Bilingualism: Language and Cognition

cambridge.org/bil

Are multiword frequency effects stronger in non-native than in native speakers?

Tomomi Ishida

Nihon Fukushi University, Aichi, Japan

Research Article

Cite this article: Ishida, T. (2024). Are multiword frequency effects stronger in non-native than in native speakers? *Bilingualism: Language and Cognition*, **27**, 295–305. https://doi.org/10.1017/S1366728923000548

Received: 28 December 2021
Revised: 9 June 2023
Accepted: 25 June 2023
First published online: 9 August 2023

Keywords:
Multiword frequency effects; Grammaticality judgment tasks; Vocabulary knowledge; Processing advantage

Corresponding author:
Tomomi Ishida, 35-6 Egemae, Okuda, Mihama-cho, Chita-gun, Aichi, 470-3233, Japan
Email: tomomii@n-fukushi.ac.jp

Abstract

This study investigated whether non-native English speakers showed a processing advantage for high-frequency multiword units (multiword frequency effects), and whether the effects differed between native and non-native speakers. Such a difference has been identified in relation to single-word processing. Native English speakers and intermediate learners of English with languages of different scripts (native speakers of Japanese and German) judged whether English multiword units were grammatical. A significant processing advantage was identified for both native and non-native participants. More importantly, the multiword frequency effects were stronger among non-native than native speakers. The discrepancy persisted even after including individual vocabulary knowledge as a predictor in the mixed-effect models. Furthermore, there was no significant different impact of the effects between two non-native groups, even though German participants responded quicker than Japanese participants. This indicates that the varying influence between L1 and L2 could be explained by within-language, not between-language, variables.

Introduction

It is clear that word frequency information has a considerable influence on first language (L1) lexical processing: people respond faster to high-frequency than low-frequency words (Inhoff & Rayner, 1986; Rayner, 1998, 2009; Rayner & Duffy, 1986). Known as word frequency effects, such effects have also been recognized in second language (L2) processing. Stronger word frequency effects among L2 speakers in comparison with L1 speakers have been widely reported in related research (in the context of an eye-tracking paradigm: Cop et al., 2015; Whitford & Titone, 2012; in picture-naming tasks: Gollan et al., 2008; in lexical decision tasks: Duyck et al., 2008); that is, the discrepancy in reaction latency between high- and low-frequency words was greater in the L2 group than in the L1 population. Diependaele et al. (2013) proposed two accounts to explain this discrepancy: "language competition" and "lexical entrenchment." The former proposes that lexical processing is slower among bilinguals because of language competition, while the latter proposes that the difference in word frequency is due to different levels of lexical exposure to the target language, that is, lexical exposure to language could be responsible for the different impacts of word frequency.

While word-level processing research has identified differential magnitudes in word frequency effects, only a few studies have addressed this issue in terms of multiword frequency effects. This is because most L2 scholars working on the mechanics of formulaic sequences have focused on the issue of whether high-frequency multiword sequences are processed and retrieved as whole units, as in L1 processing. Holistic storage, that is, whether frequent multiword strings are stored as a single word in memory, could create a processing advantage that leads to effortless and cost-efficient recognition and production of language usage (Kim & Kim, 2012). It is an unspoken assumption that, if the holistic hypothesis is supported, learning formulaic sequences would be considerably helpful in acquiring native-like proficiency in L2 learning. While L1 studies have yielded consistent results indicating that multiword sequences are stored as chunks and recognized holistically rather than word-by-word in language use, previous L2 studies have shown mixed results.

In addition to the conflicting findings, previous studies have hardly examined the extent of multiword frequency effects, as noted. Investigating their magnitude could provide valuable insights into the sensitivity of multiword frequency and shed more light on the relationship between exposure and multiword processing. Therefore, the purpose of this study is twofold: first, to verify whether L2 speakers provide clear evidence of processing advantage in multiword sequences as L1 speakers do; second, assuming this is the case, to examine whether a different degree of multiword frequency effects emerges among L1 and L2 groups and between two L2 groups. This will enable an investigation of whether exposure to the target language or language proficiency could affect the magnitude of the multiword frequency effects.

© The Author(s), 2023. Published by Cambridge University Press

Literature review

Processing advantage for multiword units in L1 and L2 visual settings

There has been an ongoing interest in the processing advantage associated with frequent multiword phrases or formulaic sequences (frequently occurring sets of words in natural discourse such as "as a result") in the linguistic study of phraseology. High-frequency multiword phrases could be described in over 40 different terms (Wray & Perkins, 2000), and Altenberg (1998) estimated that over 80% of words in a spoken corpus consisted of parts of recurrent multiword phrases.

There is substantial literature in L1 phraseology research demonstrating a frequency processing advantage; that is, frequently recurring multiword expressions were recognized faster and more accurately than less frequent ones (for compound words: Gibbs & Gonzales, 1985; for idioms: Kuperman et al., 2008). It is assumed that these formulaic sequences are stored in memory as one unit and retrieved as one unit at the time of use (Wray & Perkins, 2000). The holistic hypothesis has been supported not only by experimental studies of idioms, which have figurative and connotative meanings and function as complete units, but also by studies that dealt with lexical bundles, that is, frequently recurring clusters of words (including some strings of words) that perform as incomplete units (e.g., "in the middle of the"). For example, Tremblay et al. (2011) investigated a processing advantage for lexical bundles by conducting three self-paced reading experiments and two recall experiments; in which four- and five-word lexical bundles were read more quickly and recalled more efficiently than equivalent non-lexical bundles. Providing support for usage-based models, they interpreted the processing advantage as implying that the processing of frequent multiword phrases demanded less short-term memory storage. It seems to be a clear fact that the multiword frequency effects play a significant role in L1 processing, indicating that frequently recurring word sequences are firmly linked and stored in memory. To the best of the researcher's knowledge, no behavioral studies to date have provided evidence contradicting the multiword frequency effects relating to L1 processing advantages for multiword sequences.

Despite this consistency in L1 research, evidence in previous L2 studies has been somewhat inconclusive. Many studies have shown that high-frequency multiword sequences have a processing advantage (e.g., Conklin & Schmitt, 2008; Hernández et al., 2016; Jiang & Nekrasova, 2007; Kim & Kim, 2012; Siyanova-Chanturia et al., 2011b). In Jiang and Nekrasova (2007), for instance, English native speakers and L2 speakers with different mother tongues completed grammaticality judgment tasks, in which they were required to decide whether multiword expressions were grammatically correct. The researchers selected 26 formulaic sequences that occurred frequently and performed as coherent and complete chunks in sentences from previous corpus-based studies. Reaction latency for formulaic sequences, such as "to sum up," was compared to reaction times for non-formulaic sequence such as "to climb up." Their findings showed that formulaic sequences were recognized more quickly and accurately than non-formulaic sequences, suggesting that frequently occurring clusters were stored and memorized as single lexicalized units, supporting the holistic processing of formulaic sequences. Hernández et al. (2016) compared high-, mid-, and low-frequency multiword units to examine people's sensitivity to the distributional properties of multiword phrases. Their findings, based on a phrasal-decision task, revealed a processing advantage even when L2 groups responded to low-frequency expressions (e.g., "We have to wait" vs. "We have to leave") and confirmed that L2 participants were sensitive to the distributional properties of frequency information to the same extent as L1 participants.

