Processing Collocations in First (L1) and Second Languages (L2): Effects of Collocation- and Word-level Variables on Speakers Varying in Proficiency and Dominance

by

Rundi Guo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Psychology)
in the University of Michigan
2022

Doctoral Committee:

Professor Julie Boland, Co-Chair
Professor Emeritus Nick Ellis, Co-Chair
Professor San Duanmu
Associate Professor Ioulia Kovelman
Professor Min Wang, Xi'an Jiaotong University

Rundi Guo

guorundi@umich.edu

ORCID iD: 0000-0001-7643-3177

Dedication

This work is dedicated to all who came before me and all who will be inspired. Thank you to everyone who supported me along the way, including my advisors, members of my dissertation committee, faculty members and students in the psychology and linguistics departments, members of the Psycholinguistics Lab, all my friends and family, and my incredible partner.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Collocations are words that have a tendency to co-occur within a few words' spans, e.g., "drink coffee" and "dark chocolate" in English. Growing empirical evidence suggests that both native (L1) speakers and advanced second language (L2) learners process two-word collocations faster than unconnected word pairs, and that speakers and learners are sensitive to the frequency distributions of linguistic units beyond individual words. Here we investigate this processing advantage of collocations in L1 and L2 and unravel the factors underlying collocation processing in native speakers (of English, Exp 1; of Chinese, Exp 2) and non-native speakers (of L2 English, Exp 3). In a series of double Lexical Decision Tasks (LDT), participants were presented with pairs of letter- or character-strings and were asked to judge if both strings in a pair were correctly-spelled words in English (Exp 1 & 3) or in Chinese (Exp 2). The word pairs of all three experiments fall into four conditions: 1. collocations in English but not Chinese (e.g., honest mistake, 诚实错误), 2. collocations in Chinese (if translated verbatim) but not English (e.g., dead road, 死路), 3. collocations in both languages (e.g., deep sleep, 深睡眠), and 4. baseline controls in which the two words were unconnected in both languages (e.g., bright hand, 明亮手). Exp 1 and 2 demonstrated faster and more accurate processing of collocations in native speakers of both languages. Most interestingly, Exp 3 showed facilitated processing of English collocations for Chinese-English bilingual speakers, and the effect size increased as English experience increased. General linear regression models including distribution variables of different grain sizes in both languages revealed that highly advanced L2 speakers process collocations in a similar way to L1 speakers, with their performance being more impacted by collocation-level factors and less impacted by word-level factors, whereas the reverse pattern is found for beginner L2 speakers. This suggests that as language dominance and proficiency grows, learners become increasingly sensitive to the statistical associations relating to larger chunks – from single words to collocations and potentially other multiword units.

Keywords: collocation, formulaic language, frequency, lexical decision task, L1 influence, Phrase-Superiority Effect

Chapter 1 Introduction

It is estimated that up to 50% of the written and spoken discourse that a native speaker encounters on a daily basis consists of formulaic multi-word units (MWUs) (Howarth, 1998; De Cock et al., 1998). MWUs are defined as continuous or discontinuous sequences of words or other elements that co-occur more frequently than by chance, and they encompass a wide variety of linguistic phenomena such as idioms (*spill the beans*), lexical bundles (*is in front of the*), phrasal verbs (*put on*), irreversible binomials (*black and white*), collocations (*strong tea*), etc. Despite the heterogeneity in terms of length, structure, and transparency in meaning across the different types of MWUs, they are recognized as an important component of one's language competence (Wolter & Yamashita, 2015) and have been linked to other indicators of competence such as vocabulary knowledge (Crossley et al., 2015), spontaneous speech proficiency (Xu, 2015), and language fluency (Wray, 2002).

Language consists of units of various sizes and can be represented at different levels. From the perspective of usage-based approaches, it is the exposure to these various linguistic stimuli and the rich distributional information associated with them that gives rise to proficiency, and as a result, proficiency is a reflection of language exposure and use coupled with general cognitive factors such as blocking due to learned attention bias (Ellis, 2006; Wulff & Ellis, 2018). The human mind is naturally attuned to linguistic distributional information such as frequency and contingency, and through the (oftentimes unconscious) process of statistical learning, language learners are able to extract the underlying patterns and regularities of the language to use it more effectively (Ellis, 2002). Besides the lexical frequency effects that have been well-documented in the empirical literature (e.g., Balota & Chumbly, 1984; Barry & Seymour, 1988; McDonald & Shillcock, 2004), it appears that language learners and users are also attuned to the statistical information of linguistic units beyond individual words (e.g., Bannard & Matthews, 2008; Arnon & Snider, 2010; see Ellis, 2002 for a review). Infants as young as 11-12 months are already sensitive to the frequency of formulaic sequences despite only being in the one-word production stage (Skarabela et al., 2021). Three-year-old children are more likely to correctly produce frequent MWUs than infrequent control sequences in a sentence-repetition task (Bannard & Matthews, 2008). Such sensitivity to MWUs is also observed in adult speakers in a variety of behavioral measures such as the phrasal-decision task (Arnon & Snider, 2010) and self-paced reading (Tremblay et al., 2011). MWUs are processed faster than matched non-formulaic control phrases by both native and second language speakers (Conklin & Schmitt, 2008; Wolter & Yamashita, 2018). Many researchers believe that our baseline strategy in everyday use of language inherently relies on the expected likelihood of recurrent formulaic sequences rather than unexpected novel combinations (e.g., Wray & Perkins, 2000). In fact, for both native (NSs) and non-native speakers (NNSs), the processing and production of individual words and morphemes are enhanced when they appear in familiar multiword frames compared to when they are embedded in nonformulaic multi-word frames or presented alone (Childers & Tomasello, 2001; Arnon & Clark, 2011; Guo & Ellis, 2021).

A prominent type of MWU is collocation, which describes the lexical relation between (typically) two words that have a tendency to co-occur within a few words' span[1]. For example, in

[1] There exist various ways to define and operationalize collocations (see Gablasova et al., 2017 for review); here we mainly used the corpus-based approach, which identifies word co-occurrences by drawing on quantitative measures based in frequency and distributional patterns extracted from large corpora (e.g., Wolter & Yamashita, 2015, Gablasova et al., 2017, Öksüz et al 2021), verified with a native speaker judgment survey.

English, “drink” forms a collocation with “coffee” but not with “soup”; “dark” forms a collocation with “chocolate” but not with “coffee”. Notably, collocations are usually dictated more by convention within the language than by grammatical or semantic restrictions (Wolter & Yamashita, 2015). In the examples given above, neither “drink soup” nor “dark coffee” contains any obvious grammatical or semantic violations in English, yet they are typically not considered as commonly-used expressions by native English speakers. Due to this arbitrariness in lexical patternings, it is possible that an unconnected word-pair in one language (e.g., “drink soup” in English) forms a legitimate collocation in another (“drink soup” in Chinese, meaning “eat soup” in English), and vice versa; as it is also possible that the same two words form the same collocation in two languages (e.g., “drink coffee” in both English and Chinese), in which case the collocation is said to be *congruent* across two languages.

1.1 Investigating collocation processing in L1 and L2

Psycholinguistic research has demonstrated the psychological reality of collocations in that collocations are processed faster than unconnected word pairs by NSs and advanced NNSs. For example, Siyanova and Schmitt (2008) investigated NSs’ and NNSs’ online collocation processing using a timed subjective familiarity judgment task and found that both groups of participants responded to high-frequency collocations faster than to unconnected word-pairs, although unsurprisingly, NNSs had an overall slower reaction time (RT) and rated frequent collocations as less familiar than did NSs. Durrant and Doherty (2010) conducted a primed lexical decision task with adult NSs and found a priming effect wherein the processing of the target (second word) is facilitated when the prime and target form a high-frequency collocation, even if the two words were not semantically related. Vilkaité (2016) examined adult NSs' eye-movements when reading sentences containing two types of collocations (either the collocates were adjacent to each other, "*provide information*", or they were nonadjacent "*provide* some of the *information*") or matched control phrases. Results showed that both types of collocations were read faster than controls, with the advantage slightly larger for adjacent than nonadjacent collocations. The processing advantage of collocations has led many psycholinguists to hold the view that collocations, along with other MWUs, are processed holistically as one unit or construction, i.e., they are stored and retrieved whole from memory at the time of use, without being analyzed as individual words combined with a certain grammar (Wray, 2002: 9). If an extreme version of this holistic processing hypothesis is true, we would predict that the processing of collocations, at least for NSs, should only be affected by the characteristics and distributional information associated with the collocation level, and not affected by the characteristics and distributional information associated with the word level. A weaker version of this hypothesis might predict that the processing of collocations would be more affected by the characteristics and distributional information associated with the collocation level than those associated with the word level, while information at both levels is available for retrieval at all times.

While the appropriate use of collocations typically comes naturally for native speakers and is considered as a prerequisite for proficient language use (e.g., Wray, 2002), it is widely acknowledged that for NNSs, native-like use and competence of collocations is much more the exception than the norm. The notorious difficulty collocations can create for NNSs is well-acknowledged in research on L2 acquisition and pedagogy (Pawley & Syder, 1983). Offline paper-and-pencil tests show that learners varying in their L1 backgrounds have poor knowledge of L2 collocations, scoring an Accuracy rate of under 50% in a variety of tasks, including translation, cloze, multiple-choice, commonality judgment, etc. (e.g., Bahns & Eldaw, 1993; Siyanova & Schmitt, 2008, study 2; Nguyen & Webb, 2017). Analyses of L2 production samples, as well as learner corpora, also show that NNSs not only produce fewer collocations overall compared to native speakers (Fitzpatrick, 2006), but even advanced learners of an L2 still make frequent errors or produce atypical, albeit grammatical, word combinations (e.g., "perform a project") in their language use (Howarth, 1998; Nesselhauf, 2003; Siyanova & Schmitt, 2008, study 1; Laufer & Waldman, 2011).

According to the holistic processing view mentioned above, the difficulty observed in L2 collocation processing is attributed to fundamentally different mechanisms involved in collocation processing in L1 vs. L2: namely, whereas NSs process collocations holistically as chunks without disintegrating them into component words, NNSs are less susceptible to holistic processing, and are more prone to adopt an analytical and bottom-up processing approach, relying heavily (and sometimes solely) on the individual words and rules (Wray, 2002). Therefore, for NNSs, the characteristics and distributional information associated with the whole collocations (i.e., top-down influence) are less likely to affect the processing speed compared to the characteristics and distributional information at the single-word level.

However, this view is under debate. Emerging empirical evidence suggests that NNSs, at least those who are highly proficient in their L2, may still be sensitive to the frequencies of L2 linguistic units beyond single words during on-line processing of various types of MWUs. For example, in an eye-tracking reading experiment, Siyanova-Shanturia et al. (2011) found that both NSs and highly proficient NNSs process binomials like "bride and groom" significantly faster than their atypical reversed forms "groom and bride", while lower proficiency NNSs reading patterns were similar for both types of phrases, which the authors attributed to their lack of exposure to L2 binomials. Notably, the fact that the individual words were matched in the two forms means that the processing advantage of the typical form was likely attributable to the frequency associated with the whole phrase rather than single words. Jiang and Nekrasova (2007) examined the processing of lexical bundles in a grammaticality judgment task and found that both NSs and NNSs were faster and more accurate on frequent phrases than on infrequent phrases. In the case of collocations, Yamashita and Jiang (2010) used a timed phrasal acceptability judgment task consisting of English collocations and unconnected word pairs on English NSs and NNSs whose L1 was Japanese. Some of the collocation items were *congruent* in English and Japanese (i.e., same/similar form and meaning in both languages), while others were *incongruent* (i.e., the English form does not have a direct translation equivalent in Japanese). Results showed that both NSs and NNSs processed English collocations faster and more accurately than the baseline unconnected word pairs, replicating the collocation processing advantage in both L1 and L2. Additionally, NNSs were more accurate on congruent collocations than incongruent English-only collocations, although this advantage was not significant in the RTs. In terms of RTs, lower proficiency NNSs (but not high proficiency NNSs) had slower RTs on incongruent collocations compared to NSs. This indicates that while NNSs are sensitive to L2 collocations, the roles of L1 knowledge and L2 exposure/proficiency are equally noteworthy.

