	**0.01**	**-2.77**	**0.006****	**1.44**
	Group L2:Adjective Frequency	-0.01	0.01	-1.37	0.176	1.58
	Group L2:Noun Frequency	**-0.02**	**0.00**	**-3.51**	**0.001***	**1.80**
	Group L2:Phrase-level Frequency	0.00	0.01	0.25	0.801	2.00
	Adjective Frequency:Phrase-level Frequency	**-0.02**	**0.01**	**-2.08**	**0.038***	**1.56**
	Noun Frequency:Phrase-level Frequency	**-0.02**	**0.01**	**-2.47**	**0.014***	**1.74**
	Group L2:Adjective Frequency:Phrase-level Frequency	**-0.01**	**0.00**	**-2.79**	**0.005****	**1.60**
	Group L2:Noun Frequency:Phrase-level Frequency	**-0.02**	**0.00**	**-3.40**	**0.001***	**1.55**
REGRESSION PATH READING TIMES	**(Intercept)**	**6.26**	**0.03**	**190.99**	**< 0.001***	
	Group L2	**0.25**	**0.04**	**6.63**	**< 0.001***	**1.67**
	Adjective Frequency	**-0.06**	**0.01**	**-5.94**	**< 0.001***	**1.79**
	Noun Frequency	**-0.05**	**0.01**	**-6.04**	**< 0.001***	**1.93**
	Phrase-level Frequency	-0.01	0.01	-1.09	0.278	2.04
	Language Proficiency	-0.05	0.03	-1.52	0.136	1.46
	Phrasal Length	**0.09**	**0.01**	**12.83**	**< 0.001***	**1.33**
	Adjective Repetition	**-0.03**	**0.01**	**-5.08**	**< 0.001***	**1.16**
	Group L2:Adjective Frequency	**-0.02**	**0.01**	**-3.28**	**0.002****	**1.56**
	Group L2:Noun Frequency	**-0.01**	**0.00**	**-2.85**	**0.004****	**1.72**
	Group L2:Phrase-level Frequency	**0.03**	**0.01**	**3.97**	**< 0.001***	**2.64**
	Adjective Frequency:Phrase-level Frequency	-0.01	0.01	-1.34	0.182	1.59
	Noun Frequency:Phrase-level Frequency	0.00	0.01	-0.29	0.770	1.60
	Phrase-level Frequency:Language Proficiency	**0.01**	**0.00**	**2.67**	**0.008****	**1.54**
	Group L2:Adjective Frequency:Phrase-level Frequency	**-0.02**	**0.00**	**-4.12**	**< 0.001***	**1.56**
	Group L2:Noun Frequency:Phrase-level Frequency	**-0.01**	**0.00**	**-2.84**	**0.004****	**1.57**
SELECTIVE REGRESSION PATH READING TIMES	**(Intercept)**	**6.16**	**0.03**	**193.31**	**< 0.001***	
	Group L2	**0.27**	**0.04**	**7.34**	**< 0.001***	**1.78**
	Adjective Frequency	**-0.06**	**0.01**	**-6.11**	**< 0.001***	**1.69**
	Noun Frequency	**-0.04**	**0.01**	**-4.77**	**< 0.001***	**2.14**
	Phrase-level Frequency	-0.02	0.01	-1.74	0.083.	1.89
	Language Proficiency	**-0.07**	**0.03**	**-2.19**	**0.034***	**1.49**
	Phrasal Length	**0.10**	**0.01**	**15.50**	**< 0.001***	**1.25**
	Adjective Repetition	**-0.02**	**0.01**	**-3.69**	**0.000***	**1.24**
	Noun Repetition	**-0.02**	**0.01**	**-2.78**	**0.005****	**1.41**
	Group L2:Adjective Frequency	**-0.02**	**0.01**	**-3.43**	**0.001****	**1.54**
	Group L2:Noun Frequency	**-0.02**	**0.00**	**-3.46**	**0.001****	**1.52**
	Group L2:Phrase-level Frequency	**0.02**	**0.01**	**4.06**	**< 0.001***	**2.42**

	Estimate	SE	t/z	p	VIF
Adjective Frequency:Phrase-level Frequency	**-0.02**	**0.01**	**-2.11**	**0.035***	**1.58**
Noun Frequency:Phrase-level Frequency	0.00	0.01	-0.24	0.809	1.71
Phrase-level Frequency:Language Proficiency	0.01	0.00	1.96	0.051.	1.57
Group L2:Adjective Frequency:Phrase-level Frequency	**-0.02**	**0.00**	**-5.32**	**< 0.001*****	**1.56**
Group L2:Noun Frequency:Phrase-level Frequency	**-0.01**	**0.00**	**-3.50**	**< 0.001*****	**1.50**

TOTAL READING TIMES

	Estimate	SE	t/z	p	VIF
(Intercept)	**6.26**	**0.03**	**182.50**	**< 0.001*****	
Group L2	**0.33**	**0.03**	**9.92**	**< 0.001*****	**1.34**
Adjective Frequency	**-0.06**	**0.01**	**-5.47**	**< 0.001*****	**1.85**
Noun Frequency	**-0.04**	**0.01**	**-4.20**	**< 0.001*****	**2.18**
Phrase-level Frequency	**-0.03**	**0.01**	**-2.63**	**0.01***	**1.93**
Phrasal Length	**0.09**	**0.01**	**12.10**	**< 0.001*****	**1.27**
Phrasal Repetition	**-0.01**	**0.01**	**-2.08**	**0.04***	**1.22**
Adjective Repetition	**-0.02**	**0.01**	**-3.05**	**< 0.001***	**1.35**
Noun Repetition	**-0.02**	**0.01**	**-2.78**	**0.01***	**1.49**
Group L2:Adjective Frequency	**-0.02**	**0.01**	**-3.09**	**< 0.001***	**1.59**
Group L2:Noun Frequency	**-0.02**	**0.01**	**-3.87**	**< 0.001*****	**1.77**
Group L2:Phrase-level Frequency	**0.01**	**0.01**	**1.92**	**0.06.**	**1.91**
Adjective Frequency:Phrase-level Frequency	-0.01	0.01	-0.67	0.51	1.61
Noun Frequency:Phrase-level Frequency	0.00	0.01	0.25	0.81	1.73
Group L2:Adjective Frequency:Phrase-level Frequency	**-0.02**	**0.00**	**-4.57**	**< 0.001*****	**1.60**
Group L2:Noun Frequency:Phrase-level Frequency	**-0.02**	**0.00**	**-4.20**	**< 0.001*****	**1.53**