Conversely, some L2 studies have observed no significant processing advantages in high-frequency multiword sequences (e.g., Schmitt et al., 2004; Schmitt & Underwood, 2004; Siyanova-Chanturia et al., 2011a). Siyanova-Chanturia et al. (2011a) employed an eye movement paradigm in which L1 and L2 participants were required to read short passages containing an idiom with a figurative meaning ("at the end of the day" as in "eventually"), idioms with a literal meaning ("at the end of the day" as in "in the evening"), or novel phrases ("at the end of the war"). Two predictions were tested: (1) the L1 speaker group should show a processing advantage for idioms, and (2) L2 speakers should read figurative idioms more quickly than literal idioms. In their experiment, the L1 speakers showed shorter eye fixations on idioms and no significant differences in processing between figurative and literal idioms, while L2 participants showed no processing advantage for idioms, and figurative idioms were read slower than literal idioms, even though the L2 participants knew their meanings.

Therefore, previous studies focusing on how L2 formulaic sequences are stored and structured in the mental lexicon have yielded mixed results, which may be partially due to the experimental items. Among the three studies suggesting a null effect of processing advantage in multiword phrases, two (Schmitt & Underwood, 2004; Siyanova-Chanturia et al., 2011a) used figurative idioms. Although some of the figurative idioms they tested occur frequently and function as unitary chunks in language use, the acquisition of figurative idioms is not as simple as that of literal multiword phrases, since learners are required to understand connotative as well as literal meanings (Cooper, 1999). Even advanced learners tend to avoid using idioms (Irujo, 1993). Learners first link expressions such as "at the end of the day" with literal meanings like "in the evening" in their mental lexicon. To assimilate the figurative meaning, which is not predictable from the literal meaning of each word, they need to reconstruct the form–meaning link. Hence, it is assumed that the link between the idiom and its figurative meaning is weaker than the link between the idiom and its literal meaning. Even when learners comprehend the figurative meanings of idioms, it may be unlikely that a strong semantic link is constructed between some idiomatic expressions and their figurative meanings.

In addition to Schmitt and Underwood (2004) and Siyanova-Chanturia et al. (2011a), Schmitt et al. (2004) showed no processing advantage of frequent multiword sequences using corpus-based recurrent clusters. However, as the researchers in the study admitted, the recurrence of clusters in a corpus was not an appropriate test item to investigate the processing of formulaic sequences, due to the inclusion of phrases that functioned as incomplete units.

In summary, the L2 processing advantage of frequently used multiword expressions has been supported when the test materials were multiword sequences that were used literally, functioning as complete and coherent structural units, while existing experiments using connotative figurative idioms have produced conflicting results.

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

Different frequency effects on L1 and L2 recognition

It has been widely reported that L1 and L2 visual word processing differ. The existing literature has demonstrated the difference in lexical processing between L1 and L2 within the same participants or between L1 and L2 speakers, as well as between adult and child bilinguals (Brysbaert et al., 2017; Cop et al., 2015; Duyck et al., 2008; Lemhöfer et al., 2008; Schröter & Schroeder, 2018; Tiffin-Richards & Schroeder, 2015; Whitford & Titone, 2012).

One of the significant findings of these studies was that the word frequency effects were greater for L2 than L1. For example, Diependaele et al. (2013) investigated this difference by reanalyzing Lemhöfer et al.'s (2008) data from a word identification task, comparing L1 English speakers and three different L2 groups (Dutch, French, and German). The test material contained 1,025 English words. In the masked word identification task, participants were asked to press a button as they recognized a word that gradually appeared on the computer screen. The findings showed significantly stronger word frequency effects for the three different L2 groups. The researchers proposed two accounts to explain these different magnitudes: "language competition" and "lexical entrenchment." The language competition account postulated that lexical processing is slower among bilinguals than monolinguals because of competition between L1 and L2 in bilinguals' memories; they need to deal with two lexicons and compare similar or overlapping forms across languages, when a word is orthographically or phonologically similar in the mother tongue and target language. For instance, when German–English bilinguals recognized a word such as "June" in English, they needed to distinguish the word from seven cross-linguistic orthographic neighbors like "jung," as well as some English neighbors like "tune." Furthermore, Perry et al., (2007) claimed that recognizing low-frequency words with high-frequency neighbors would require even more processing time. Hence, the language competition hypothesis is based on the bilingual processing mechanism; no matter how much an L2 speaker is exposed to the target language, the stronger word frequency effects will persist. Moreover, within such an account, Duyck et al. (2008) assumes larger L2 frequency effects in bilinguals with an L1 with the same script (e.g., Dutch–English bilinguals) but not in bilinguals whose L1 has a different script (e.g., Chinese–English bilinguals) because of "shared bins" for same-script languages (but also see Mor & Prior, 2020, 2022).

In contrast, the lexical entrenchment hypothesis proposes that the difference in word frequency effects can be explained by language exposure. This is based on the lexical representation of use-based theory, in which the amount of language exposure is a critical factor in lexical processing. From this perspective, Cop et al. (2015) argued that lower levels of L2 exposure mean that subjective word frequency for L2 speakers differs from objective frequency defined by, for example, corpus-based word frequency. Therefore, their lack of exposure to low-frequency words contributes to the increased word frequency effects among L2 speakers. A gap between subjective and objective word frequency could produce a discrepancy, especially in the case of low-frequency words. According to the lexical entrenchment account, with increasing language exposure or native-like proficiency, the discrepancy in the word frequency effects would become very subtle or disappear altogether: there should be no significant difference between L1 and L2 speakers in lexical processing with similar language proficiency. To verify the influence of exposure to the target language, Diependaele et al. (2013) included vocabulary size as a predictor in their statistical analysis and found that the significant interaction of word frequency effects disappeared. They argued that the discrepancy in word frequency effects between L1 speakers and three L2 groups could be accounted by language proficiency and concluded that there was no qualitatively different mechanism between L1 and L2 lexical processing, supporting the lexical entrenchment perspective. Whitford and Titone (2012) and Cop et al. (2015), on the other hand, found that objective L2 proficiency or vocabulary knowledge had no significant impact on L2 speakers' greater word frequency effects compared to L1 speakers. In summary, even though the evidence for stronger word frequency effects among L2 speakers is conclusive, no consensus has been reached about its mechanism.