Extending this line of research, Wolter and colleagues (Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018) conducted a series of experiments using timed phrasal acceptability judgment tasks involving NSs' and NNSs' on-line processing of congruent and incongruent collocations, demonstrating the congruence effect in the RTs. Namely, whereas NSs processed the congruent and incongruent collocations equally well, and both faster than baseline unconnected word-pairs, NNSs processed congruent collocations the fastest, followed by incongruent collocations, followed by baseline. Importantly, Wolter and colleagues also investigated the effects of lexical (single-word) frequency and collocation frequency on the NSs' and NNSs' performance, finding that across the board (for both NSs and NNSs and for all types of collocations), both word frequency and collocation frequency were significantly predictive of RTs, with word frequency having a more pronounced effect on NNSs than NSs, and collocation frequency having a more pronounced effect on NSs than on NNSs. Nonetheless, both NSs and NNSs were affected by the frequencies at both levels, suggesting some overlap between the L1 and L2 collocation processing mechanisms. However, it is still unclear whether the observed discrepancies between NSs and NNSs are indeed due to fundamental qualitative differences (e.g., Wray, 2002), or if they can be accounted for by quantitative differences in exposure (e.g., Siyanova-Shanturia et al., 2011).

Besides lexical and collocation frequency, contingency is also believed to play an important role in the acquisition and processing of MWUs (e.g., Ellis, 2006; Gries & Ellis, 2015). Contingency has been measured in various ways, e.g., Mutual Information (MI) scores (Durrant & Doherty, 2010), t-score (Wolter & Gyllstad, 2011), log Dice scores (Gablasova et al., 2017), ΔP (Gries & Ellis, 2015), G^2 (Dunning, 1993), etc. The present study adopted MI as it has been shown to have high construct validity (Gries, 2022)[2]. In collocations, contingency can be understood as the strength of statistical association between the two component words. Behavioral studies confirmed the facilitation effect of contingency on on-line collocation processing using the similar phrasal acceptability judgment paradigm. For example, Yi (2018) found both NSs and advanced NNSs sensitive to both collocation frequency and collocation strength (measured by MI). In fact, advanced NNSs demonstrated even stronger sensitivity to these two measures than did L1 speakers. Interestingly, the speakers' sensitivity to such language statistics was independent of

[2] For an in-depth review and comparison of different association measures and their validity, see Gries (2022).

their cognitive and linguistic aptitude as measured by six aptitude tests which loaded separately onto implicit language aptitude, explicit language aptitude, and working memory capacity. Öksüz et al (2021) also concluded that both NSs and advanced NNSs were affected by collocation strength, irrespective of how it was operationalized (MI vs. log Dice scores), although unlike Yi (2018), the two groups of speakers in Öksüz et al (2021) showed comparable sensitivity to both collocation frequency and collocation strength. In addition, Öksüz et al. (2021) also did not find any significant difference between the NSs' and NNSs' sensitivity to word-level frequency, contrary to Wolter and Yamashita (2018) and Fang and Zhang (2021). For both speaker groups in Öksüz et al (2021), the effect of word frequency information becomes weaker as collocation frequency increases. Taken together, it seems that language learners and users are remarkably attuned to both word-level and collocation-level distribution patterns in the processing of collocation even when they have relatively limited exposure to a language, although the nature of the differences between NSs and NNSs in their sensitivity to the different grain-sizes is less clear.

1.2 Additional considerations for L2 collocation processing

Language speakers' sensitivity to word frequency, collocation frequency, and association strength seems to be ubiquitous for collocation processing in both NSs and NNSs. However, for NNSs, an additional piece of the puzzle that calls for more empirical attention is the role of language proficiency and/or dominance. Does the NNS's proficiency in the L2 play a facilitative role in the processing of L2 collocations, and crucially, does the learners' sensitivity to the various distribution statistics at different linguistic levels gradually change as they become increasingly proficient or dominant in their L2? Several studies on L2 collocation processing have examined the effect of L2 proficiency by dividing NNSs into proficiency groups (e.g., Yamashita & Jiang, 2010; Wolter & Yamashita 2015; Ding & Reynolds, 2017; Wolter & Yamashita 2018; Fang and

Zhang, 2021). Unsurprisingly, high proficiency NNSs generally processed L2 collocations faster and more accurately than low proficiency NNSs. In terms of the potential interaction between the learners' proficiency and their sensitivity to the various types of distribution information, Wolter and Yamashita (2018) found that while collocation frequency has a bigger effect on NSs than on NNSs, it was the low proficiency NNSs that drove the interaction, and that high proficiency NNSs demonstrated sensitivity to collocation frequency to a similar extent as did the NSs. Word frequency, conversely, had a more prominent effect on both NNSs groups (which did not differ) than on NSs. Fang and Zhang (2021) observed a three-way interaction between L2 proficiency, word frequency, and collocation frequency. Namely, when the frequency of a collocation is low, both low proficiency and high proficiency NNSs rely more on word frequency than did NSs. As the collocation frequency goes up, while the high proficiency NNSs are increasingly more affected by collocation frequency in a similar fashion to NSs, the low proficiency NNSs continue to rely more on word frequency. The interpretation of the three-way interaction is not as straightforward, but the trend is consistent with Wolter and Yamashita's (2018) observation that as NNSs become advanced in their L2, they tend to rely more on the distribution information such as frequency associated with the entire collocation as their NS counterparts do, although the results regarding NNSs' sensitivity to the word-level distribution information and its interaction with L2 proficiency is less clear. Since most of the studies examining L2 proficiency only included two categorical levels (e.g., high proficiency vs. low proficiency, or ESL vs. EFL, etc), with much of what constitutes "proficiency" largely uncontrolled for, research with a better defined and variegated proficiency factor is still needed to unveil its more nuanced effects and its interaction with various distributional variables. Adding more levels and variability of the proficiency measure would also help us gain insights into the development of L2 collocation processing based on a more refined L2 growth trajectory that better represents the incremental nature of L2 exposure.

Another noteworthy issue unique to collocation processing in L2 is the role of L1 knowledge. There are a plethora of studies investigating the effect of L1 influence on the processing of collocations and other MWUs, but the operationalization of L1 influence has almost always been L1-L2 congruence (e.g., Yamashita & Jiang, 2010; Ding & Reynolds, 2017; Wolter & Yamashita 2015, 2018; Fang & Zhang, 2021). These studies confirmed the influence of L1 knowledge by demonstrating the congruence effect, namely, when the frequency of collocations is held constant, congruent collocations are recognized and processed better than collocations that do not have cross-language equivalence (i.e., an overlap between L1 and L2 aids the online processing of L2 collocation); this suggests knowledge of L1 collocations can be utilized in the processing of the L2 to some extent. Importantly, this congruence effect seems to diminish as L2 proficiency increases (Yamashita & Jiang, 2010; Ding & Reynolds, 2017).

In the present study, we aim to identify a more complete picture of the role of L1 by looking at other indicators of L1 influence and how they interact with L2 proficiency. It would be interesting to see if the impact of L1 collocation knowledge on L2 collocation processing manifests in other ways. One example is the "false-friend" effect, i.e., will the processing of an unconnected word-pair in L2 be affected if the two words' verbatim translations form a collocation (i.e., a "false-friend") in the L1? For example, will a Chinese-English bilingual process an unconnected English word-pair such as "cold joke" (which forms a Chinese collocation with a meaning equivalent to the English collocation "dad joke") differently from a word-pair that is unconnected in either language? This false-friend effect has been examined on the lexical level where it is more widely known as the false-cognate effect. For example, Cañizares-Álvarez and Gathercole (2019) found that Spanish L1 English L2 NNSs had more difficulties rejecting false-cognates in cloze (multiple-choice) tasks, especially when the L2 form is highly frequent (relative to L1). Several studies also investigated false-friend effects in MWUs but have found mixed results. Carrol, Conklin & Gyllstad (2016) examined the eye-movements by Swedish L1 English L2 NNSs as they read English sentences containing idioms that either only exists in English (L2-only), only exist in Swedish and are translated verbatim into English (L1-only), or exist in both languages (congruent). Despite the inherent reluctance for L2 speakers to translate L1-only items when they seem idiomatic (e.g., Kellerman, 1986), the authors found that both the L1-only false-friends and the congruent collocations were read faster than the L2-only collocations, suggesting a false-friend advantage. Carrol and Conklin (2014, 2017) subsequently found a similar processing advantage in L1-Chinese English NNSs for Chinese idioms translated verbatim into English. Likewise, Ueno (2009) used a primed lexical decision task and found that L1-Japanese English NNSs responded more quickly to unconnected English word pairs that were translations of L1 Japanese collocations (like "forgive marriage"). Contrarily, Wolter and Yamashita (2015, 2018) also examined L1-Japanese English NNSs using a similar distribution of items (English-only, Japanese-only, congruent, baseline) and yet found no processing advantage for Japanese-only false-friend items irrespective of L2 proficiency. These inconsistent findings call for further research.

Another indicator of L1 influence that is worth examining is the potential effect of L1 frequency and other distributional statistics on the processing of the L2 translational equivalent. This will be referred to as the "L1 carry-over effect". In a corpus analysis of L2 learners' writing samples, Paquot (2017) found that the learners' preferred use of three-word lexical bundles was highly correlated with the frequency of the translation equivalent form in the learners' L1, suggesting that at least in off-line processing and production, learners draw knowledge from their

L1, including the statistical knowledge of how words go together. Wray and Perkins (2000) found a similar pattern in L2 speech, whereby less proficient L2 users tend to overuse L1 collocational patterns. However, to our knowledge, the only study that investigated the effect of L1 frequency in an on-line task involving translation equivalent of collocations in the L2 failed to find any evidence of the carry-over effect of L1 frequency on the processing of congruent collocations (Wolter & Gyllstad, 2013). Since these results were based on a relatively homogeneous group of highly proficient L2 speakers, the lack of L1 frequency carry-over effect might not be generalizable to learners with lower proficiency levels. Due to the paucity of empirical research on this topic, it remains unclear whether NNSs draw on the L1 distributional statistics during the on-line processing of L2 collocations, and if so, whether this tendency changes with growing proficiency and L2 exposure.

1.3 The present study

Our current research explores the processing of L1 and L2 collocations in three experiments each driven by a main goal:

- First (Exp 1), we wish to replicate the previously observed collocation advantage in English to see if collocations, compared to unconnected word-pairs, are psychologically privileged in English NSs. Additionally, we aim to identify important word-level and collocation-level distribution variables contributing to this processing advantage.
- Second (Exp 2), we aim to extend the findings on English to Mandarin Chinese (hereafter, Chinese): will the word-level and collocation level patterns found for English generalize to a typologically different language?

- Third (Exp 3), we examine whether NNSs of English across proficiency levels are also sensitive to the statistical information at different grain sizes of language, and if so, whether the level of sensitivity to the different distribution variables changes as language proficiency increases.