FIXATION COUNT

	Estimate	SE	t/z	p	VIF
(Intercept)	**0.98**	**0.03**	**29.33**	**< 0.001*****	
Group L2	**0.30**	**0.03**	**9.38**	**< 0.001*****	**1.19**
Adjective Frequency	**-0.05**	**0.01**	**-4.30**	**< 0.001*****	**2.16**
Noun Frequency	**-0.04**	**0.01**	**-3.10**	**< 0.001***	**2.56**
Phrase-level Frequency	**-0.03**	**0.01**	**-2.41**	**0.02***	**2.16**
Phrasal Length	**0.09**	**0.01**	**10.84**	**< 0.001*****	**1.38**
Adjective Repetition	**-0.02**	**0.01**	**-2.52**	**0.01***	**1.32**
Noun Repetition	**-0.03**	**0.01**	**-3.61**	**< 0.001*****	**1.45**
Group L2:Adjective Frequency	**-0.01**	**0.01**	**-1.94**	**0.05.**	**1.89**
Group L2:Noun Frequency	**-0.02**	**0.01**	**-3.05**	**< 0.001***	**2.06**
Group L2:Phrase-level Frequency	**0.01**	**0.01**	**2.00**	**0.05***	**2.20**
Adjective Frequency:Phrase-level Frequency	0.00	0.01	0.05	0.96	1.68
Noun Frequency:Phrase-level Frequency	0.01	0.01	0.95	0.34	1.86
Group L2:Adjective Frequency:Phrase-level Frequency	**-0.02**	**0.01**	**-2.72**	**0.01***	**1.60**
Group L2:Noun Frequency:Phrase-level Frequency	**-0.01**	**0.01**	**-2.42**	**0.02***	**1.76**

SKIPPING PROBABILITY

|	Estimate	SE	t/z	p	VIF
(Intercept)	**-2.93**	**0.13**	**-23.01**	**< 0.001*****	
Group L2	**-0.62**	**0.13**	**-4.92**	**< 0.001*****	**1.00**
Adjective Frequency	**0.08**	**0.04**	**2.00**	**0.05***	**1.70**
Noun Frequency	0.01	0.04	0.26	0.79	1.37
Phrase-level Frequency	0.02	0.04	0.51	0.61	1.60
Phrasal Length	**-0.15**	**0.04**	**-3.89**	**< 0.001*****	**1.45**
Adjective Repetition	**-0.11**	**0.04**	**-2.87**	**< 0.001***	**1.29**

*p < .05. **p < .01. ***p < .001

The main effect of adjective frequency was consistent across all measures.

Higher adjective frequency facilitated phrase reading, leading to shorter reading times, fewer fixations, and increased skipping. Group x adjective frequency was observed in SRPRT, RPRT, and TRT. Phrase-level frequency x adjective frequency was only found in FRRT. Significant three-way interactions among group, adjective frequency, and phrase-level frequency emerged in most measures (except skipping probability and first-fixation duration). For L2 speakers, higher adjective frequency resulted in faster phrase reading, and this facilitative effect became stronger as phrase-level frequency increased (see Figure. 2). In contrast, L1 speakers showed a facilitative effect of adjective frequency primarily when phrase-level frequency was low, particularly in later measures (e.g., TRT), suggesting different integration dynamics of lexical and phrase-level information across groups.