With respect to multiword linguistic constructions, there have been few studies of different frequency effect magnitudes (e.g., Hernández et al., 2016; Jiang & Nekrasova, 2007). Hernández et al. (2016) investigated decision-making in phrasal judgment tasks by comparing the multiword frequency effects of one L1 group and two L2 groups (the immersion exposure group and the classroom exposure group). The L2 groups revealed a similar pattern to L1 participants in terms of sensitivity to the distributional properties of multiword phrases. Contrary to single word studies, larger multiword frequency effects for L2 speakers than for the L1 group were not observed in this study. They failed to observe a significantly different response latency caused by exposure to the target language. They pointed out that this may have been due to the frequency range of the experimental items. Their phrasal judgment tasks employed high-, mid-, and low-frequency four-word compositional items whose mean frequency difference was much smaller than the word frequency difference in previous studies. Therefore, the differences in the multiword frequency effects between L1 and L2 speakers require further research.

Present study

With the limitations and conflicting findings of previous studies in mind, the first goal of this study is to address the issue of processing advantage when processing recurring multiword units. The second goal is to determine whether the relative sizes of multiword frequency effects differ among one L1 and two L2 groups with different scripts.

Rather than relying on phrasal judgment tasks, the present research employs the grammaticality judgment task used by Jiang and Nekrasova (2007), because L2 participants might not fully understand the concept of "phrases." The current research expands on the existing literature in three ways. First, as the existing literature employed a much smaller number of experimental items than studies investigating single word frequency effects, the present study increases the number of experimental multiword units. Second, the present study recruited L2 participants with lower proficiency than those in previous studies, so that there would be a greater proficiency difference between the L1 and L2 participants. For instance, the L2 group in Kim and Kim (2012) consisted of highly proficient L2 English speakers who obtained B2 or C1 in the Common European Framework of Reference for Languages (CEFR). The participants in the present study were lower-intermediate learners of English. Third, there were two L2 groups in this study, differentiated according to the orthographies and language structures of their mother tongue: the first L2 group comprised Japanese non-native speakers of English (JNNS) with a non-alphabetical L1 script and different

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

writing system, while the second group consisted of German non-native speakers of English (GNNS) with an alphabetical L1 and similar canonical word order, so as to observe language competition.

The hypothesis in this study is that multiword frequency effects for both the L1 and L2 would be observed, demonstrating the processing advantage of very frequent multiword expressions, in terms of quick and accurate visual recognition, since literally coherent and complete structural multiword units are used in this experiment. More specifically, the study also addresses whether the magnitude of these effects differs between L1 and L2 groups. If greater multiword frequency effects emerge among L2 participants, the results will provide two possible explanations relating to the aforementioned accounts of the mechanism behind single word frequency effects. If the language competition account explains multiword frequency effects, then German–English participants should show stronger multiword frequency effects than the Japanese–English group, because they need to distinguish the target language from an orthographically and structurally similar L1. This logic follows that of Duyck et al. (2008), who predicted that L2 groups with different scripts would show null or little effects due to no explicit orthographical language conflicts in participants' memories. If this assumption can expand beyond the L2 single-word level to multiword processing, the GNNS would demonstrate larger relative multiword frequency effects, compared to the JNNS, because their L1's impact via German language structures and the shared Roman alphabet may lead to language competition. In particular, if incongruent multiword sequences in the L2 that do not share the same structural unit as the L1 are confusing to distinguish, they may distract from processing, causing cognitive load.

By contrast, the lexical entrenchment account in word frequency effects assumes that all L2 populations with less exposure to the target language will show greater frequency effects. Based on this, the prediction is that Japanese and German groups would show stronger multiword frequency effects, compared to the L1 group, depending on the amount of their exposure to the L2.

Methods

Participants

The participants in the present study comprised one group of 30 native English speakers (NS) and two groups of non-native speakers (NNS; 30 Japanese–English and 28 German–English), who were all residing in Japan at the time of the experiment. Data for one English participant were removed due to procedural failure, and one participant was replaced for the experiment. The 30 L1 participants were mainly undergraduate or postgraduate students (22 males, 8 females) who came to Japan as exchange students. Participant characteristics are presented in Table 1. JNNS had studied English for more than six years in Japanese formal education. They were either undergraduate or postgraduate students (18 males and 12 females). GNNS were mostly international students in Japan (16 males and 12 females); 26 GNNS had spent less than three months in Japan, and only one GNNS had stayed in Japan for more than one year[1]. Participants' English vocabulary knowledge was estimated by LexTALE (Lemhöfer & Broersma, 2012), with average scores of 93.90 (SD 6.15), 71.08 (SD 7.93), and 80.64 (SD 10.79) for NS, JNNS, and GNNS, respectively. The NS group had more extensive vocabulary knowledge over JNNS (t (58) =12.27, $p < .001$) and GNNS (t (56) = 5.80, $p < .001$). There was a significant difference between JNNS and GNNS (t (56) = 3.73, $p < .001$). The study met the requirements and gained the approval of the Ethics Committee of Nihon Fukushi University, in Japan, concerning empirical studies with human participants.

Table 1. Characteristics of three groups (standard deviation)

	NS (N = 30)	JNNS (N = 30)	GNNS (N = 28)
Age	26.03 (8.85)	25.28 (7.29)	25.38 (5.43)
LexTALE	93.90 (6.15)	71.08 (7.93)	80.64 (10.79)
Age started English learning		11.07 (2.71)	9.11 (2.06)
Self-reported proficiency			
Speaking		4.72 (2.09)	7.61 (1.57)
Listening		5.66 (1.95)	8.32 (1.12)
Reading		6.45 (1.61)	8.68 (.82)
Writing		5.34 (1.90)	7.36 (1.39)

Note. Self-report proficiency on a 10-point scale; NS = native speakers; JNNS = Japanese non-native speakers of English; GNNS = German non-native speakers of English.

Materials and design

There were 46 high-frequency multiword linguistic constructions, 46 low-frequency multiword linguistic constructions, and 46 ungrammatical sequences. The multiword recurring sequences used in the test were regarded as a coherent unit in a sentence and did not include incomplete sequences such as "I don't think you." Idiomatic expressions were also excluded from the list of test items. This is because, when the author asked 20 JNNS not involved in the experiment to translate the figurative and literal meanings of 20 idioms used by Siyanova-Chanturia et al. (2011a) to test the comprehensibility of idiomatic expressions, the survey revealed that, apart from idioms that were identical between their mother tongue (Japanese) and the target language (English), the percentage of L2 participants who correctly identified figurative meanings was considerably low.

The high-frequency word sequences (three or four words) were taken from previous studies (e.g., Biber, 2009; Jiang & Nekrasova, 2007; Schmitt et al., 2004). The low-frequency multiword linguistic constructions were grammatical expressions formed by replacing one word of the high-frequency multiword sequences with another word that had a similar frequency and word length. For instance, "time" in "for the first time" was substituted by "half" to make "for the first half." Based on the British National Corpus (BNC Consortium, 2001), there were no significant differences in the substituted words (e.g., "time" vs. "half") in terms of lexical frequency (4.75 and 4.55 for mean logarithmic frequency of different words, $t(89.31) = 1.21$, $p > .05$) or word length (4.91 and 4.83 for mean word length of different words, $t(76.62) = .27$, $p > .05$). The mean high-multiword frequency was 23.97 per million (SD 21.89), and the mean low-multiword frequency was .93 per million (SD 1.51), according to the BNC and the Corpus of Contemporary American English (COCA; Davies, 2008). In addition, 46 ungrammatical phrases were

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

created for the study, and their ungrammaticality was reviewed by the native English speakers who participated.