Across the three experiments, we compare and contrast the processing of collocations by NSs and by NNSs at different points in their L2 development trajectory in terms of their sensitivity to the different distribution variables, and determine whether NSs' and NNSs' processes are fundamentally different (holistic vs. analytical) or are they underpinned by the same factors and mechanisms. We will also consider the potential influence of L1 knowledge and its interaction with L2 proficiency.

To answer these questions, we adopted a double lexical decision task (LDT) similar to the one in Wolter and Yamashita (2015) for all experiments. In a classic double LDT, in each trial participants see two strings of letters simultaneously, one above the other, and are asked to press a YES key if both are correct spelled words and a NO key if at least one of them is not a word (Meyer & Schvaneveldt, 1971). The main finding of a classic LDT is if the two words are semantically related to each other (e.g., bread, butter), the RT to determine the lexical statuses of the two words is shorter than if the two words are unrelated (e.g., nurse, butter). This paradigm demonstrates the general cognitive principles of priming and spreading activation, whereby the processing of one word automatically activates words and concepts that relate to them and thus gives them a head start in processing. Although there is a wealth of research on semantic priming (for an overview, see e.g., McNamara, 2004), less is known about the priming between words that "belong together", as Wray (2002) described, in a statistical relation of forming a collocation. We chose the double LDT over the more commonly used phrasal acceptability judgment task in the hope to minimize the effect of strategizing and the use of explicit knowledge about the language grammar during the on-line processing of collocations, as a participant's attention is not being explicitly drawn to the fact that some of the word-pairs they see form collocations, allowing the capture of the automatic/implicit aspect of collocation processing.

The participants in our study are either NSs of English (Exp 1), NSs of Chinese (Exp 2), or NNSs of English (Exp 3). Since Chinese and English are unrelated, studying L1 Chinese speakers eliminates any confounds in processing due to the existence of cognates or extensive language borrowing, which might have been a problem in earlier studies (e.g., Wolter & Gyllstad, 2011, 2013 on English and Swedish; Wolter & Yamashita, 2015, 2018 on English and Japanese; Öksüz et al., 2021 on English and Turkish, etc.) Currently, there is very little research on collocation processing in English and in Chinese by L1 Chinese speakers learning English as an L2 varying in their language proficiency and dominance. The present study aims to fill this gap.

Chapter 2 Experiment 1

To recap, Exp 1 investigates L1 collocation processing in English NSs, aiming to replicate the previously observed processing advantage of collocations and identifying important word-level and collocation-level distribution variables that contribute to the collocation advantage in English NSs.

2.1 Methods

2.1.1 Items Development

To reduce uncontrollable variance, a single list of items was used in all three experiments. Because the items are kept consistent in all three experiments, the description of item development in this paragraph will not be repeated in the methods sections of Exp 2 and 3. With the same items, we created two language versions of the double LDT task, one with the instructions and items all in English (to be used in exp 1 and 3), and one with the instructions and items all in Simplified Chinese (to be used in Exp 2). The list of items was formulated with L1 Chinese speakers in mind – developed by the first author along with two Chinese-English bilingual research assistants. Critically, the word-pairs belong to one of the following conditions: 1) English collocations whose verbatim translations in Chinese do not form acceptable collocations in Chinese (aka, English-only, n=36, e.g., "bite bullet"); 2) Chinese collocations whose verbatim translations in English do not form acceptable collocations in English (aka, Chinese-only, n=36, e.g., "wash photo"); 3) word pairs that form acceptable collocations in both languages (aka, congruent, n=36, e.g., “leather jacket”); 4) unconnected word pairs in both languages (aka, baseline, n=108, e.g., “answer soup”), and 5) word pairs that contain one nonword (aka, nonce trials, n=216, e.g., “glude juice”)[3]. For the English version of the task to be used in Exp 1, the English-only and congruent items are coded as “collocations”, and the Chinese-only and baseline items are coded as “non-collocations”. For the Chinese version of the task to be used in Exp 2, the Chinese-only and congruent items are coded as “collocations’, and the English-only and baseline items are coded as “non-collocations”.

To verify the collocation status or the lack thereof, we first checked the collocation frequency of all the items in the Corpus of Contemporary American English (COCA; Davies, 2008-) using the “collocates” function[4]. Given the tremendous size of the corpus, most non-collocation items registered a small number of occurrences (e.g., the Chinese-only collocation “new hand” has 150 occurrences in COCA). To ensure that these occurrences are due to random idiosyncrasies and not systematic collocational patterns, we then administered a native speaker phrasal acceptability judgment Qualtrics survey for all the critical items[5]. Sixty-one (61) native speakers of English were recruited from Amazon Mturk and were asked to rate on a scale of 1-7 how often they use/encounter them in their everyday interaction with other native speakers and exposure to the language (1= Never or extremely rare; 7 = All the time or very often). They were encouraged to not overthink or analyze the grammar and only use their “native speaker instinct”. A total of 24 catch items were interspersed between test items to filter out low-quality data.

[3] All collocation words were checked with a lexical range breakdown using the VocabProfile program from the Compleat Lexical Tutor website (Cobb, accessed Apr. 2020) to ensure that they are frequent enough to be known to the ESL participants to be recruited for future experiments

[4] for a more in-depth description of how the items are developed, including the span for the collocation searches, the collection of the frequency data and the calculation of MI, refer to the Appendix, Section 1.2

[5] a determiner (the, a, or an) was inserted for most of the verb-noun items to make them grammatical (e.g., the item “break record” becomes “break the record” in the survey).

Participants (n=15) who did not rate “1” for all of the catch items were eliminated. As informed by the survey results, all of the items were kept in the experiment. The means and standard deviations (SDs) of the subjective ratings for items in each condition can be found in Table 1.

Because one of our goals is to examine the underlying factors that contribute to the ease of collocation processing, we included the following word-level and collocation-level variables: word frequency (lemma frequency of each of the two words in COCA), collocation frequency (lemma frequency of the collocation in COCA), Mutual Information (MI, calculated based on the word frequencies, collocation frequency, and the size of the corpus), and word length (letter count). Table 1 summarizes Exp 1 stimuli in the different conditions. The full stimuli list can be assessed from osf.io.

Table 1. English LDT stimuli and characteristics for Exp 1.

Condition		NSRatingAvg	FreqWord1	FreqWord2	FreqColloc	MI	LengthWord1	LengthWord 2
English-only	mean	6.09	80,410.22	129,124.19	850.72	4.98	5.06	5.58
	SD	0.99	59,868.98	125,502.08	1,048.44	2.58	1.19	1.75
Congruent	mean	6.32	201,260.72	151,371.33	2,239.81	5.03	4.94	5.42
	SD	0.83	413,436.89	139,481.41	3,562.56	3.10	1.33	1.54
Chinese-only	mean	1.39	273,293.25	151,482.97	48.75	-1.29	4.78	5.50
	SD	0.68	459,236.66	356,504.17	116.02	2.79	1.62	1.92
baseline	mean	1.11	184,988.06	143,992.83	22.71	-4.22	4.93	5.50
	SD	0.46	363,953.01	230,652.78	97.68	2.88	1.39	1.73

2.1.2 Participants

Forty-four (44) English monolingual NSs were recruited through the online research platform Prolific.co[6]. They were between the ages of 18 and 35, were United States citizens currently living in the United States, agreed to the statement "I only know English", and had not participated in the item norming survey. They were paid on the average rate of $9.50/hr for the LDT ($4.75 for 30 minutes) and $9.60/hr for an exit survey ($1.60 for 10 minutes) which collected information on their demographics and language background to further ensure that the participants were truly monolingual in English.

2.1.3 Procedure

Due to COVID-19 restrictions during the time of the experiment, data collection was fully remote. The experiment was programmed in PsychoPy3 and was published online to Pavlovia.org. Participants were able to access the Pavlovia.org link through the Prolific.co interface and complete the task on their own computer browsers. The PsychoPy script is available on osf.io[7].

The main experiment consisted of a double LDT in which participants were presented with two letter-strings one above the other in each trial. The instruction screen at the beginning of the program informed participants that their task was to judge as quickly as possible whether both of the letter strings are English words or not. Note that participants were not instructed to attend to the connection between the two words and were not made aware of the potential collocation status in some trials. Participants were asked to press "M" on their keyboard if both letter strings are words and "Z" if at least one of them is not a word. They could see the response coding on the screen for the entirety of the experiment. For each trial, the letter strings would remain on the

[6] 41 participants remained after eliminating abnormal and/or low effort participants. For more detail on outlier removal, refer to SM 2

[7] Accessed here: https://osf.io/9cba8/?view_only=2ae8f588db354ed2841528d323f8a8f5

screen until a response was made. The response would trigger an interim screen that asked participants to press the space key to advance to the next trial. Throughout the experiment, participants were able to see the number of trials left and were allowed to take a short break whenever they needed to before pressing the space key. After each incorrect response, they would see a warning screen that says "WRONG" which would cause a one-second delay before the space key can be pressed. This was to incentivize them to strive for accuracy so that they could complete the task in the shortest amount of time. RTs and Accuracy rates were logged by PsychoPy. The experiment began with 16 practice trials which included words and non-words that were not used in the 432 test trials described in Item Development. The trial order was individually randomized. Typically, the entire LDT task took 25-30 minutes for the English monolingual NSs.

2.2 Results

All data were analyzed with mixed-effects modeling using the "lme4" package (R package: lme4, RRID:SCR_015654, version 1.1-13, Bates et al., 2015) in R (version 3.3.3, R Core Team, 2020). The models were fitted by REML, with random effects specified for participants and items. To determine which random effects to include, specifically whether or not to include the random slopes for participants for each fixed-effect predictor, we ran two versions for each model, one with only participant random intercepts, and one further adding a participant random slope for the most important theoretical variables, i.e., collocation status and/or MI[8]. We report the model with only the random intercept for participants unless adding the random slope significantly improves model fit and does not introduce convergence issues, in which case we report the latter. This model

[8] For models including both collocation status and MI, we each ran three versions with participant random slopes: 1) for collocation status only, 2) for MI only, and 3) for both collocation status and MI, as well as a version without any random slopes. The model with the best fit is reported.

building and selection process applies to all the GLM analyses reported for all the experiments in this paper, and will not be described again in the individual models[9].

We first ran a linear mixed model with English collocation status, word 1 log frequency, word 2 log frequency, word 1 length, and word 2 length as fixed-effect predictors of the log Winsorized RT[10] with no interaction terms (Model EN-1, Appendix Table 6). In this model, English collocation status was defined as a categorical variable with two levels: English collocations (English-only and congruent items) and English non-collocations (Chinese-only and baseline items). All the other language variables were continuous: word frequencies were lemma frequencies (log-transformed) obtained from COCA, and word lengths were the number of letters. Participant random slope for collocation status was included. Both collocation status and word 1 log frequency had a significant effect on RT. RT was faster for collocations vs. non-collocations, and for items whose first words were more frequent. The Accuracy analysis (Model EN-1*, Appendix Table 7) further revealed a significant effect of word 1 length in addition to the expected facilitatory effect of word 1 frequency: intriguingly, Accuracy is higher when word 1 is longer.