Figure 2
Three-way interaction among Group, Phrase-level frequency, and Adjective frequency in the MWS region.
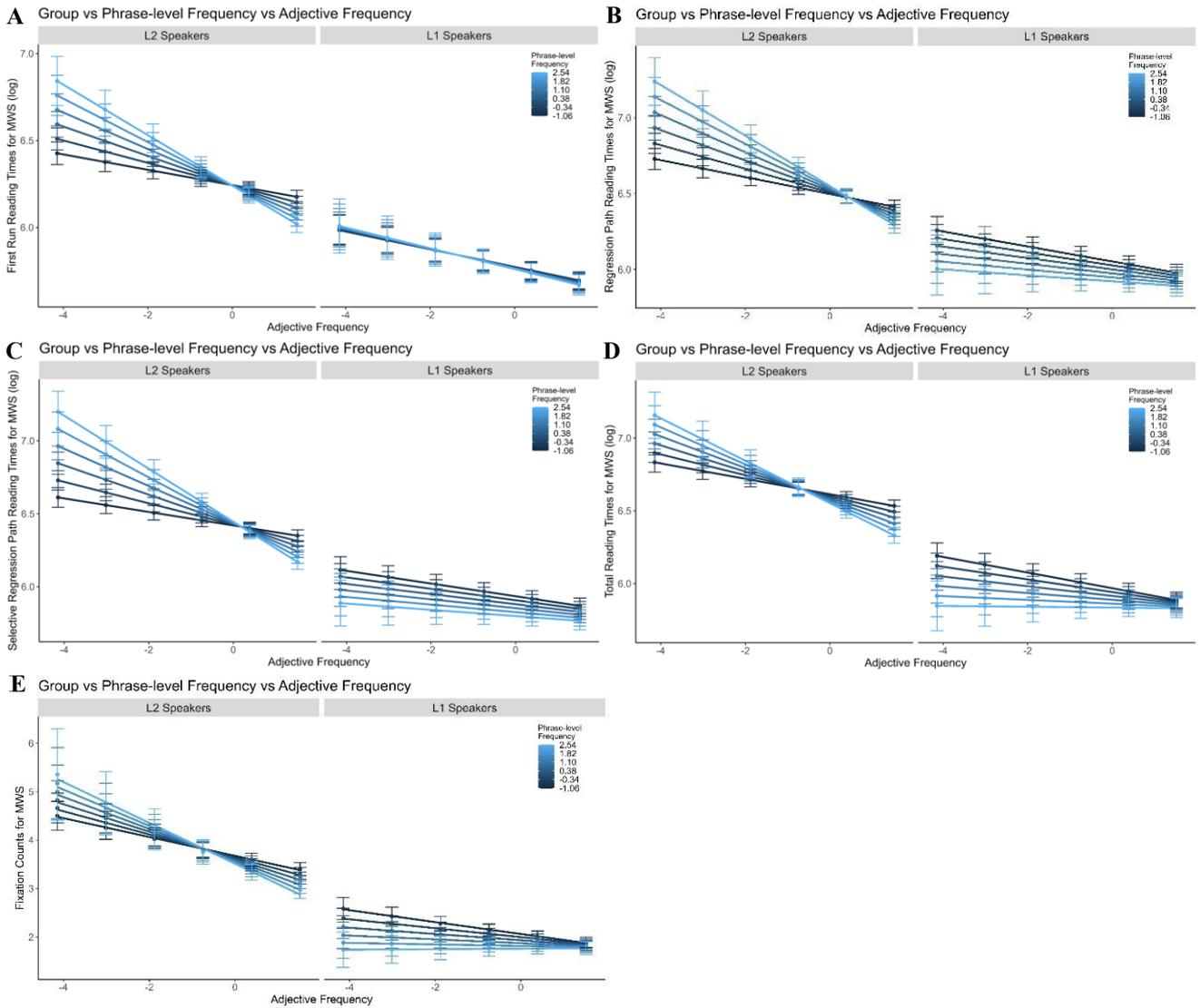
The main effect of noun frequency emerged in most measures (except for first fixation and skipping probability). Higher noun frequency led to shorter reading times and fewer fixations overall. Group x noun frequency was found in most measures as well (except for first fixation and skipping probability). Phrase-level frequency x noun frequency only emerged in FRRT. Significant three-way interactions between group, noun frequency, and phrase-level frequency were observed in the same measures as in the main effect of noun frequency. For L2 speakers, noun frequency facilitated phrase

reading, and this effect increased as phrase-level frequency rose, indicating that frequent phrases amplified word-level faciliation (see Fig. 3). For L1 speakers, the facilitative effect of noun frequency was most evident when phrase-level frequency was low, again showing a contrasting pattern of frequency integration between groups.
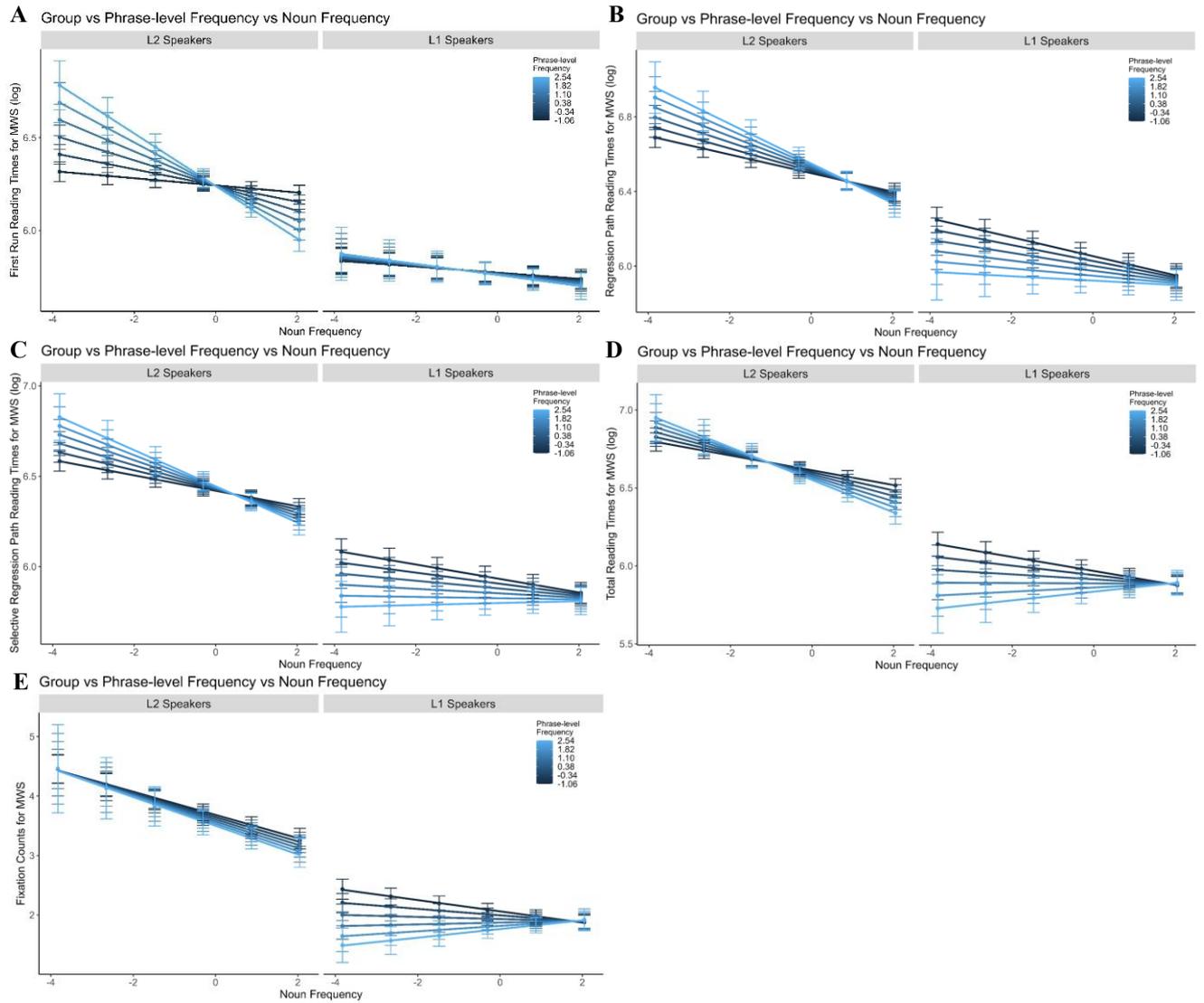
Figure 3