Two initial counterbalanced lists were developed for 138 experimental items. Each list consisted of 23 high-frequency successive multiword sequences, 23 low-frequency successive multiword sequences, and 46 ungrammatical sequences. Each participant saw one of the two sets so that each multiword sequence occurred only once per participant. To avoid a priming effect through reading a similar sequence, the list contained only one of the paired multiword successive expressions. For instance, "for the first time" and "for the first half" were presented in different sets. In addition to the main test items, 46 filler sequences were used. Each participant judged grammaticality for 23 high-frequency multiword sequences, 23 low-frequency multiword sequences, 46 ungrammatical sequences, and 46 grammatical filler expressions.

Procedure

The participants took the test individually, in a quiet room. Participants read an information sheet and provided written informed consent prior to the experiment. They were asked to fill out a questionnaire about their language skills and educational background. Then, they were told to make a grammatical-nongrammatical discrimination for each multiword sequence on the screen by pressing a button ("YES" or "NO") as quickly and accurately as possible. The participants' reaction times and errors were recorded using DMDX software (Forster & Forster, 2003). The order of stimulus presentation was randomized. Words appeared in lowercase. Twenty practice trials were conducted prior to the main experimental session[2]. At the onset of each trial, a fixation point (asterisk) appeared at the center of the screen for 600ms. No feedback was provided during any trial. After the grammaticality judgment task, all participants took the LexTALE test to check their vocabulary knowledge.

Data analysis

All incorrect responses were excluded from the latency analysis. Grammatical decision latencies of less than 400ms or greater than 4000ms[3], and any responses that were 2.5 standard deviations away from the individual participant mean, were excluded. This involved 3.52%, 2.74%, and 3.39% of the data for the NS, JNNS, and GNNS groups, respectively.

To analyze the correct responses, a linear mixed-effect model (LME) with cross random-effect factors was conducted using R version 4.1.1 (R Development Core Team, 2013) and the R package *lme4* (Baayen et al., 2008). The models included predictors, random slopes of the predictors, and random intercepts associated with participants and items. All the remaining reaction times were log-adjusted. The mixed-effect models contained fixed effects for group (NS, JNNS, and GNNS), logarithmic multiword frequency (continuous), logarithmic frequency of differing words (e.g., "time" and "half" in "for the first ____"), and list. Group was a categorical factor, and NS was set as the reference. List was a categorical factor referring to two different counterbalanced lists displayed on the screen (versions A and B). All continuous variables were centered. The Pearson's correlations of the continuous predictor variables were low *(all* $r < .11$*)*. A plot was created using the *effects* package.

The best model was chosen by backward selection, including the interaction of group with multiword frequency and the main effects of group and multiword frequency, since the main focus of the study was to investigate the discrepancy of the multiword frequency effects among the three groups. The selection procedure of the best model took the following steps. First, I developed a converged model with maximal fixed and random structures, including these variables and their interactions. Then, several converged models were created by simplifying fixed structures or removing random slopes one by one. Finally, the best model was chosen with the lowest value according to the Akaike information criterion (AIC) using the anova command to compare the models. The relevant supplementary materials (study materials, predictors' characteristics, R scripts) are available from the Open Science Framework (OSF; https://osf.io/tfs3e/).

Regarding the analyses of error proportions, a generalized mixed-effects model (GLMM) with cross random-effect factors was created for binomial data. Contrast coding was used for data accuracy (-.5 = incorrect, .5 = correct). The best model was chosen as the analyses of reaction times were performed. The optimizer (BOBYQA) was adopted to avoid convergence failure.

To examine the processing of multiword frequency effects, four separate analyses were performed: (1) the first model assessed the processing advantage of the frequently occurring multiword units for each group; (2) the second model included the main effects of group and multiword frequency, and their interactions, to explore the different influences of multiword frequency effects between NS and NNS; (3) the third model was designed to address the role of individual vocabulary knowledge by adding LexTALE scores as fixed effects to the second model; (4) the fourth model aimed to observe the differences of the multiword frequency effects between two L2 participant groups who do not typologically and orthographically share language structures and scripts.

Results

Multiword frequency effects in L1 and L2

Table 2 summarizes the mean grammaticality decision latencies and error rates from the three participant groups. For the first LME model to clarify the multiword frequency effects and examine sensitivity to the distributional properties of multiword phrases for each group, separate analyses, including multiword frequency and word frequency as fixed effects, were administered. As predicted, for NS, a main effect of multiword frequency was significant in reaction time (estimate = -.04, $SE = .00$ $t = -9.14$, $p < .001$). There was a main effect of the multiword frequency effects in the reaction time for the two NNS groups as well (JNNS: estimate = -.05, $SE = .01$, $t = -8.23$, $p < .001$, GNNS: estimate = -.05, $SE = .01$ $t = -8.70$, $p < .001$). The findings indicated that high-frequency multiword sequences have a processing advantage over low-frequency multiword expressions. The error rate analyses provided the resemble results, showing a main effect of multiword frequency for both L1 and two L2 groups (NS: estimate = 1.17, $SE = .22$, $z = 5.38$, $p < .001$, JNNS: estimate = .83, $SE = .16$, $z = 5.09$, $p < .001$, GNNS: estimate = .66, $SE = .22$, $z = 3.06$, $p < .001$).

Of primary importance for the current study was the relative size of performance in processing multiword sequences among the three groups. The second model was fitted by including multiword frequency, word frequency and all three groups. As Table 3 illustrates, the effects of multiword frequency were significant ($t = -7.70$, $p < .001$). The NS group responded faster than the two

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

Table 2. Mean reaction times (in milliseconds) and error rate (%) of each group with standard deviation in parentheses

	NS RT	NS Error	JNNS RT	JNNS Error	GNNS RT	GNNS Error
High-frequency multiword expressions	894 (320)	1.03	1246 (458)	2.75	1018 (397)	2.64
Low-frequency multiword expressions	1065 (420)	9.57	1520 (565)	12.46	1262 (514)	9.64
Ungrammatical expressions	1300 (528)	10.80	1780 (698)	18.33	1466 (310)	12.34

Note. RT = reaction times; NS = native speakers; JNNS = Japanese non-native speakers of English; GNNS = German non-native speakers of English.

NNS groups (JNNS: $t = 6.24$, $p < .001$; GNNS: $t = 2.48$, $p < .05$). More importantly, the magnitude of the multiword frequency effects differed between NS and JNNS ($t = -2.28$, $p < .05$), and NS and GNNS ($t = -2.09$, $p < .05$). Figure 1 presents the sensitivity to multiword frequency for each participant group. The two NNS groups showed greater multiword frequency effects than the NS group.