To explore whether the word-level variables impacted collocations and non-collocations differently, we added the interactions between English collocation status and each of the language variables to the RT model[11]. To determine the appropriate interactions to include, we created interim models by adding one interaction at a time and conducted model comparisons using the anova() function in R to see if adding any single interaction would make the model significantly

---

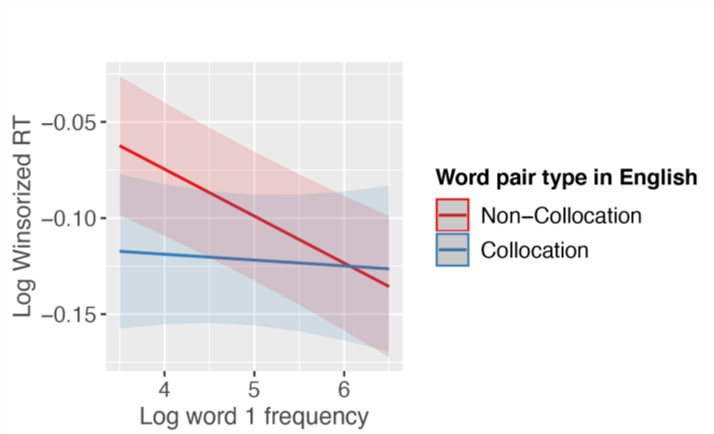
[9] We analyzed both the RT and the Accuracy of the LDT, but the ceiling effect of Accuracy rendered the Accuracy analyses less informative than the RT analyses. Thus, in the main text, we will focus on the output from RT analyses and only include selected Accuracy analyses. The full analyses on RT and Accuracy can be found in Appendix 4.

[10] Before analysis, we trimmed each participant's extreme RTs using the Winsorize function, whereby the smallest (<.05 percentile) and largest values (>.95 percentile) were replaced with the value at the .05 or .95 cut-off line, and log-adjusted all the values. After Winsorization, we filtered out the nonce trials, so the analyses only focused on critical items.

[11] The mirroring Accuracy analysis is not reported because adding the interaction terms did not improve model fit from Model EN-1* [χ 2(df = 4) = 5.700, p = 0.2229].

better at capturing the data than Model EN-1, aka the simple model. If any interim model provided a better fit than the simple model, it was then compared with the full model that included the interactions between collocation status and every language variable to see if there were additional appropriate interaction terms. Through this process, we found that the model with the interaction between English collocation status and word 1 frequency had the best fit (Model EN-1a, Appendix Table 8). Adding this interaction significantly improved model fit from the simple model EN-1 [χ 2(df = 1) = 5.222, p = 0.0223], while adding the rest of the interactions did not afford a better-fitting model than Model EN-1a [χ 2(df = 3) = 1.5068, p = 0.6807]. As shown in Figure 1, while both collocation status and word 1 frequency were still significant, word 1 frequency had a larger facilitation effect on non-collocation trials than collocation trials.

Figure 1. Effects of word pair type and log word 1 frequency on log Winsorized RT for English NSs



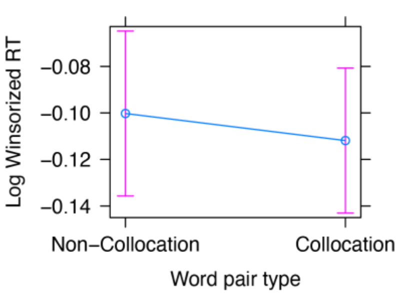
Next, to examine whether the participants' LDT performance is sensitive to the gradation of the collocation effect, i.e., whether a stronger (or a more frequent) collocation leads to an increasingly faster RT, we added the continuous predictor English MI score, i.e., a measurement of collocation association strength, to Model EN-1 (Model EN-2). Note that here we chose MI score over collocation frequency as the single collocation-level continuous predictor and did not include both due to the inevitable high inter-correlation between collocation frequency and word

1 frequency [r(8854) = .238, p = .000] and between collocation frequency and word 2 frequency [r(8854) = .314, p = .000]. The addition of MI score significantly improved model fit [χ 2(df = 1) = 6.1497, p = 0.01314]. Results from this model (Appendix Table 9) showed that there was indeed a facilitation effect of MI score: RT decreased as MI score increased. The facilitation effect of word 1 frequency remains the same as it was in Model EN-1. The main effect of the categorical collocation status is no longer significant. This is not surprising because the two variables are two sides of the same coin, i.e., a high MI score is what makes two words a collocation vs. a non-collocation, and thus the two factors would compete for effects in the same model. The fact that adding MI renders collocation status insignificant suggests that MI is the overriding predictor of RT. The mirroring Accuracy analysis (Model EN-2*) is in Appendix Table 10.

Following the same procedure for identifying appropriate interactions, we included an interaction term between word 1 frequency and MI (Model EN-2a, Appendix Table 11)[12]. As shown in Figure 2, the main effect of word 1 frequency and its interaction with MI remained the same pattern as they were in Model EN-1a: while word 1 frequency had an overall facilitation effect on RT, this effect becomes smaller as MI increases.

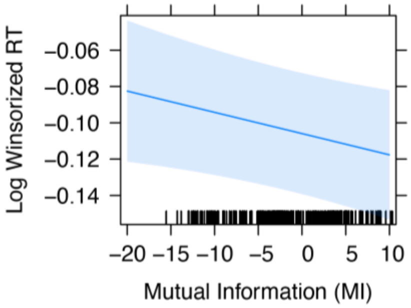
Figure 2.Effects of English language variables on log Winsorized RT for English NSs

a.) Word Pair Type

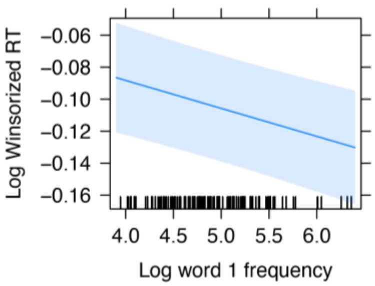


[12] The mirroring Accuracy analysis is not reported because adding the interaction terms did not improve model fit from Model EN-2* [χ 2(df = 4) = 3.982, p = 0.4084].

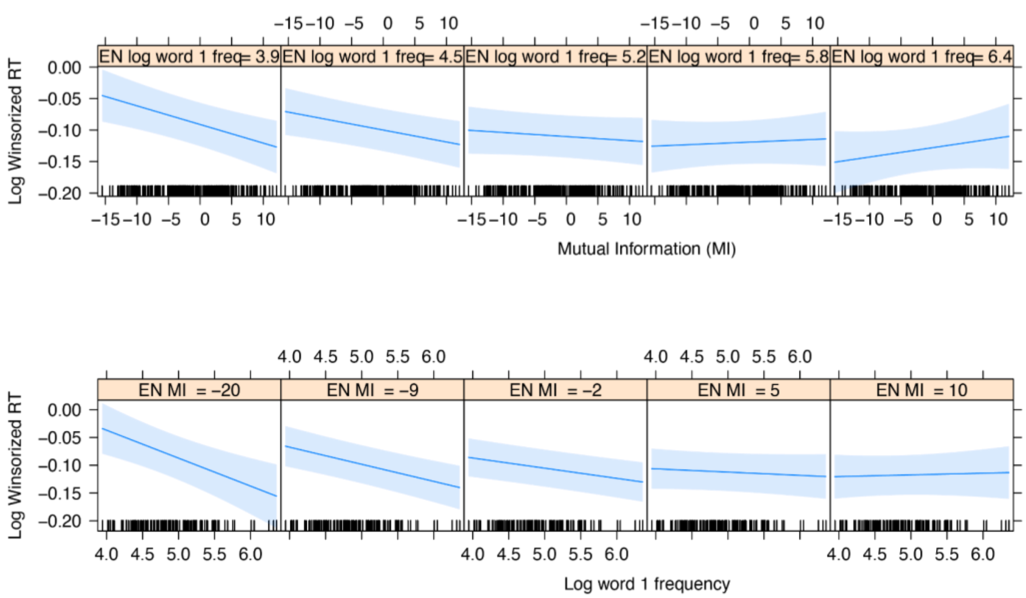
b.) MI



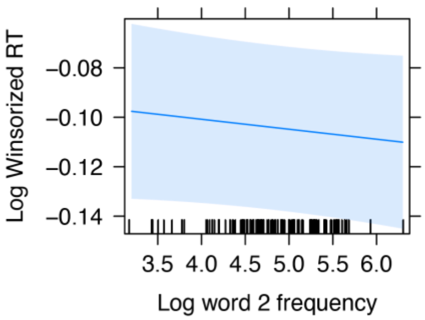
c.) Log word 1 frequency



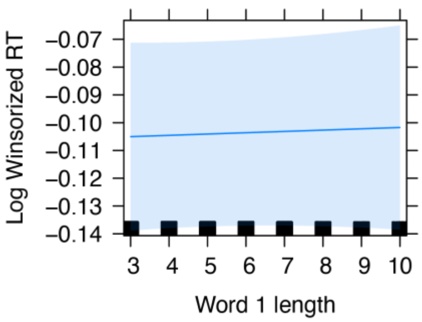
d.) MI: Log word 1 frequency



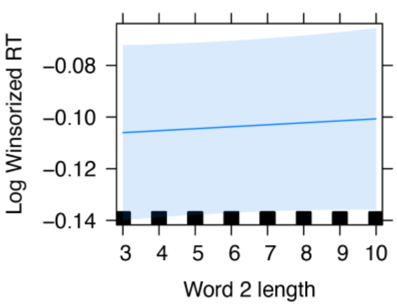
e.) Log word 2 frequency



f.) Word 1 length



g.) Word 2 length



2.3 Discussion

As expected, English NSs demonstrated a processing advantage of collocations over unconnected word pairs in an on-line processing setting, replicating the previously observed collocation facilitation effects in English NSs (e.g., Siyanova & Schmitt, 2008; Yamashita and Jiang, 2010; Vilkaité, 2016). Consistent with Yi (2018), our NS participants also showed graded sensitivity to the strength of association of the collocations as measured by MI.

Notably, word 1 frequency also had a facilitation effect on the LDT performance, suggesting that information associated with individual component words are at least partially activated during the processing of the word pairs. Additionally, its interaction with collocation-level variables such as collocation status and MI suggests that as the two words are more strongly paired, the effect of the frequencies of individual words becomes weaker. Therefore, consistent with Öksüz et al’s (2021) observation, for English NSs, the effect of MI is more prominent than the effect of word frequency for highly collocational items.

Chapter 3 Experiment 2

Exp 2 investigates L1 collocation processing of Chinese and aims to extend the findings of Exp 1 to a language in which the collocation phenomenon is vastly understudied. To our knowledge, there has not been a single study examining Chinese NSs' processing of Chinese collocations - a gap that Exp 2 aims to address.

3.1 Methods

The items used in the Chinese LDT were verbatim translations of the English items used in Exp 1. The translations were cross-checked by three Chinese-English bilingual speakers. The item verification process was largely the same as Exp 1. We conducted a phrasal acceptability judgment Qualtric survey with the items and instructions all in Chinese and recruited 32 Chinese native speakers through word-of-mouth to complete the survey. Five (5) participants who did not answer "1" for all the catch items were excluded. No item was eliminated through this norming process. Note that for the Chinese stimuli, Chinese-only and congruent items were coded as "collocations," and English-only and baseline items as "non-collocations." The collocation and word frequencies were gathered from the Chinese Web 2017 Simplified Corpus (zhTenTen17; Jakubíček et al., 2013) from Sketch Engine (for more information, see Kilgarriff et al., 2014). Word length in Chinese was operationalized as stroke count rather than letter/character count due to the non-linear characteristics of the orthography. A stroke is a line (straight or curved; horizontal, vertical, or diagonal) completed each time the writing instrument is lifted from the written surface. Since Chinese does not have letters, and since most Chinese words contain either one or two characters, the number of characters would not have been sufficiently informative. Stroke count has been found to predict the ease of processing of written Chinese: characters with more strokes are processed more slowly. These effects are robust and are similar to the effect of letter count in alphabetic languages (see Sze et al., 2015 for a review on lexical variables in Chinese visual processing). Table 2 summarizes the characteristics of the stimuli used in Exp 2.

Table 2. Chinese LDT stimuli and characteristics for Exp 2.