Three-way interaction among Group, Phrase-level frequency, and Noun frequency in the MWS region.

Phrase-level frequency exerted main effects in measures of overall reading behaviour: TRT and fixation counts. Higher phrase-level frequency was generally associated with shorter reading times and fewer fixations. However, phrase-level frequency interacted systematically with both noun and adjective frequency, and these effects differed by group (see details above).

Finally, language proficiency affected FRRT and SRPRT. Higher proficiency was associated with shorter phrase reading times, reflecting more efficient lexical integration. A significant phrase-level frequency × proficiency interaction in RPRT indicated that high phrase-level frequency was linked to shorter reading times among less proficient readers (see Fig. 4).

Figure 4
Two-way interaction between phrase-level frequency and language proficiency in the MWS region.
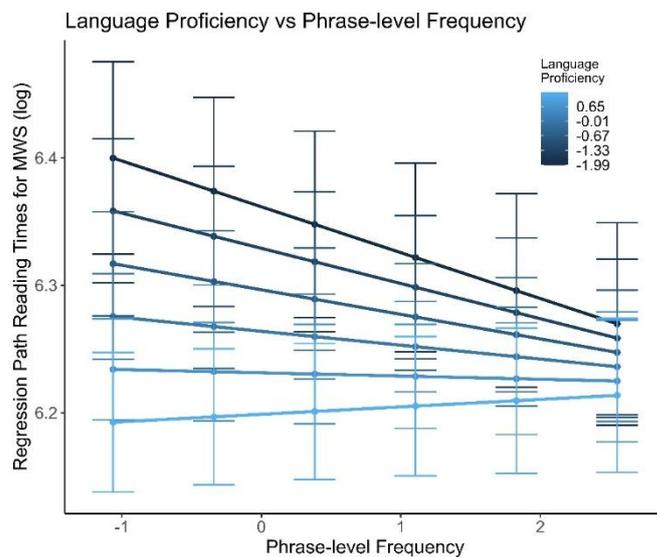
Discussion

Prior research provides compelling evidence that the frequency of multi-word sequences (MWS) influences both language acquisition (Arnon & Clark, 2011) and language processing (Arnon & Snider, 2010), as well as demonstrating effects of multiple distributional frequencies, including individual word and phrase-level frequencies (Öksüz et al., 2021, 2024; Shao et al., 2019). To our knowledge, this is the first study to examine these multiple frequency effects in the context of natural reading for both L1 and L2 speakers.

This work first aims to explore the influence of word- and phrase-level frequency on MWS, i.e., two-word pairs. As predicted, both word-level and phrase-level frequency exerted robust facilitative effects on reading behaviour. Adjectives and nouns with higher frequency were read more quickly, had fewer and shorter fixations, and were skipped more often than low-frequency words. Similarly, collocations with higher phrase-level frequency were associated with reduced fixation durations and fewer fixations, particularly in later measures such as TRT. Contrary to our prediction, the phrase-level frequency effect emerged in earlier measures and indeed became larger in later measures.

The second objective was to explore the interaction between word- and phrase-level frequency. The third goal was to examine whether L2 speakers different interactions between phrase- and word-level frequency compared to L1 speakers. The predicted interaction between word-level and phrase-level frequency was observed but manifested differently across groups. In line with our hypothesis, L1 speakers showed stronger faciliatory word frequency effects when phrase-level frequency was low and reduced or inhibitory frequency effects when phrase-level frequency was high. This pattern suggests that for familiar, high-frequency MWS, L1 readers rely on

holistic, top-down processing rather than on the lexical properties of individual words. For L2 speakers, however, the pattern diverged from predictions. Rather than a reduction in word frequency effects at high phrase-level frequency, L2 speakers exhibited additive facilitation: word frequency effects became stronger as phrase-level frequency increased. This indicates that L2 readers benefit jointly from lexical and phrasal familiarity, relying on both sources of frequency-based information. These findings suggest that while L1 readers flexibly shift from bottom-up to top-down processing depending on phrase familiarity, L2 readers engage in a more uniformly frequency-driven, compositional mode of processing. Consistent with predictions, L2 speakers were more sensitive to frequency effects than L1 speakers. L2 speakers showed stronger facilitative effects of both word-level and phrase-level frequency.

The last aim was to investigate whether proficiency increased the MWS frequency effect. As predicted, higher proficiency was associated with overall faster and more efficient reading, reflected in shorter regression-path durations and higher skipping probabilities. The interactions between proficiency and phrase-level frequency revealed a more nuanced pattern. Less proficient readers showed stronger facilitative effects of phrase-level frequency. In contrast, more proficient readers exhibited weaker or even inhibitory effects.