With respect to error rate analysis, the two NNS groups made significantly more errors than NS (JNNS: $z = -3.04$, $p < .001$; GNNS: $z = -2.32$, $p < .05$). Like the findings of the reaction times, the logarithmic multiword frequency interacted with group. The frequency difference of multiword sequences tends to affect more for two NNS groups (JNNS: $z = -2.72$, $p < .05$; GNNS: $z = -3.28$, $p < .001$).

The influence of individual vocabulary knowledge differences

To investigate whether the interaction of the multiword frequency with group could be modulated by the individual different variables across all three groups, an additional analysis was conducted, taking vocabulary knowledge into account. As Table 4 shows, vocabulary knowledge did not eliminate the interaction between group (NS vs. JNNS) and multiword frequency ($t = -1.99$, $p < .05$) and between group (NS vs. GNNS) and multiword frequency ($t = -2.34$, $p < .05$), indicating that the different impact between L1 and the two L2 groups persisted even after individual vocabulary knowledge was included in the statistical analysis. Note that there was no interaction between vocabulary knowledge and multiword frequency; vocabulary knowledge did not have a main effect, suggesting that increased vocabulary knowledge did not lead to smaller multiword frequency effects. This leads to a possible argument that individual vocabulary knowledge has less influence on multiword frequency effects than that on single-word processing.

The model for error rate found the main effect of multiword frequency, indicating that lower multiword frequency expressions predicted more errors than higher ones. An interaction between multiword frequency and group NS vs. GNNS was found ($z = -2.23$, $p < .05$), but there was no interaction of multiword frequency and group NS vs. JNNS ($z = -.90$, $p = .37$) after adding the individual difference variables. Moreover, the vocabulary knowledge did not interact with multiword frequency in error analysis. This implies that vocabulary knowledge has limited influence on processing multiword expressions.

L2 group difference with different script

To further explore the different degrees of multiword frequency effects and clarify whether the two NNS groups differed in the relative size of the multiword frequency effects, a separate analysis containing only two L2 groups' data with the multiword frequency, word frequency, and vocabulary knowledge as predictors, was performed. Table 5 summarizes the main effects and interactions in reaction latency for JNNS and GNNS. Reaction times were significantly shorter for GNNS than for JNNS ($t = 2.29$, $p < .05$). However, the interaction of logarithmic multiword frequency with group was not significant ($t = -.28$, $p = .78$), meaning that sensitivity to multiword frequency does not differ between two NNS groups even though there was a significant discrepancy of vocabulary knowledge. One can predict that participants with larger vocabulary knowledge tend to have overall faster reaction times, but that increased vocabulary knowledge in L2 does not make the multiword frequency effects smaller. The accuracy analysis found the main effect of the multiword frequency to be only ($z = 4.05$, $p < .001$). All remaining predictors and interactions were not significant.

Discussion

In this study, two research questions were addressed regarding the processing of recurrent multiword strings among L1 and L2 groups. First, confirmation was needed of previous results showing that both L1 and L2 speakers demonstrated frequency effects for multiword units. The second aim was to observe differences in the multiword frequency effects by comparing reactions among one NS group and two NNS groups whose mother tongues differed typologically and orthographically in scripts and language structures.

Data from the grammaticality judgment task revealed that NS and NNS were sensitive to multiword frequency. The higher the frequency of multiword expressions, the faster and more accurately the participants responded. This finding is highly compatible with the existing research, which demonstrated L1 and L2 sensitivity to the distributional properties of multiword frequency (Hernández et al., 2016). Although the present study did not directly address the holistic processing of multiword expressions, the results indicated that individual words in high-frequency multiword units were strongly linked and stored in memory. These findings pose a challenge to some previous studies that did not show a processing advantage for high-frequency phrasal expressions. These conflicting results can be partially attributed to the differences in experimental items used (Jiang & Nekrasova, 2007). Siyanova-Chanturia et al. (2011a), for instance, found that non-native participants' processing of idioms and novel phrases were very similar, even though they took care to choose idioms that were well-known and familiar to their L2 group. The lack of processing advantage for idioms in their eye-tracking experiment can be attributed to the links between literal and figurative meanings and orthographic units. When learning an L2 idiom, NNS first learn its literal meaning through L1 translation and link the multiword expression to its L1 equivalent. In learning the figurative meaning, a new link is created between the idiom and its figurative meaning, which may require a cognitive load.

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

Table 3. Main effects and interactions for three participant groups

	Reaction Time				*Accuracy*			
Fixed effects	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*z*	*p*
(Intercept)	2.95	.02	148.25	<.001	4.37	.33	13.32	<.001
List	.03	.02	1.45	.15	−.40	.33	−1.22	.22
Log Multiword Frequency	−.04	.01	−7.70	<.001	1.22	.17	7.22	<.001
Group1	.06	.02	2.48	<.05	−.62	.27	−2.32	<.05
Group2	.15	.02	6.24	<.001	−.80	.26	−3.04	<.001
Log Multiword Frequency × Group1	−.01	.00	−2.09	<.05	−.50	.15	−3.28	<.001
Log Multiword Frequency × Group2	−.01	.00	−2.28	<.05	−.39	.14	−2.72	<.05
Control variable								
Log Word Frequency	.01	.01	.71	.48	−.22	.32	−.70	.48
Log Multiword Frequency× Log Word Frequency	.00	.01	.14	.89				
Group1 × Log Word Frequency	−.02	.01	−2.41	<.05				
Group2 × Log Word Frequency	.00	.01	.40	.69				
Log Multiword Frequency× Group1 × Log Word Frequency	.00	.01	.43	.67				
Log Multiword Frequency× Group2× Log Word Frequency	.03	.01	3.16	<.001				
Random effects	Variance	SD			Variance	SD		
Item (intercept)	.00	.04			1.19	1.09		
Log word frequency	.00	.03						
Subject (intercept)	.01	.09			.25	.50		

Note. Group1 = native English speakers vs. German non-native English speakers;
Group2 = native English speakers vs. Japanese non-native English speakers.
*RT Formula: log rt~list + log multiword frequency*group*log word frequency+(1|subject)+(1 + log word frequency|item); rt = reaction times;*
*Accuracy Formula: error~list + group*log multiword frequency + log word frequency+(1|subject)+(1|item)*

The findings in the present study provided affirmative evidence of greater multiword frequency effects for L2 groups, compared to the L1 group. These results are consistent with the previously identified single-word frequency effects, but they are dissimilar to previous research on processing multiword sequences. This may be due to the number of experimental items and the participants' language proficiency levels. For instance, Hernández et al. (2016) and Jiang and Nekrasova (2007), who failed to observe a discrepancy in multiword frequency effects, used a smaller number of multiword expressions, and participants in their studies were at a higher proficiency level compared to participants in the present research. This can be taken as an indication that multiword frequency effects' differences might be less salient than the differences in single-word frequency effects, which were identified by existing studies even when using only a small number of experimental words (Duyck et al., 2008). Schröter and Schroeder (2018) argued that difficulties recognizing L2 low-frequency words caused a greater influence of single-word frequency effects for NNS because of the limited exposure to low-frequency words. Taken together, recognizing and retrieving a single low-frequency word is likely to be more time-consuming than processing low-frequency multiword sequences composed of high-frequency individual single words.