Condition		NSRatingAvg	FreqWord1	FreqWord2	FreqColloc	MI	LengthWord1	LengthWord 2
Chinese-only	mean	6.06	4,019,072.14	1,603,214.72	56,435.61	6.46	9.5	12.83
	SD	0.69	9,045,074.08	3,142,405.77	84,686.85	3.18	4.48	5.04
Congruent	mean	6.18	2,002,282.50	2,177,146.14	83,032.61	7.03	12.44	13.44
	SD	0.87	3,508,699.52	3,142,405.77	181,367.66	3.56	5.26	6.32
English-only	mean	1.85	801,242.89	1,527,628.22	279.64	-0.69	13.78	15.36
	SD	1.12	1,637,906.29	2,195,308.68	548.69	2.59	5.88	4.42
baseline	mean	1.37	2,092,651.63	1,769,329.69	201.29	-2.77	11.91	13.88
	SD	0.39	5,337,557.15	2,966,019.07	603.53	2.67	5.49	5.38

3.1.1 *Participants and procedure*

173 Chinese NSs who did not participate in the norming survey were recruited through either Prolific.co, word-of-mouth, online posters distributed by the fourth author to students at her university in China, or through a Chinese colleague whose students participated in the experiment as a class assignment[13]. The background demographics will be explained in-depth in Exp 3, in

[13] 163 out of 173 participants remained after eliminating abnormal and/or low effort participants. For more detail on outlier removal, refer to SM 2

which the same participants participate. The experiment procedure was the same as Exp 1, except that the Chinese version of the LDT program was used. Participants typically finished the task in 25-30 minutes and were paid $10.28/hr for their participation.

3.2 Results

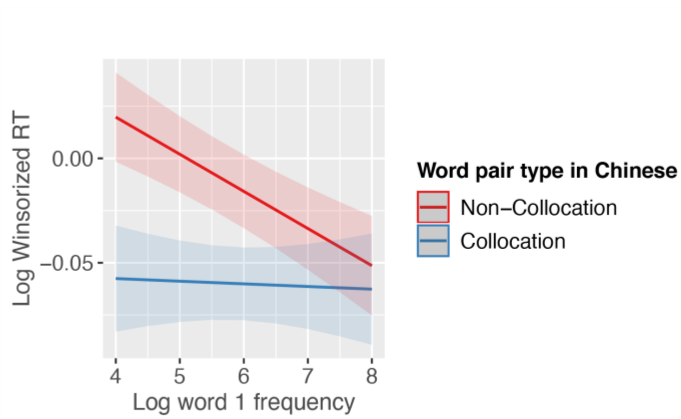
The analyses of Exp 2 mirrored those in Exp 1. After Winsorization and log-adjustment of the RTs, we first ran a model (Model CH-1, Appendix Table 12) that included the fixed-effects of word 1 log frequency, word 2 log frequency, word 1 stroke count, word 2 stroke count, and Chinese collocation status (Chinese collocation, i.e., Chinese-only and congruent items, and Chinese non-collocations, i.e., English-only and baseline items) with no interactions[14]. As per the model selection process outlined in the Results for Exp 1, no random slopes were included in this model. Word frequencies were lemma frequencies (log-transformed) obtained from zhTenTen17. Total stroke counts of each of the words are obtained from an online automated stroke counter[15] and were verified by a native Chinese speaker. Model CH-1 showed that Chinese collocation status was a significant predictor of RT: RT was faster for collocations than non-collocations. All other language variables were also significant. RT decreases as word 1 and word 2 frequency increases and as word 1 and word 2 stroke count decreases. The mirroring Accuracy analysis in Appendix Table 13.

[14] For Exp 2 analyses, we included an ad hoc length pattern variable to adjust for the unique prosodic preference effect in Chinese collocation processing. Most Chinese words are monosyllabic or bisyllabic, and thus any two-word phrases can have four length patterns: 1+2, 2+2, 2+1, 1+1, where “1” stands for a monosyllabic word and “2” stands for a bisyllabic word. These length patterns are found to impact the acceptability of the phrases due to preferred prosodic cadences in the language: namely, verb-noun phrases with the 2+1 pattern and modifier-noun phrases with the 1+2 pattern are less prosodically preferred than other patterns and are thus considered less acceptable as collocations (Duanmu, 2012; Qin & Duanmu, 2017; Duanmu et al., 2018). This length pattern variable turned out to have a significant main effect on RT measures in all Exp 2 models whereby RT is longer for unpreferred patterns, while it did not interact with or change the pattern of significance of any other variables in the model. But because this variable was not of central concern of this study and its inclusion did not change the main conclusions, the set of analyses involving it is not reported or discussed here. The full analysis output will be available upon request.

[15] Accessed from: http://www.gaoshukai.com/lab/0031/en.html

To explore whether the language variables impacted collocations and non-collocations differently, we used the same procedure for the analyses in Exp 1 and similarly found that for the RT analysis, the model with the interaction between Chinese collocation status and word 1 frequency had the best fit (Model CH-1a, Appendix Table 14). The results show that in addition to the main effects of collocation status and all the language variables, the interaction between collocation status and word 1 frequency was also significant, replicating the pattern of the English LDT data (Model EN-1a). In both Chinese and English NSs' LDT performance, word 1 frequency had a lesser facilitation effect on the RTs for collocations than on the RTs for non-collocations (Figure 3). The mirroring Accuracy analysis in Appendix Table 15.

Figure 3. Effects of word pair type and log word 1 frequency on log Winsorized RT for Chinese NSs



Similar to the English LDT analyses, we added the corpus-derived Chinese MI scores as an additional predictor to gauge the gradation of the collocation effect (Model CH-2, Appendix Table 16)[16]. As expected, in addition to the main effect of collocation status, MI also positively predicted RT. The main effects of word 1 frequency, word 2 frequency, word 1 stroke count, and

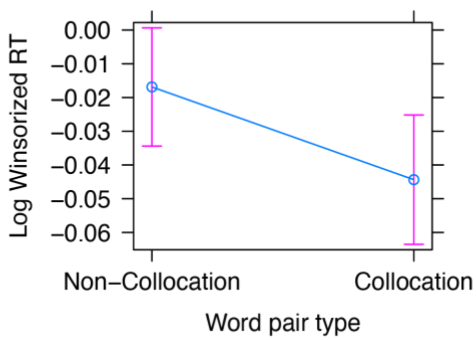
[16] Again, collocation frequency was not included because of the same multicollinearity issues: collocation frequency was correlated with word 1 frequency [r(35669) = .398, p = .000] and word 2 frequency [r(35669) = .327, p = .000].

word 2 stroke count all remained significant in the expected direction. The mirroring Accuracy analysis in Appendix Table 17.

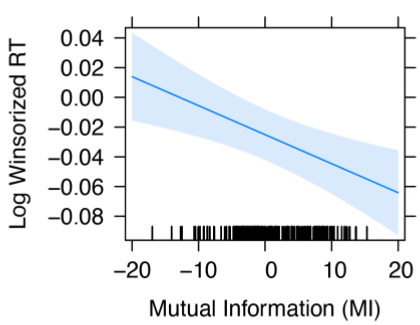
Following the same procedure for identifying interactions, we found an interaction between MI and word 1 frequency, which was marginal in the RT analysis (Model CH-2a, Appendix Table 18) and significant in the Accuracy analysis (Model CH-2a*, Appendix Table 19). As shown in Figure 4, the direction of the interaction replicated the pattern found in the English L1 data: the facilitation effect of word 1 frequency diminishes as MI increases for both RT and Accuracy. The main effects of all the language variables remained significant. Additionally, Accuracy analysis also revealed an interaction between MI and word 2 stroke count: the facilitation effect of MI on Accuracy becomes larger when the second word is more complex (see Appendix Figure 7, h).

Figure 4. Effects of Chinese language variables on log Winsorized RT for Chinese NSs.

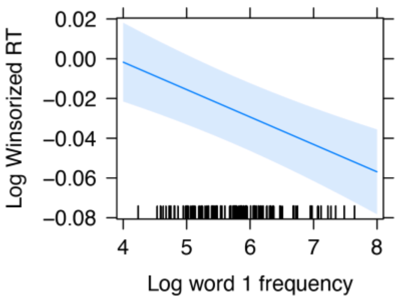
a.) Word Pair Type



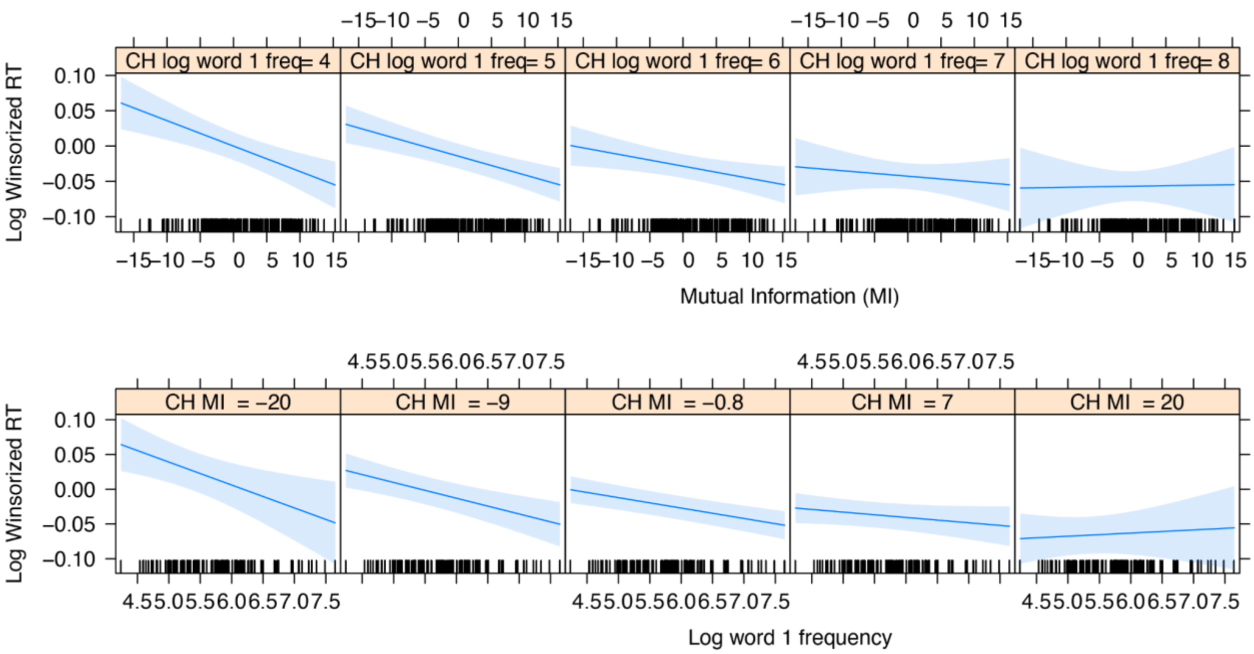
b.) MI



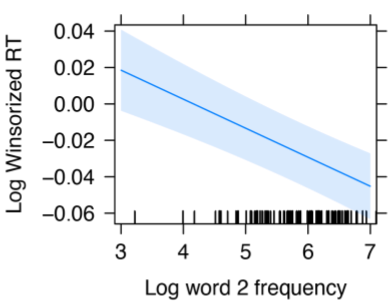
c.) Log word 1 frequency



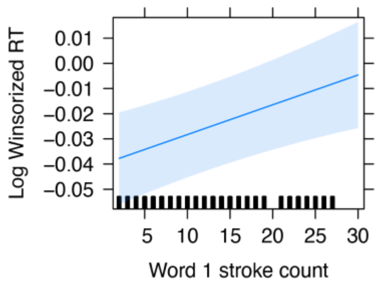
d.) MI: Log word 1 frequency



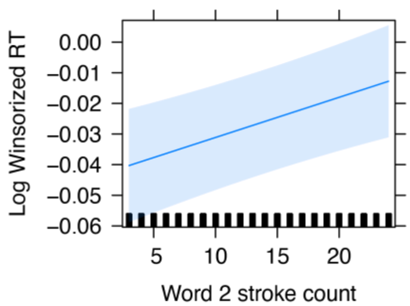
e.) Log word 2 frequency



f.) Word 1 length



g.) Word 2 length



3.3 Discussion

The results from Exp. 2 successfully extended the findings on English NSs to NSs of the typologically distinct Chinese language. Chinese NSs not only demonstrated the collocation advantage, but also the graded sensitivity to collocation association strength, lending support to the universality of the psychological reality of collocations. Like their English-speaking counterparts, the Chinese NSs are also less affected by word 1 frequency as the word-pairs increase in their collocation association strength. Interestingly however, the Chinese NSs demonstrated sensitivity to all the word-level variables - both the frequencies and the lengths (i.e., stroke counts) of both word 1 and word 2, and not just word 1 frequency observed in English NS. Furthermore, except for word 1 frequency, these word-level variables uniformly impacted all word pairs regardless of collocation strength, demonstrated by the lack of interactions. This sensitivity to word-level variables could indicate that the processing of collocations is subjected to an interim mechanism by which the information associated with individual component words is retrieved. We will further discuss this in the general discussion.