Phrase-level Frequency Effect

This study demonstrates a clear MWS processing advantage: high phrase-level frequency collocations are read more quickly than low-frequency ones by both L1 and L2 speakers, replicating previous studies (e.g., Arnon & Snider, 2010; Carrol & Conklin, 2021; Öksüz et al., 2021; Siyanova & Schmitt, 2008; Sonbul, 2015; Vilkaitė, 2016; Vilkaitė & Schmitt, 2019; Wolter & Gyllstad, 2011, 2013). This indicates that language users are sensitive not only to word-level frequency but also to the

frequency of larger linguistic constructions (Arnon et al., 2017; Arnon & Snider, 2010). The findings are therefore consistent with usage-based approaches to L1 and L2 acquisition (Bybee, 1998; Ellis & Wulff, 2020; Tomasello, 2003, 2009), which propose that learners acquire a repertoire of constructions of varying form, size, and complexity, all shaped by frequency. Moreover, the phrase-level frequency effect reflects sensitivity to word co-occurrence patterns, in line with statistical learning accounts that highlight how both L1 and L2 speakers track recurring distributional regularities (Romberg & Saffran, 2010; Saffran & Kirkham, 2018; Wang & Saffran, 2014). Taken together, these results support a single-system view of language (Elman, 2009; McClelland et al., 2010). The parallel frequency effects for both words and larger MWS make it difficult to argue that the two are stored or processed in qualitatively different systems (contra Pinker & Ullman, 2002). Instead, the findings suggest continuity across linguistic units, with frequency shaping representations at multiple levels of granularity.

Additionally, the phrase-level frequency effect speaks to ongoing debates about how multi-word information is accommodated within computational models of eye-movement control, particularly the E-Z Reader model (Reichle et al., 2003, 2011). E-Z Reader assumes that lexical processing is largely driven by the frequency of individual words and unfolds in a predominantly serial fashion, with attention allocated to one word at a time. Under this framework, lexical processing of word n + 1 typically begins only after word n has reached a sufficient level of lexical completion, aside from limited cases involving very short or highly frequent words (Reichle & Schotter, 2020).

At first glance, effects of phrase-level frequency might appear difficult to reconcile with a strictly word-by-word architecture. However, recent work has

suggested that multi-constituent sequences can, under certain conditions, be represented and processed as single lexical units, thereby remaining compatible with serial processing assumptions (e.g., Zang et al., 2019). From this perspective, highly frequent phrases may function similarly to long words in the lexicon, without fundamentally undermining the core mechanisms of the model.

Our findings extend this line of inquiry by showing that co-occurrence statistics modulate reading behaviour even when moving beyond a small set of highly frequent, lexicalised expressions. Because our operationalisation captures a broader range of distributional relationships—including relatively low-frequency co-occurrences—our results suggest that sensitivity to MWS frequency may be more general and gradient than previously assumed. At the same time, we acknowledge that corpus-based measures of co-occurrence provide indirect evidence regarding the time course of serial attention allocation.

Stronger Phrase-level Effect in L2

This study is the first to demonstrate, using eye movements on a big natural corpus, that phrase-level frequency effects are stronger in L2 than in L1 speakers. This result is consistent with earlier findings using the phrasal-decision paradigm (Ishida, 2024; but see Hernández et al., 2016) and extends earlier evidence of larger word-level frequency effects in L2 (e.g., Duyck et al., 2008; Whitford & Titone, 2017). Importantly, these phrase-level effects emerged already in earlier measures of reading (FFRT and FRRT) and continued to more global and composite measures of reading (e.g., TRT). In addition, the current results replicate earlier observations, showing that L2 speakers are more strongly influenced by word-level frequency than L1 speakers (Cop et al., 2015; Lemhöfer et al., 2008; Sui et al., 2025). Collectively, these findings indicate that both L1 and L2 speakers rely on word- and phrase-level

frequency during reading, with L2 speakers being more strongly affected.

These results can be explained by existing theoretical models, such as such as lexical entrenchment (e.g., Diependaele et al., 2013), the BIA+ model (Dijkstra & van Heuven, 2002), or the "weaker links" hypothesis (Gollan et al., 2005), all of which predict larger frequency effects in L2 speakers due to reduced language exposure, which influence either the strengthen of representations or the activation speed (Diependaele et al., 2013; Gollan et al., 2005). Extending this logic to phrase-level processing, these models can also account for the observed interaction between language proficiency and phrase-level frequency in the current study. That is, since L2 speakers with lower language proficiency typically have less exposure to the language, they are expected to exhibit larger phrase-level frequency effects compared to more proficient participants, analogous to the patterns at word-level frequency. This assumption aligns with prior findings that word frequency effects are typically larger in L2 than in L1, and, within L1, are more pronounced for less proficient readers, reflecting their greater reliance on frequently encountered linguistic patterns during processing (e.g., Mor & Prior, 2022).

Interaction between Word and Phrase-level Frequencies

The interaction between word- and phrase-level frequency across both speaker groups suggests a shift in the dynamics of MWS processing, with distinct manifestations in each group. On the one hand, L1 speakers showed stronger faciliatory word frequency effects when MWS were uncommon and reduced or inhibitory frequency effects when MWS were frequent. This suggests that highly frequent MWS become entrenched in memory and are accessed more holistically, reducing reliance on their constituent words. Consistent with Öksüz et al. (2021, 2024) with their L1 speakers and Arnon and Cohen (2014), participants were less

sensitive to word-level frequency when collocational frequency was higher, supporting a gradient of lexicalisation whereby repeated exposure promotes entrenchment. Crucially, unlike previous studies (Arnon & Cohen, 2014; Öksüz et al., 2021, 2024), the present study found that for highly frequent MWSs, word-level effects not only diminished but became inhibitory. This supports the view that constituent and holistic representations co-exist along a continuum of cognitive prominence: at the high-frequency end, entrenched MWS are retrieved as unified lexical units, and activation of their components may be not only unnecessary but also disruptive. On the other hand, L2 speakers exhibited additive facilitation, with word frequency effects increasing as phrase-level frequency increased. This pattern suggests that L2 readers benefit jointly from lexical and phrasal frequency, relying on both sources of frequency-based information more than L1 speakers during reading. In other words, L2 speakers primarily engage in bottom-up, even though their reading is also modulated by phrase-level frequency.