Concerning similarities and discrepancies between single-word and multiword frequency effects, one possible account for multiword frequency effects appears to be consistent with that for single-word frequency effects. Diependaele et al. (2013) examined word frequency effects from the perspective of the lexical entrenchment explanation, positing that word frequency effects could be attributed to differences in exposure to the target language, and to the language competition account, hypothesizing that lexical competition in memory causes these effects. The findings in this study provide straightforward support for the lexical entrenchment account, as they show that language competition cannot explain the different multiword frequency effects between the L1 and two L2 groups. The language competition hypothesis predicts that NNS with the same alphabet orthographies would show stronger multiword frequency effects than NNS with different orthographies, due to the greater competition in the mental lexicon. However, the lexical entrenchment account, based on the usage-based model, maintains that, in terms of storing, there is no qualitative difference in the lexical processing mechanism, emphasizing the effects of lexical exposure. The significant interactions of multiword frequency and both groups found in this experiment are consistent with the view of the lexical entrenchment account that all L2 participants with less exposure

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

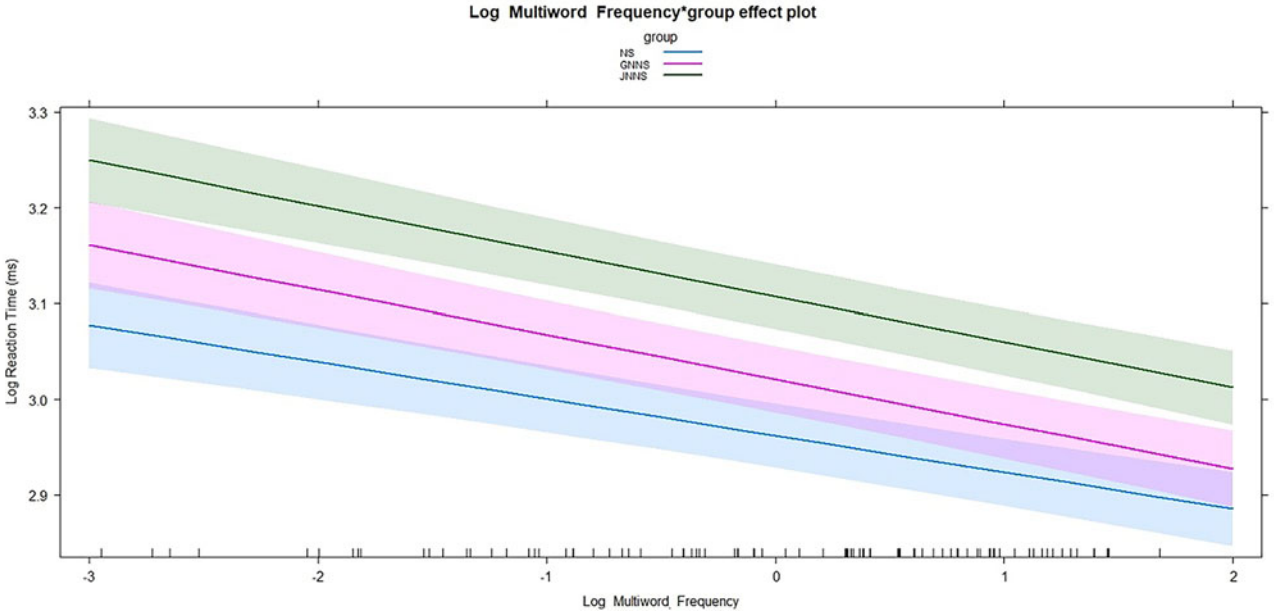


Figure 1. The interaction between log multiword frequency and group
Note. NS = native speakers; JNNS = Japanese non-native speakers of English; GNNS = German non-native speakers of English.

Table 4. Main effects and interactions for three participant groups with individual vocabulary knowledge

	Reaction Time				*Accuracy*			
Fixed effects	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*z*	*p*
(Intercept)	2.98	.03	87.96	<.001	4.01	.30	13.41	<.001
Log Multiword Frequency	−.04	.01	−7.11	<.001	1.10	.19	5.92	<.001
Group1	.04	.04	1.04	.30	−.45	.31	−1.45	.15
Group2	.08	.04	1.84	.07	−.48	.38	−1.26	.21
Vocabulary Knowledge	.00	.00	−.58	.56	.01	.01	1.13	.26
Log Multiword Frequency × Group1	−.01	.00	−2.34	<.05	−.39	.17	−2.23	<.05
Log Multiword Frequency × Group2	−.01	.01	−1.99	<.05	−.19	.22	−.90	.37
Log Multiword Frequency ×Vocabulary Knowledge	.00	.00	−.56	.58	.01	.01	1.19	.23
Group1 × Vocabulary Knowledge	.00	.00	−.18	.86				
Group2 × Vocabulary Knowledge	.00	.00	−1.03	.30				
Control variable								
Log Word Frequency	.00	.01	−.19	.85	−.16	.32	−.50	.62
Random effects	Variance	SD			Variance	SD		
Item (intercept)	.00	.04			1.19	1.09		
Log Word Frequency	.00	.02						
Subject (intercept)	.01	.09			.25	.50		
Log Word Frequency	.00	.02						

Note. Group1 = native English speakers vs. German non-native English speakers;
Group2 = native English speakers vs. Japanese non-native English speakers.
*RT Formula: log rt~log word frequency + log multiword frequency*group + vocabulary knowledge*group + log multiword frequency*vocabulary knowledge+(1 + word frequency|subject)+(1 + word frequency |item); rt = reaction times;*
*Accuracy Formula: error~log multiword frequency*vocabulary knowledge + group*log multiword frequency + log word frequency +(1|subject)+(1|item)*

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

Table 5. Main effects and interactions for Japanese and German participant groups with individual vocabulary knowledge

	Reaction Time				*Accuracy*			
Fixed effects	Estimate	*SE*	*t*	*p*	Estimate	*SE*	*z*	*p*
(Intercept)	3.03	.02	166.80	<.001	3.58	.26	13.76	<.001
Log Multiword Frequency	−.05	.01	−7.88	<.001	.68	.17	4.05	<.001
Group	.06	.02	2.29	<.05	−.11	.28	−.39	.70
Vocabulary Knowledge	.00	.00	−1.28	.20	.01	.01	.49	.63
Log Multiword Frequency × Group	.00	.01	−.28	.78	.18	.16	1.10	.27
Log Multiword Frequency × Vocabulary Knowledge	.00	.00	.15	.88	.01	.01	.89	.37
Group × Vocabulary Knowledge	.00	.00	−1.15	.26				
Log Multiword Frequency × Group × Vocabulary Knowledge	.00	.00	−.64	.52				
Control variable								
Log Word Frequency	−.02	.01	−1.17	.24	−.17	.35	−.50	.62
Group × Log Word Frequency	.02	.01	2.21	<.05				
Random effects	Variance	SD				Variance	SD	
Item (intercept)	.00	.05				1.38	1.17	
Subject (intercept)	.01	.08				.27	.52	
Log Multiword Frequency	.00	.01						