Chapter 4 Experiment 3

Exp 3 investigates L2 collocation processing for NNSs of English. We determine whether the collocation advantage observed in NSs still holds true for NNSs across various levels of proficiency and language dominance and if NNSs are also sensitive to the statistical information at the word-level and collocation-level. We also examine if the level of sensitivity at the different grain-sizes of language is modulated by participants' language experience, i.e., whether the contribution of different distribution variables changes as language proficiency and dominance increases. Additionally, we compare and contrast the processing of collocations by NSs and NNSs, taking into consideration the potential influence of L1 knowledge and its interaction with L2 proficiency and dominance.

4.1 Methods

4.1.1 Participants and procedure

We recruited the same 173 Chinese NSs who participated in Exp 2 to participate in Exp 3 because they were all sequential bilinguals speaking Mandarin Chinese as their L1 and learning English as their L2[17]. They all completed the Chinese-English Bilingual Language Profile (BLP) administered in Chinese (translated from English by the first author) along with the experiment. The survey is deposited at osf.io. The BLP is an instrument for assessing language dominance

[17] The order in which the participants completed Exp 2 and Exp 3 was counterbalanced. Participants were instructed to complete both tasks in the set order within a week's time, and we checked the completion timestamps for each participant to ensure that the correct order was followed. To control for the order effect, for models in Exp 3, we ran a version with order as an additional fixed effect factor. Because the order effect was not significant in any models, we are reporting the models without this factor.

through self-reports (Birdsong, 2012). By taking into account a variety of linguistic variables in four broad aspects: language history, use, proficiency, and attitudes, the BLP produces a general bilingual profile and a continuous dominance score for a bilingual's two languages, and in this case, Chinese and English. We are using language dominance in lieu of proficiency because dominance is a more comprehensive measure of language background which proficiency is a component of. The task was the same English version of the LDT used in Exp 1. The NNSs took 25-40 minutes to complete the LDT.

The participants naturally stratify into 5 language groups according to how they were recruited. First, the highly advanced ESL participants (ADV+, n=19) were those recruited from Prolific.co, all of whom were born in China and living in the United States at the time of recruitment. Due to the nature of the Prolific.co subject pool, many Chinese-English bilinguals recruited there are more dominant in English than in Chinese. To exclude simultaneous bilinguals (English and Chinese were both L1s), as well as sequential bilinguals who have a low level of Chinese literacy and whose Chinese has been greatly attrited (self-report Chinese reading proficiency ≤ 2), we used selected questions in the language history and proficiency categories of the BLP as prescreening criteria. Potential participants completed the BLP as the first step and were paid an average rate of $9.50/hr. Only those who fit the criteria were invited to participate in the experiment and were paid $10.28/hr for the LDT. The second group was the advanced ESL group (ADV, n=42), consisting of Chinese International undergraduate or graduate students studying in the United States or Canada, recruited through word-of-mouth and online advertisements. Third, the intermediate ESL participants (INT, n=77) were Chinese undergraduate students recruited from a top-tier university in China. Fourth, there were two beginner groups both consisting of students from a technical school in China who participated in the experiment for course credits of their English language class. Prior to participation, these students were given an Oxford placement test to gauge their English background. Those who scored 27 and above formed the improving beginner ESL group (BEG+, n=68), and those who scored below 27 formed the true beginner ESL group (BEG, n=50). Except for the ADV+ prolific participants who completed the BLP before the experiment, everyone else completed the BLP as an exit survey. The BLP survey took around seven minutes to complete. Table 3 shows how the BLP dominance scores were distributed across the five pre-defined groups. The full summary of the BLP survey is in Appendix 3.

Table 3. Number of participants in each group and mean English and Chinese dominance scores for each group across different criteria (max dominance score = 218).

NNS Groups	# of Participants[18]	Chinese dominance score		English dominance score	
		Mean	SD	Mean	SD
ADV+	17	162.76	23.09	137.34	22.96
ADV	33	188.85	13.05	101.94	26.02
INT	44	201.28	11.38	60.98	26.99
BEG+	38	198.19	10.90	57.91	19.81
BEG	17	200.17	9.83	53.47	24.06

4.2 Results

We first ran a simple linear mixed-effects model with English collocation status, word 1 log frequency, word 2 log frequency, word 1 length, word 2 length, and the English Dominance Score from the BLP as fixed-effect predictors of the log Winsorized RT (Model ESL-1, Appendix Table 20). No random slope was included. Model ESL-1 showed a significant effect of English

[18]149 out of 173 participants remained after eliminating abnormal and/or low effort participants. For more detail on outlier removal, refer to SM 2

collocation status for NNSs. Similar to the English NSs in Exp 1, NNSs processed English collocation items (English-only and congruent items) faster than English non-collocation items (Chinese-only and baseline items). But unlike the English NSs, where the only significant word-level predictor was word 1 frequency, the NNSs' RTs were significantly impacted by both the word 1 and word 2 frequencies and lengths. RT was faster as the frequencies of the two words increased and as the lengths of the two words decreased. Unsurprisingly, English dominance also had a significant effect: RT was faster as English dominance increased.

Similar to the analyses in previous experiments, we added the English MI score to Model ESL-1 to see if the collocation effect is graded. However, this new model rendered both collocation status and MI score insignificant. Since in the English NS data, MI score was found to be the overriding factor, we decided to replace collocation status with MI rather than including both in the same model so that they would not compete for the same effects[19]. In other words, this model (Model ESL-2, Appendix Table 21) included English MI, word 1 log frequency, word 2 log frequency, word 1 length, word 2 lengths, and English Dominance Score as the fixed-effect predictors. All the effects observed in Model ESL-1 remained. The facilitation effect of MI score also emerged: RT was faster as the MI score of the word pairs increased.

To see if MI affects RT differently depending on the characteristics of the words in the word-pairs, we ran another model (Model ESL-2a, Appendix Table 22) with two-way interactions between MI and all the word-level language variables, i.e., word 1 and word 2 length, and word 1 and word 2 log frequency, as well as between MI and dominance. In this model, the main effect of MI is no longer significant; however, the effect emerges as English dominance goes up, as indicated by the significant MI:dominance interaction. The main effects of word 1 and word 2

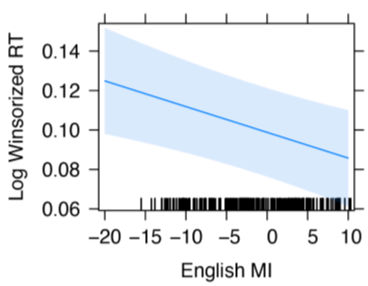
[19] This is justified in model comparison using the anova() function in R, i.e., the full model with both collocation status and MI score did not have a better fit than the reported model that included MI score only [χ 2(df = 1) = 0.2173, p = 0.6411].

length, as well as word 1 and word 2 frequency, all remained as expected. Interestingly, there was a marginal interaction between MI and word 1 length: the facilitatory effect of MI was much greater when word 1 was shorter.

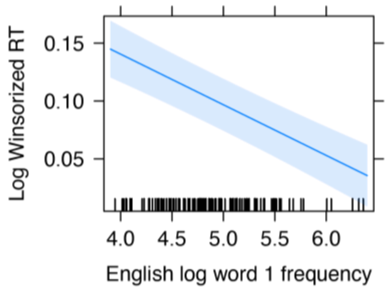
Next, to examine whether the effects of MI and other language variables depend on English dominance, we further included in the model two-way interactions between dominance and all the language variables (Model ESL-2b, Appendix Table 23). In addition to the significant effects in Model ESL-2a, Model ESL-2b revealed significant interactions between English dominance and English MI: the facilitatory effect of MI on RT becomes larger as dominance increases. Dominance also significantly interacted with word 2 length, word 1 frequency, and word 2 frequency. As dominance grows, word 2 length becomes less inhibitory, and word 1 and word 2 frequency become less facilitatory. This overall pattern shows that as a learner becomes increasingly dominant in English, their RT is more and more impacted by MI, and less and less impacted by individual word level variables. These effects are illustrated in Figure 5. The mirroring Accuracy analysis in Appendix Table 24.

Figure 5. Effects of English language variables and BLP English Dominance on log Winsorized RT for English NNSs.