The observed divergence in how L1 and L2 speakers integrate word- and phrase-level frequency information may be understood within the framework of compensatory processing (Stanovich, 1986). Readers with lower processing efficiency compensate for weaker lexical or automatic processes by relying more heavily on contextual or distributional cues. Applied to the present findings, L2 speakers, who generally exhibit less automatised access to MWSs and slower integration, appear to recruit multiple sources of frequency information simultaneously. Their reliance on both word-level and phrase-level frequency may thus reflect a compensatory mechanism aimed at reducing processing effort through the use of redundant predictive cues.

Our findings for L2 speakers diverge from those reported by Öksüz et al. (2021,

2024), potentially reflecting differences in task demands and the degree of representational entrenchment across groups. In their phrase-decision task, both L1 and L2 speakers showed reduced noun frequency effects at high phrase-level frequency, suggesting reliance on stored MWS during isolated access. In contrast, our eye-tracking data revealed additive effects for L2 readers only, indicating that their weaker MWS representations promote concurrent use of lexical and phrasal frequency cues during incremental comprehension. While L2 readers can rely on phrase-level frequency in controlled decision tasks efficiently, natural reading requires continuous lexical, syntactic, and semantic integration, prompting the recruitment of multiple frequency-based sources to support fluent processing.

Our findings also provide some evidence for the Multi-Constituent Unit Hypothesis (Zang, 2019; Zang et al., 2021, 2023, 2024). Language users can store lexical entries not only for individual words but also for highly frequent MWS such as idioms and familiar four-character phrases. Such co-occurring words can be processed as unified lexical units, showing that reading is flexible and dynamically shaped by lexical status rather than rigidly constrained by orthographic word boundaries. Our results also extend this account by demonstrating that the degree of lexicalisation is moderated by phrase-level frequency: not all MWS are lexicalised to the same extent. Moreover, while higher phrase-level frequency increases the prominence of MWS representations, information about constituent words remains available, suggesting a layered system in which holistic and componential representations co-exist.

Therefore, our findings challenge the notion that MWS are represented strictly holistically (e.g., Wray, 2008). Under a strictly holistic account, speakers would not retain access to the constituent elements, and word-level frequency effects-nor interactions between word-level and phrase-level frequency-should not arise.

However, our data show MWS processing appears to rely on a dynamic representational system, involving interconnected representations at both the word level and the MWS level. These representations are bidirectionally linked, allowing activation to spread between constituents and the sequence. This interpretation resonates with psycholinguistic accounts of idiom and complex word processing (Kuiper et al., 2007; Kuperman et al., 2009; Sprenger et al., 2006; Sandra et al., 1999). For example, studies of idiom production and comprehension have shown that constituent lemmas remain activated during processing, even when the expression has a stored, conventionalised meaning (e.g., Sprenger et al., 2006; Sandra et al., 1999). In production, idioms appear to be retrieved via a superlemma that specifies the phrasal meaning while simultaneously activating their constituent words, suggesting parallel access to both phrasal and word-level representations.

Conclusion

This study examined how word-level and phrase-level frequency jointly shape the reading of adjective–noun MWS in L1 and L2 English speakers. We found clear effects of lexical and phrasal frequency in both L1 and L2 readers, as well as systematic interactions between them. Readers dynamically integrate both sources of frequency information as they move through a text. For L1 speakers, word-level frequency mattered most when phrase-level frequency was low and diminished or reversed for highly frequent phrases, consistent with more MWS level processing. L2 speakers showed the opposite pattern: lexical frequency effects strengthened as phrase-level frequency increased, indicating that frequent MWS boost lexical facilitation. Overall, L2 speakers showed larger frequency effects, with proficiency further modulating sensitivity to distributional cues. These findings imply that language users are sensitive to multiple levels of frequency information, and their

reliance on lexical versus phrasal cues shifts with language experience and exposure.

Reference

Arnon, I., & Christiansen, M. H. (2017). The role of multiword building blocks in explaining L1–L2 differences. *Topics in cognitive science*, 9(3), 621-636.

Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth–Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, 7(2), 107-129.

Arnon, I., & Priva, U. C. (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., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of memory and language*, 62(1), 67-82.

Arnon, I., McCauley, S. M., & Christiansen, M. H. (2017). Digging up the building blocks of language: Age-of-acquisition effects for multiword phrases. *Journal of Memory and Language*, 92, 265-280.

Barr, D. J. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in psychology*, 4, 328.

Bates, D., Maechler, M., Bolker, B., Walker, S., Christensen, R. H. B., Singmann, H., ... & Bolker, M. B. (2015). Package 'lme4'. convergence, 12(1), 2.

Biber, D., Johansson S., Leech G., Conrad S. & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.

Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, 1(1), 9.

Bybee, J. (1998). *The emergent lexicon*. In Chicago Linguistic Society (Vol. 34, No. 2, pp. 421-435).

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

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and speech*, 63(1), 95-122.

Christiansen, M. H., & Arnon, I. (2017). More than words: The role of multiword sequences in language learning and use. *Topics in cognitive science*, 9(3), 542-551.

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and brain sciences*, 39, e62.

Clifton, C., Jr., Ferreira, F., Henderson, J. M., Inhoff, A. W., Liversedge, S. P., Reichle, E. D., & Schotter, E. R. (2016). Eye movements in reading and information processing: Keith Rayner's 40 year legacy. *Journal of Memory and Language*, 86, 1–19. https://doi.org/10.1016/j.jml.2015.07.004

Contreras Kallens, P., & Christiansen, M. H. (2022). Models of language and multiword expressions. *Frontiers in Artificial Intelligence*, 5, 781962.

Cop, U., Dirix, N., Drieghe, D., & Duyck, W. (2017). Presenting GECO: An eyetracking corpus of monolingual and bilingual sentence reading. *Behavior research methods*, 49(2), 602-615.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic bulletin & review*, 22(5), 1216-1234.