Note. Group = German non-native English speakers vs. Japanese non-native English speakers.
*RT Formula: log rt~log multiword frequency*group*vocabulary knowledge + log word frequency *group+(1 + log multiword frequency|subject)+(1|item); rt = reaction times;*
*Accuracy Formula: error~group*log multiword frequency + log multiword frequency*vocabulary knowledge + log word frequency+(1|subject)+(1|item)*

to the target language would show greater effects. In the present study, both JNNS with orthographically and structurally different L1 and GNNS with similar L1 to English showed stronger frequency effects of recurrent word combinations. It should also be noted that no significant difference in the multiword frequency effects between the two L2 groups was observed. The results endorse the ideas of the lexical entrenchment account.

Diependaele et al. (2013) argued that the strength of lexical representations in L2 was weaker than in L1. It was assumed that subjective frequency in L2 was lower than (corpus-based) objective frequency, and there were greater gaps between subjective and objective frequencies among L2 speakers, especially in the low-frequency range, because of the resting levels of words. This account can also be applied to multiword frequency effects; subjective frequencies for multiword chunks in L2 are usually lower because of less exposure to the language, compared to L1. As Figure 1 reveals, the gaps in reaction times among the three participant groups in the lower-frequency range were greater than those in the high-frequency range. The differences in the subjective and objective frequencies of low-frequency multiword units seemed to result in a steep increase in reaction latencies and accuracy.

An additional explanation might be needed for the different magnitudes of the multiword frequency effects between L1 and L2 because the multiword units used in this grammaticality judgment experiment did not consist of low-frequency individual words. Reaction latencies for multiword sequences seemed to depend on the time required to analyze the compound words in the multiword sequences. Jiang and Nekrasova (2007) pointed out that no syntactic analysis would be required if high-frequency multiword expressions are lexicalized and processed as a chunk. It is reasonable to assume that participants need to focus on one-by-one analysis as multiword frequency decreases. The participants would analyze the syntactic information of the low-frequency multiword sequences when making a grammatical judgment but would not need to do so for the high-frequency multiword units. Analyzing syntactic forms appears to cause the differences in reaction latency and accuracy. The findings of the present study clearly indicated that the more frequent the multiword sequences were, the faster they were responded to because less syntactic analysis was necessary. According to Wood (2010), when a sequence acquires a formulaic-like status with frequent usage and production, it leads to automatization of the string and memorizing it as a piece of procedural knowledge in which syntactic analysis is no longer conducted. Wolter and Gyllstad (2011, 2013) defined such automatization as entrenchment, representing a process wherein a structural construction becomes automated into a unit. The entrenchment involves the degree to which activation of a multiword sequence is a highly automated process through L2 language input and output. The more strongly entrenched a structure is, the less specific attention is required when recognizing and producing frequently used specific language structures. Wolter and Gyllstad (2011, 2013) also suggested that the disuse or infrequent use of a certain structure is likely to result in negative effects on the entrenchment. Therefore, the processing of low-frequency multiword sequences,

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press

in which individual single words are weakly linked, may be time-consuming because of the syntactic analysis that they necessitate. In the future, to answer the question of whether syntactic analysis no longer occurs when recognizing frequently recurring multiword expressions, further additional data, such as eye-tracking or electroencephalogram (EEG) data, will be beneficial.

With respect to the effects of language exposure or proficiency, L2 researchers have used vocabulary knowledge as a measurement of exposure to the target language, due to strong correlations between vocabulary knowledge and language exposure, and the difficulty in accurately measuring lexical exposure by self-reports. Diependaele et al. (2013) stated that vocabulary knowledge can explain the differences in word frequency effects across NNS (among French, German, and Dutch English L2 bilinguals). However, in the present study, vocabulary knowledge did not eliminate the interaction between multiword frequency and groups. The differences in the multiword frequency effects persisted even after taking vocabulary knowledge into account. This result agrees with some previous studies (e.g., Cop et al., 2015) and indicates that individual vocabulary knowledge is not a satisfactory explanation for the differing multiword frequency effects among groups. Moreover, no evidence emerged from the current study to obtain the interaction of vocabulary knowledge and logarithmic multiword frequency or the main effect of multiword frequency. Emerging evidence revealed that increased individual vocabulary knowledge does not seem to reduce multiword frequency effects.

An important point to note is that GNNS participants responded to multiword sequences significantly faster than the JNNS participants but did not significantly differ in the degree of multiword frequency effects. Given the comparable degree of the effects among two L2 participants, the data in the current study provide evidence that the effects' differences between L1 and L2 could be ascribed to within-language characteristics. Even though the GNNS group with more extensive vocabulary knowledge generally had considerable orthographical access to the shared alphabet in their daily lives, no reduced multiword frequency effects were observed. It becomes possible to hypothesize that increased vocabulary knowledge accelerates recognition but does not necessarily decrease multiword frequency effects. One can argue that between-language factors do not appear to be a critical factor in explaining different multiword frequency effects.

Conclusion

This study was undertaken with the goal of determining whether there would be different magnitudes of multiword frequency effects among L1 and L2 multiword frequency effects, in addition to observing the processing advantage of frequently occurring sets of words. Similar to the findings of Jiang and Nekrasova (2007) and Hernández et al. (2016), this study observed a processing advantage for high-frequency multiword units and sensitivity to the frequency of multiword sequences. This study also identified a significant difference in the multiword frequency effects between L1 and L2 English speakers. The multiword frequency effects were greater for L2 speakers than for L1 speakers, similar to the single-word level. These findings support the lexical entrenchment account. Nonetheless, individual vocabulary knowledge did not eliminate the differences in multiword frequency effects, leading to the assumption of the possibility that it was not the explanatory factor.

This study revealed that no significant discrepancies in sensitivity to the distributional property of multiword frequency were found in two L2 populations. Enlarged vocabulary knowledge would produce accelerated visual recognition for multiword units but does not lead to decreased multiword frequency effects. The lack of significant influence appears to indicate that multidimensional proficiency measures related to collocational and syntactic knowledge, as well as wide vocabulary knowledge need to be tested to uncover the explanation for the different impacts of multiword frequency effects. It is also of value to note that the subjective frequency of multiword sequences among L1 and L2 groups differs from corpus-based objective frequency. Therefore, measures of subjective frequency, such as familiarity rate, should be used in future research.