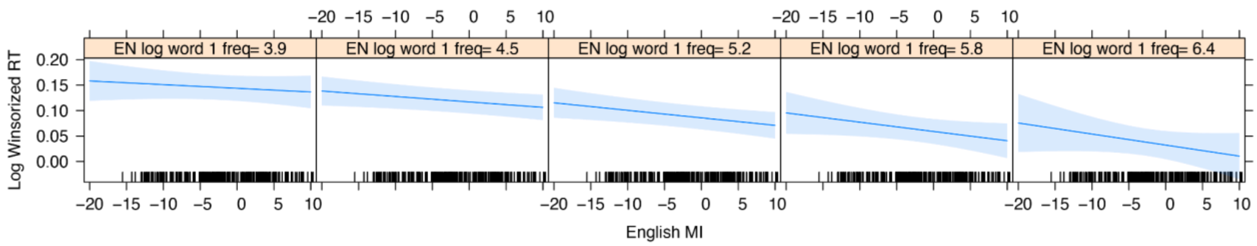
a.) English MI



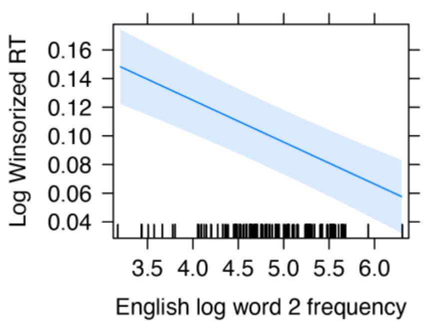
b.) Log word 1 frequency



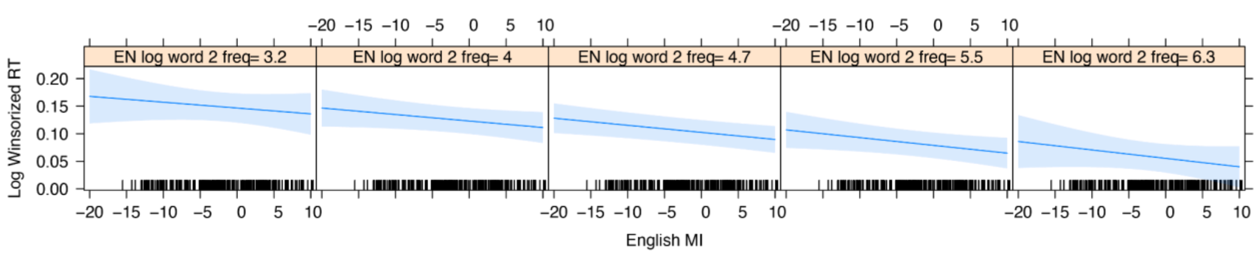
c.) MI : log word 1 frequency



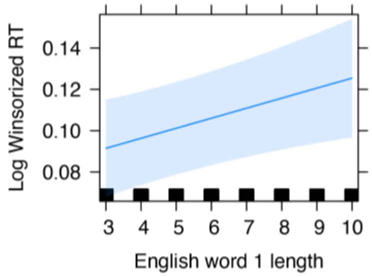
d.) Log word 2 frequency



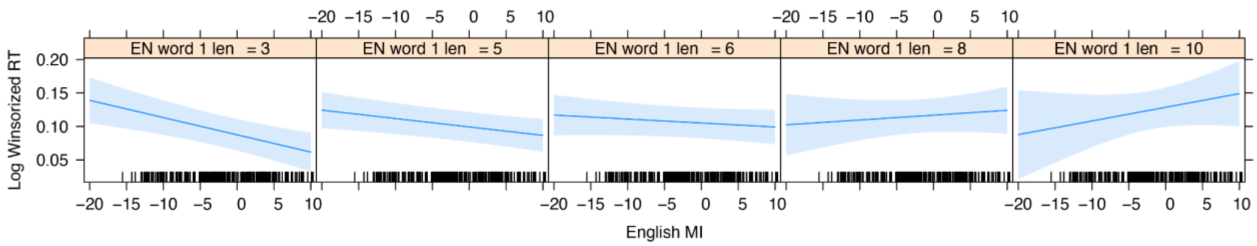
e.) MI : log word 2 frequency



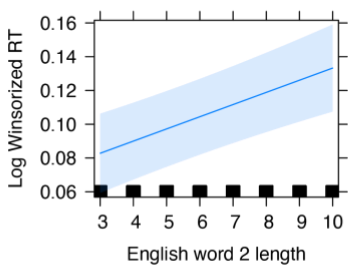
f.) Word 1 length



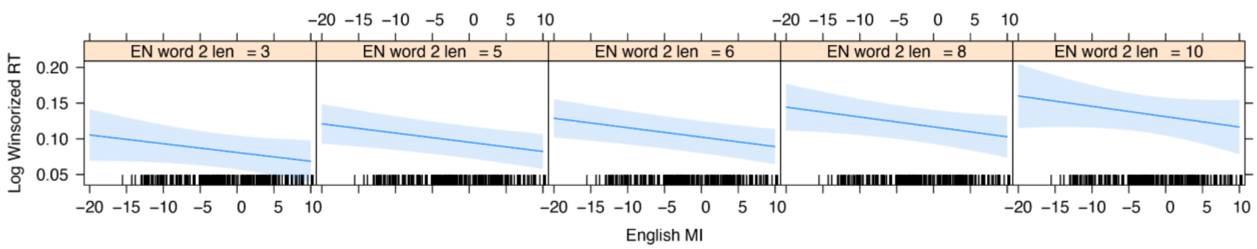
g.) MI : word 1 length



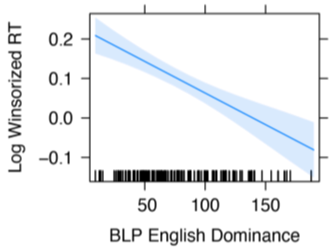
h.) Word 2 length



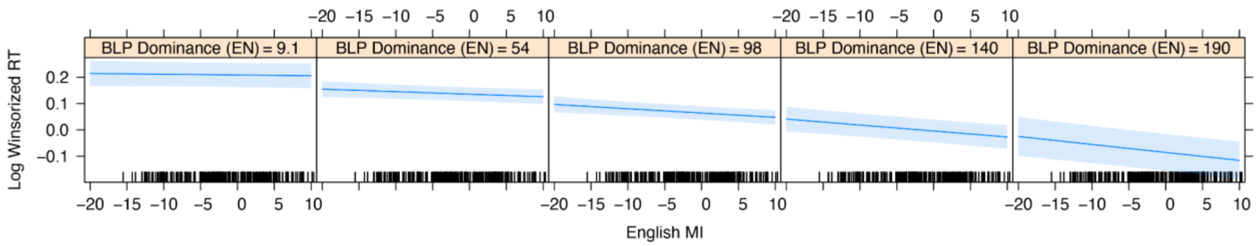
i.) MI : word 2 length



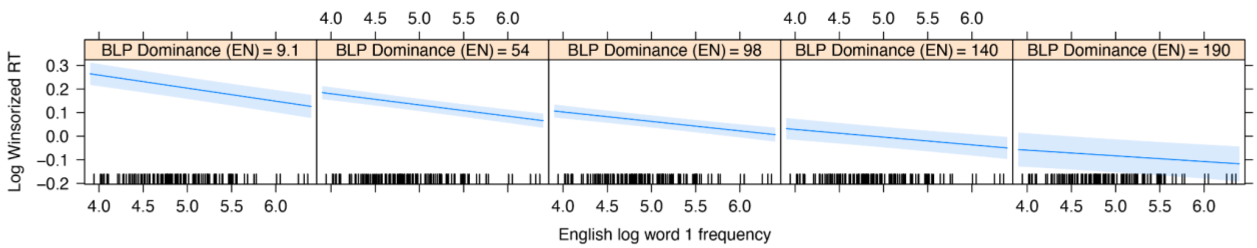
j.) BLP English Dominance



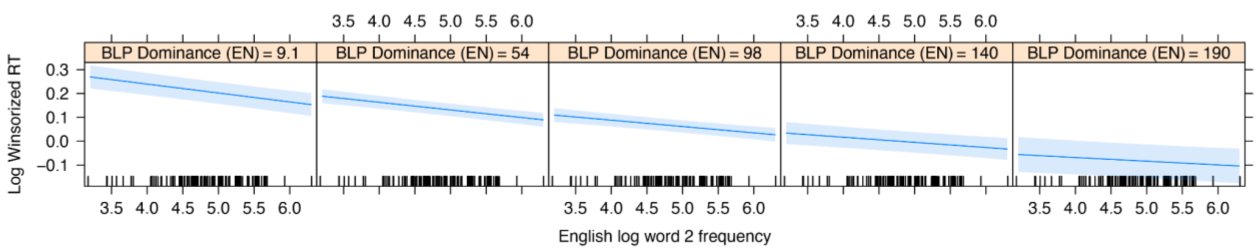
k.) BLP English Dominance : MI



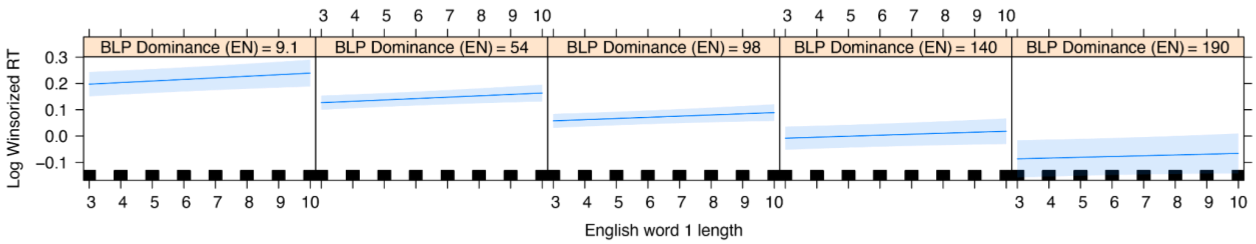
l.) BLP English Dominance : log word 1 frequency



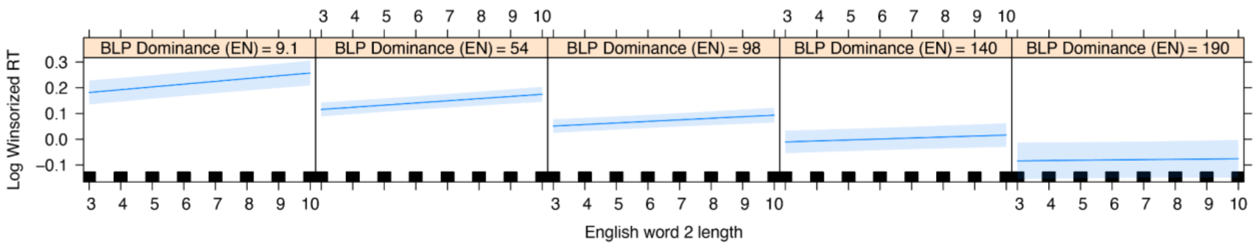
m.) BLP English Dominance : log word 2 frequency



n.) BLP English Dominance : word 1 length



o.) BLP English Dominance : word 2 length



4.2.1 Investigating L1 Influence

To examine L1 influence, we included in the analyses the English NSs' data collected from Exp 1 as a baseline control to compare to the NNSs' data. Because of this addition, we were no longer able to use the BLP Dominance variable, as it did not apply to the English monolingual NSs. Thus, we replaced the English Dominance score with the predetermined grouping factor to gauge English language background. The five language groups of NNSs (ADV+, ADV, INT, BEG+, and BEG) were the ones described in the Participants and procedure section, adding another level for monolingual English NSs. To see if the RT in the English LDT was impacted by the NNSs' knowledge and experience of Chinese, we included in our linear mixed-effects model all the language variables obtained from both the English corpus and the Chinese corpus, along with their interactions with language groups. We also introduced the 4-level categorical variable word-pair type (i.e., Congruent, Chinese-only, English-only, or baseline items) and its interaction with language groups[20] (Model ESL-3, Appendix Table 25). In other words, besides the random effects, Model ESL-3 includes the MI of the collocation in English (English MI), word 1 and word 2 frequencies and lengths in English (English word 1 frequency, English word 2 frequency, English word 1 length, English word 2 length), the MI of the translation of the collocation in Chinese (Chinese MI), the frequencies and stroke counts of the translations of word 1 and word 2 in Chinese (Chinese word 1 frequency, Chinese word 2 frequency, Chinese word 1 stroke count, Chinese word 2 stroke count), word-pair type, the language group variable and its two-way interaction with all the above-mentioned variables. No random slopes were included.

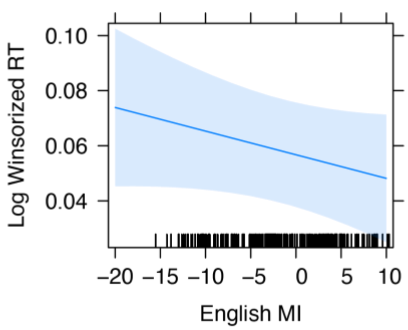
Model ESL-3 output showed a marginal facilitation effect of English MI and a significant effect of English word 1 frequency in the expected direction. As shown in Figure 6, the NNS groups were all significantly different from the NS group, with the RT incrementally decreasing as language experience increased. There was a small interaction between Group and English MI: compared to NSs, NNSs showed numerically smaller facilitation effects of English MI; this interaction was marginally significant for the BEG+ group. Group also interacted with English word 1 and word 2 lengths: compared to NSs, the INT and BEG NNSs were significantly more impacted by English word 1 length, and the ADV, INT, BEG+, and BEG NNSs were significantly more impacted by English word 2 length. The interactions between Group and English word 1 and word 2 frequencies were also significant: compared to NSs, the INT, BEG+, and BEG NNSs were increasingly more impacted by English word 1 frequency, and the ADV, INT, BEG+, and BEG NNSs were increasingly more impacted by English word 2 frequency. Importantly, there were interactions between groups and two Chinese language variables: Chinese MI and Chinese word

[20] Since previous analyses revealed that MI and collocation status tended to have mirroring effects, and that MI was the more impactful factor, we did not include both MI and collocation status in the same model and we only report the model with MI here.