Cowie, A. P. (Ed.). (1998). *Phraseology: Theory, analysis, and applications*. OUP Oxford.

Davies, Mark. (2004) British National Corpus (from Oxford University Press).

   Available online at https://www.english-corpora.org/bnc/.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in

   first-and second-language word recognition: A lexical entrenchment

   account. *Quarterly journal of experimental psychology*, 66(5), 843-863.

Dijkstra, T., & Van Heuven, W. J. (2002). The architecture of the bilingual word

   recognition system: From identification to decision. *Bilingualism: Language

   and cognition*, 5(3), 175-197.

Durrant, P. & Doherty, A. (2010). Are high-frequency collocations psychologically

   real? Investigating the thesis of collocational priming. *Corpus Linguistics and

   Linguistic Theory*, 6(2), 125-155. https://doi.org/10.1515/cllt.2010.006

Ellis, N. C., & Wulff, S. (2020). *Usage-based approaches to L2 acquisition*.

   In Theories in second language acquisition (pp. 63-82). Routledge.

Elman, J. L. (2009). On the meaning of words and dinosaur bones: Lexical knowledge

   without a lexicon. *Cognitive science*, 33(4), 547-582.

Fox J, Weisberg S (2019). *An R Companion to Applied Regression*, Third edition.

   Sage, Thousand Oaks CA. https://www.john-fox.ca/Companion/.

Goldberg, A. E. (2006). *Constructions at work: The nature of generalization in

   language*. Oxford University Press.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005).

   Bilingualism affects picture naming but not picture classification. *Memory &

   cognition*, 33(7), 1220-1234.

Green, P., & MacLeod, C. J. (2016). SIMR: An R package for power analysis of

   generalized linear mixed models by simulation. *Methods in Ecology and

   Evolution*, 7(4), 493-498.

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

Howarth, P. (1998). Phraseology and second language proficiency. *Applied linguistics*, 19(1), 24-44.

Ishida, T. (2024). Are multiword frequency effects stronger in non-native than in native speakers?. *Bilingualism: Language and Cognition*, 27(3), 295-305.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & psychophysics*, 40(6), 431-439.

Jiang, S., & Siyanova–Chanturia, A. (2023). The processing of multiword expressions in L1 and L2 Chinese: Evidence from reaction times and eye movements. *The Modern Language Journal*, 107(2), 565-605.

Kaakinen, J. K., & Hyönä, J. (2005). Perspective effects on expository text comprehension: Evidence from think-aloud protocols, eyetracking, and recall. *Discourse processes*, 40(3), 239-257.

Kaakinen, J. K., Hyönä, J., & Keenan, J. M. (2002). Perspective Effects on Online Text Processing. *Discourse Processes*, 33(2), 159–173. https://doi.org/10.1207/S15326950DP3302_03

Kuiper, K., Van Egmond, M. E., Kempen, G., & Sprenger, S. (2007). Slipping on superlemmas: Multi-word lexical items in speech production. *The Mental Lexicon*, 2(3), 313-357.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology: Human Perception and*

Performance, 35, 876–895. https://doi.org/10.1037/a0013484

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, 82, 1-26.

Lemhöfer, K., Dijkstra, T., Schriefers, H., Baayen, R. H., Grainger, J., & Zwitserlood, P. (2008). Native language influences on word recognition in a second language: a megastudy. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 12.

Lenth, R., (2018). Package 'lsmeans'. The American Statistician, 34(4), 216-221.

Li, H., Paterson, K. B., Warrington, K. L., & Wang, X. (2022). Insights into the processing of collocations during L2 English reading: Evidence from eye movements. *Frontiers in Psychology*, 13, 845590. https://doi.org/10.3389/fpsyg.2022.845590

Li, H., Warrington, K. L., Pagán, A., Paterson, K. B., & Wang, X. (2021). Independent effects of collocation strength and contextual predictability on eye movements in reading. *Language, Cognition and Neuroscience*, 36(8), 1001–1009. https://doi.org/10.1080/23273798.2021.1922726

Li, W., Bassetti, B., & Frisson, S. (2025). Changes in word order do not eliminate the collocation advantage: An eye-tracking study of L1 and L2 speakers. *Bilingualism: Language and Cognition*, 1-15.

McClelland, J. L., Botvinick, M. M., Noelle, D. C., Plaut, D. C., Rogers, T. T., Seidenberg, M. S., & Smith, L. B. (2010). Letting structure emerge: connectionist and dynamical systems approaches to cognition. *Trends in cognitive sciences*, 14(8), 348-356.

Miellet, S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French: An evaluation of the E-Z Reader model.

Psychonomic Bulletin & Review, 14, 762–769. doi: 10.3758/BF03196834

Miellet, S., Sparrow, L., & Sereno, S. C. (2007). Word frequency and predictability effects in reading French: An evaluation of the EZ Reader model. *Psychonomic bulletin & review*, 14(4), 762-769.

Miller, L. M., Cohen, J. A., & Wingfield, A. (2006). Contextual knowledge reduces demands on working memory during reading. *Memory & Cognition*, 34, 1355–1367. doi:10.3758/BF03193277

Mor, B., & Prior, A. (2022). Frequency and predictability effects in first and second language of different script bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(9), 1363–1383. https://doi.org/10.1037/xlm0000927

Nesselhauf, N. (2003). The use of collocations by advanced learners of English and some implications for teaching. *Applied linguistics*, 24(2), 223-242.

Öksüz, D., Brezina, V., & Rebuschat, P. (2021). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*, 71(1), 55-98.

Öksüz, D., Brezina, V., Monaghan, P., & Rebuschat, P. (2024). Individual word and phrase frequency effects in collocational processing: Evidence from typologically different languages, English and Turkish. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 50(8), 1287.

Pinker, S. (1999). *Words and rules: The ingredients of language*. Basic Books.