Acknowledgments. This research was supported by a Grand-in-Aid for Scientific Research from the Japan Society for the Promotion of Science (No. 18K00780). The author is most grateful to Mayumi Kajiura and Masatoshi Sugiura for their help in data collection and to two anonymous reviewers for their comments on earlier versions of the paper.

Competing interest. The author declares that they have no conflict of interest.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728923000548

Notes

1 The follow-up survey revealed that participants had between six months and seven years' experience studying Japanese, so their Japanese language proficiency could be estimated at the beginner through intermediate levels.
2 The number of practice trials was based on Duyck et al.'s (2008) experiment. The practice stimuli were 10 grammatical and 10 ungrammatical phrases that were not used in the main experiment.
3 The cut-off set was based on Jiang and Nekrasova (2007) who examined multiword processing.

References

Altenberg, B. (1998). On the phraseology of spoken English: The evidence of recurrent word-combinations. In A. P. Cowie (Ed.), *Phraseology: Theory, Analysis and applications* (pp. 101–122). Oxford University Press.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Biber, D. (2009). A corpus-driven approach to formulaic language in English: Multi-word patterns in speech and writing. *International Journal of Corpus Linguistics*, *14*(3), 275–311. https://doi.org/10.1075/ijcl.14.3.08bib.

BNC Consortium. (2001). The British national corpus (second version). Shogakukan corpus network [Distributor]. http://scnweb.jkn21.com/BNC2/.

Brysbaert, M., Lagrou, E., & Stevens, M. (2017). Visual word recognition in a second language: A test of the lexical entrenchment hypothesis with lexical decision times. *Bilingualism: Language and Cognition*, *20*(3), 530–548. https://doi.org/10.1017/S1366728916000353.

Conklin, K., & Schmitt, N. (2008). Formulaic sequences: Are they processed more quickly than nonformulaic language by native and nonnative speakers? *Applied Linguistics*, *29*(1), 72–89. https://doi.org/10.1093/applin/amm022.

Cooper, T. C. (1999). Processing of idioms by L2 learners of English. *TESOL Quarterly*, *33*(2), 233–262. https://doi.org/10.2307/3587719.

Cop, U., Keuleers, E., Drieghe, D., & Duyck, W. (2015). Frequency effects in monolingual and bilingual natural reading. *Psychonomic Bulletin and Review*, *22*(5), 1216–1234. https://doi.org/10.3758/s13423-015-0819-2.

Davies, M. (2008). The Corpus of Contemporary American English (COCA). https://www.english-corpora.org/coca/.

Diependaele, K., Lemhöfer, K., & Brysbaert, M. (2013). The word frequency effect in first- and second-language word recognition: A lexical entrenchment account. *The Quarterly Journal of Experimental Psychology*, *66*(5), 843–863. https://doi.org/10.1080/17470218.2012.720994.

Duyck, W., Vanderelst, D., Desmet, T., & Hartsuiker, R. J. (2008). The frequency effect in second-language visual word recognition. *Psychonomic Bulletin and Review*, *15*(4), 850–855. https://doi.org/10.3758/PBR.15.4.850.

Forster, K. I., & Forster, J. C. (2003). DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods, Instruments, and Computers*, *35*(1), 116–124. https://doi.org/10.3758/bf03195503.

Gibbs, R. W., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition*, *20*(3), 243–259. https://doi.org/10.1016/0010-0277(85)90010-1.

Gollan, T. H., Montoya, R. I., Cera, C., & Sandoval, T. C. (2008). More use almost always means a smaller frequency effect: Aging, bilingualism, and the weaker links hypothesis. *Journal of Memory and Language*, *58*(3), 787–814. https://doi.org/10.1016/j.jml.2007.07.001.

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. https://doi.org/10.1080/23273798.2016.1152389.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception and Psychophysics*, *40*(6), 431–439. https://doi.org/10.3758/bf03208203.

Irujo, S. (1993). Steering clear: Avoidance in the production of idioms. *IRAL – International Review of Applied Linguistics in Language Teaching*, *31*(3), 205–220. https://doi.org/10.1515/iral.1993.31.3.205.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal*, *91*(3), 433–445. https://doi.org/10.1111/j.1540-4781.2007.00589.x.

Kim, S. H., & Kim, J. H. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL Quarterly*, *46*(4), 831–841. https://doi.org/10.1002/tesq.66.

Kuperman, V., Bertram, R., & Baayen, R. H. (2008). Morphological dynamics in compound processing. *Language and Cognitive Processes*, *23*(7–8), 1089–1132. https://doi.org/10.1080/01690960802193688.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid Lexical Test for advanced learners of English. *Behavior Research Methods*, *44*(2), 325–343. https://doi.org/10.3758/s13428-011-0146-0.

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), 12–31. https://doi.org/10.1037/0278-7393.34.1.12.

Mor, B., & Prior, A. (2020). Individual differences in L2 frequency effects in different script bilinguals. *International Journal of Bilingualism*, *24*(4), 672–690. https://doi.org/10.1177/1367006919876356.

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

Perry, C., Ziegler, J. C., & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, *114*(2), 273–315. https://doi.org/10.1037/0033-295X.114.2.273.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372.

Rayner, K. (2009). The Thirty-Fifth Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461.

Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory and Cognition*, *14*(3), 191–201. https://doi.org/10.3758/BF03197692.

R Development Core Team. (2013). *R: A Language and Environment for Statistical Computing*. http://www.r-project.org. R Foundation for Statistical Computing.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Language learning and language teaching,* (Vol. 9). *Formulaic sequences: Acquisition, processing and use* (pp. 173–189). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.9.10sch.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127–151). John Benjamins Publishing Company. https://doi.org/10.1075/lllt.9.08sch.

Schröter, P., & Schroeder, S. (2018). Differences in visual word recognition between L1 and L2 speakers: The impact of length, frequency, and orthographic neighborhood size in German children. *Studies in Second Language Acquisition*, *40*(2), 319–339. https://doi.org/10.1017/S0272263117000201.

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

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011b). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multiword sequences. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(3), 776–784. https://doi.org/10.1037/a0022531.

Tiffin-Richards, S. P., & Schroeder, S. (2015). Word length and frequency effects on children's eye movements during silent reading. *Vision Research*, *113*(A), 33–43. https://doi.org/10.1016/j.visres.2015.05.008.

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing advantages of lexical bundles: Evidence from self-paced reading and sentence recall tasks. *Language Learning*, *61*(2), 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x.

Whitford, V., & Titone, D. (2012). Second-language experience modulates first- and second-language word frequency effects: Evidence from eye movement measures of natural paragraph reading. *Psychonomic Bulletin and Review*, *19*(1), 73–80. https://doi.org/10.3758/s13423-011-0179-5.

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. https://doi.org/10.1093/applin/amr011.

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. https://doi.org/10.1017/S0272263113000107.

Wood, D. (2010). *Formulaic language and second language speech fluency: Background, evidence and classroom applications*. Continuum.

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

https://doi.org/10.1017/S1366728923000548 Published online by Cambridge University Press