Pinker, S., & Ullman, M. T. (2002). The past and future of the past tense. *Trends in cognitive sciences*, *6*(11), 456-463.

Powell, R. (2009). Contingency and convergence in macroevolution: a reply to John Beatty. *The Journal of Philosophy*, 106(7), 390-403.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological bulletin*, 124(3), 372.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. Memory & cognition, 14(3), 191-201.

Rebuschat, P., & Williams, J. N. (Eds.). (2012). *Statistical learning and language acquisition* (Vol. 1). Walter de Gruyter.

Reichle, E. D., & Schotter, E. R. (2020). A computational analysis of the constraints on parallel word identification. *In Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 42).

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The EZ Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and brain sciences*, 26(4), 445-476.

Reichle, E. D., Tokowicz, N., Liu, Y., & Perfetti, C. A. (2011). Testing an assumption of the E-Z Reader model of eye-movement control during reading: Using event-related potentials to examine the familiarity check. *Psychophysiology*, 48(7), 993-1003.

Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1(6), 906-914.

Saffran, J. R., & Kirkham, N. Z. (2018). Infant statistical learning. *Annual review of psychology*, 69(1), 181-203.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002, February). *Multiword expressions: A pain in the neck for NLP. In International conference on intelligent text processing and computational linguistics* (pp. 1-15). Berlin, Heidelberg: Springer Berlin Heidelberg.

Sandra, D., Frisson, S., & Daems, F. (1999). Why simple verb forms can be so difficult to spell: The influence of homophone frequency and distance in Dutch. *Brain and language*, 68(1-2), 277-283.

Sereno, S. C., & Rayner, K. (2003). Measuring word recognition in reading: eye movements and event-related potentials. Trends in cognitive sciences, 7(11), 489-493.

Schilling, H. E., Rayner, K., & Chumbley, J. I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory & cognition*, 26(6), 1270-1281.

Shao, Z., Van Paridon, J., Poletiek, F., & Meyer, A. S. (2019). Effects of phrase and word frequencies in noun phrase production. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 45(1), 147.

Shi, J., Peng, G., & Li, D. (2023). Figurativeness matters in the second language processing of collocations: Evidence from a self-paced reading experiment. *Language Learning*, 73(1), 47-83.

Siyanova, A., & Schmitt, N. (2008). L2 learner production and processing of collocation: A multi-study perspective. *Canadian modern language review*, 64(3), 429-458.

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.

Sprenger, S. A., Levelt, W. J., & Kempen, G. (2006). Lexical access during the production of idiomatic phrases. *Journal of memory and language*, 54(2), 161-184.

Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of

individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21(4), 360–407. https://doi.org/10.1598/RRQ.21.4.1

Sui L, Woumans E, Duyck W, Dirix N. The word frequency effect in first- and second-language reading by Chinese and Dutch bilinguals. *Bilingualism: Language and Cognition*. 2025;28(1):232-247. doi:10.1017/S136672892400035X

Sui, L., Dirix, N., Woumans, E., & Duyck, W. (2023). GECO-CN: Ghent Eye-tracking COrpus of sentence reading for Chinese-English bilinguals. *Behavior Research Methods*, 55(6), 2743-2763.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Harvard University Press.

Tomasello, M. (2009). *The usage-based theory of language acquisition*. In The Cambridge handbook of child language (pp. 69-87). Cambridge Univ. Press.

Veldre, A., & Andrews, S. (2014). Lexical quality and eye movements: Individual differences in the perceptual span of skilled adult readers. *Quarterly Journal of Experimental Psychology*, 67(4), 703-727.

Vilkaitė, L. (2016). Are nonadjacent collocations processed faster?. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(10), 1632.

Vilkaitė, L., & Schmitt, N. (2019). Reading collocations in an L2: Do collocation processing benefits extend to non-adjacent collocations?. *Applied Linguistics*, 40(2), 329-354.

Vilkaitė-Lozdienė, L. (2022). Do different morphological forms of collocations show comparable processing facilitation? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 48(9), 1328–1347. https://doi.org/10.1037/xlm0001130

Vilkaitė-Lozdienė, L., & Conklin, K. (2021). Word order effect in collocation processing. *The Mental Lexicon*, 16(2-3), 362-396.

Wang, T., & Saffran, J. R. (2014). Statistical learning of a tonal language: The influence of bilingualism and previous linguistic experience. *Frontiers in psychology*, 5, 953.

Wickham, H. (2016). Data analysis. In ggplot2: elegant graphics for data analysis (pp. 189-201). Cham: Springer international publishing.

Wolter, B., & Gyllstad, H. (2011). Collocational links in the L2 mental lexicon and the influence of L1 intralexical knowledge. *Applied Linguistics,* 32(4), 430-449.

Wolter, B., & Gyllstad, H. (2013). Frequency of input and L2 collocational processing: A comparison of congruent and incongruent collocations. *Studies in Second Language Acquisition*, 35(3), 451-482.

Wolter, B., & Yamashita, J. (2018). Word frequency, collocational frequency, L1 congruency, and proficiency in L2 collocational processing: What accounts for L2 performance?. *Studies in second language acquisition*, 40(2), 395-416.

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

Zang, C. (2019). New Perspectives on Serialism and Parallelism in Oculomotor Control During Reading: The Multi-Constituent Unit Hypothesis. *Vision*, 3(4), 50. https://doi.org/10.3390/vision3040050

Zang, C., Fu, Y., Bai, X., Yan, G., & Liversedge, S. P. (2021). Foveal and parafoveal processing of Chinese three-character idioms in reading. *Journal of Memory and Language*, 119, 104243.

Zang, C., Wang, S., Bai, X., Yan, G., & Liversedge, S. P. (2024). Parafoveal

processing of Chinese four-character idioms and phrases in reading: Evidence

for multi-constituent unit hypothesis. *Journal of Memory and Language*, 136.