2 frequency. Chinese MI had a facilitation effect on RT for the BEG+ and BEG NNSs, compared to NSs. Chinese word 2 frequency had a significant facilitation effect on RT for the INT and BEG NNSs and a marginal facilitation effect for the ADV and BEG+ NNSs, compared to NSs. Additionally, there was an interaction between congruence and group: compared to NSs, BEG+ and BEG NNSs had slower RTs for congruent items compared to items in the other conditions. None of the other main effects or interactions in this model were significant[21]. The mirroring Accuracy analysis (Appendix Table 26, Figure 8) revealed an additional interaction between conditions and group: compared to NSs, BEG and BEG+ NNSs had lower accuracy for Chinese-only items than other items.

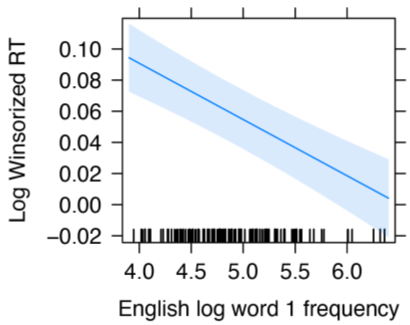
Figure 6. Effects of English and Chinese language variables, Group, and Condition on log Winsorized RT for English NNSs.

a.) English MI

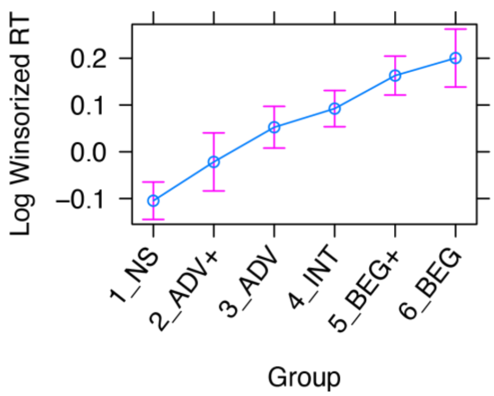


[21] Due to the concern that the transfer effect may depend on the running order (i.e., participants who completed the Chinese task in Exp 2 before completing the English task in Exp 3 may experience heightened Chinese-to-English transfer), we ran an additional analysis that included the interactions between order and all the Chinese language variables and between order and condition. None of the interactions were significant, i.e., the effect of transfer did not depend on the order. The inclusion of interaction terms did not provide better fit [χ 2(df = 8) = 7.268, p = 0.508], and thus this model is not reported.

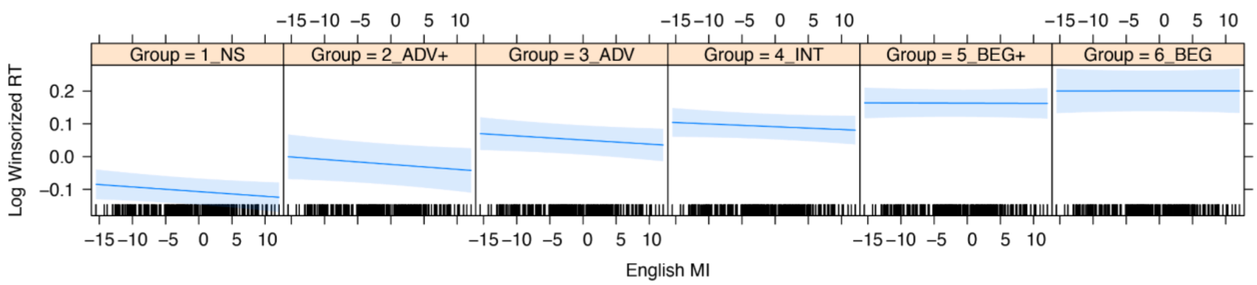
b.) English log word 1 frequency



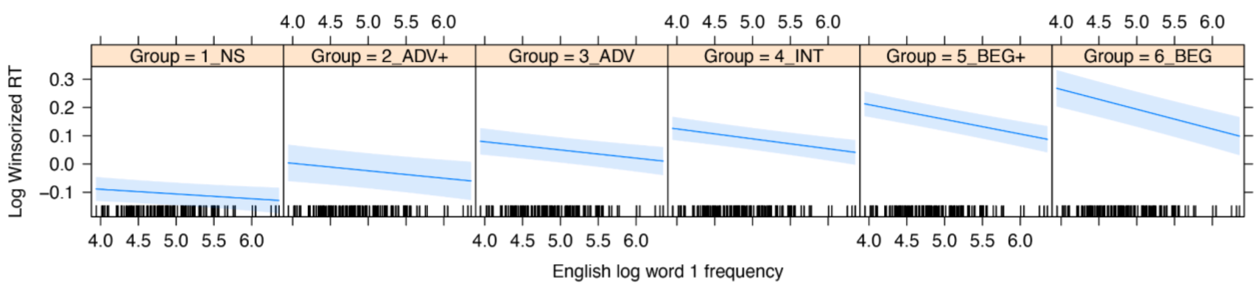
c.) Group



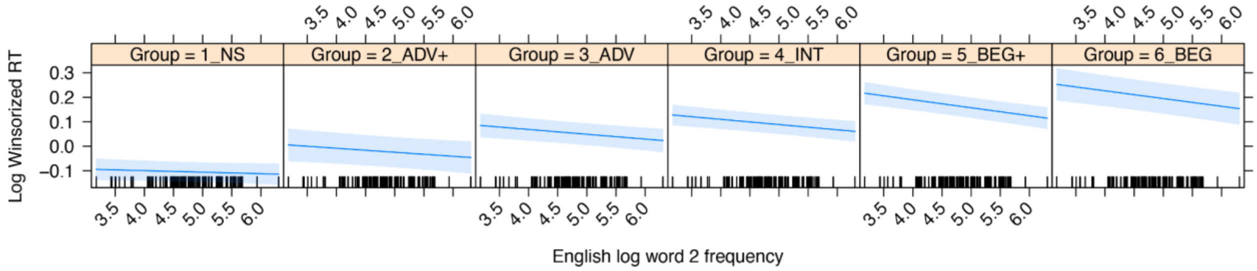
d.) English MI : Group



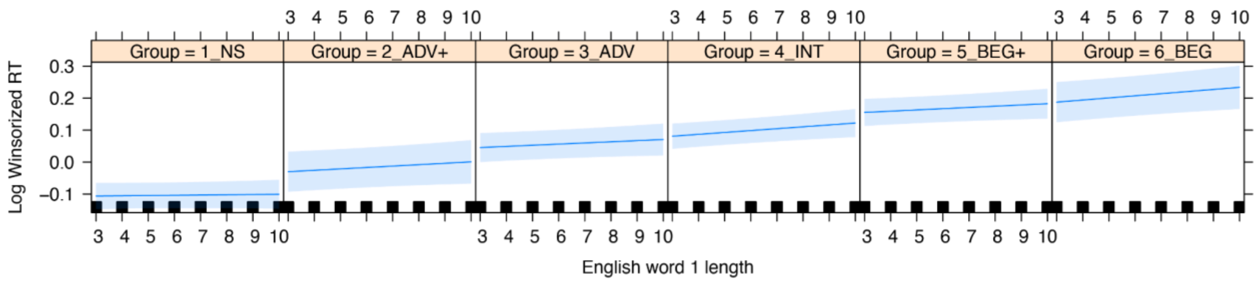
e.) English log word 1 frequency : Group



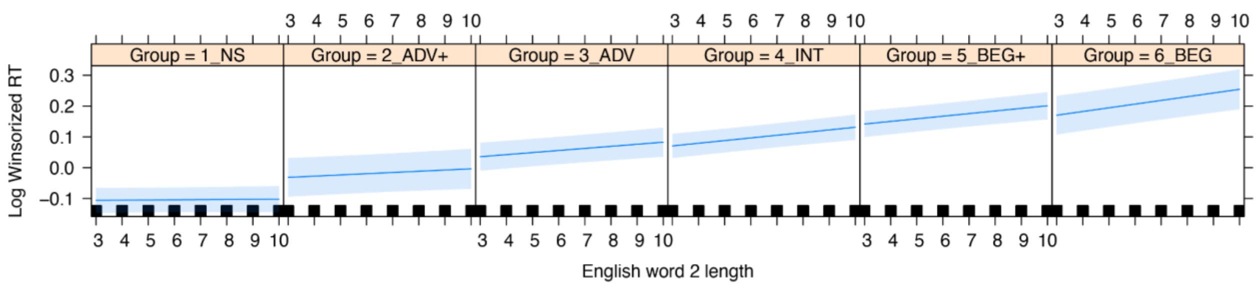
f.) English log word 2 frequency : Group



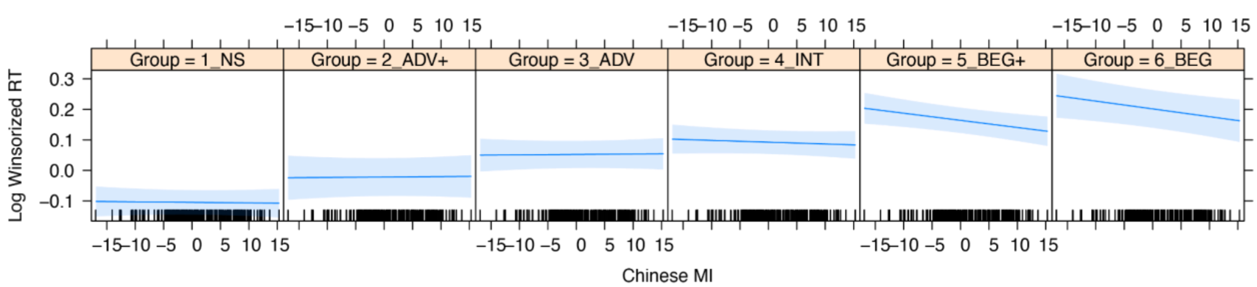
g.) English word 1 length : Group



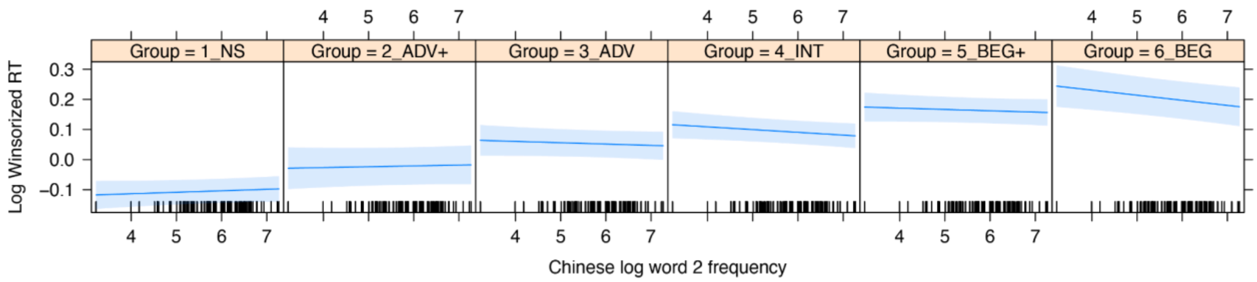
h.) English word 2 length : Group



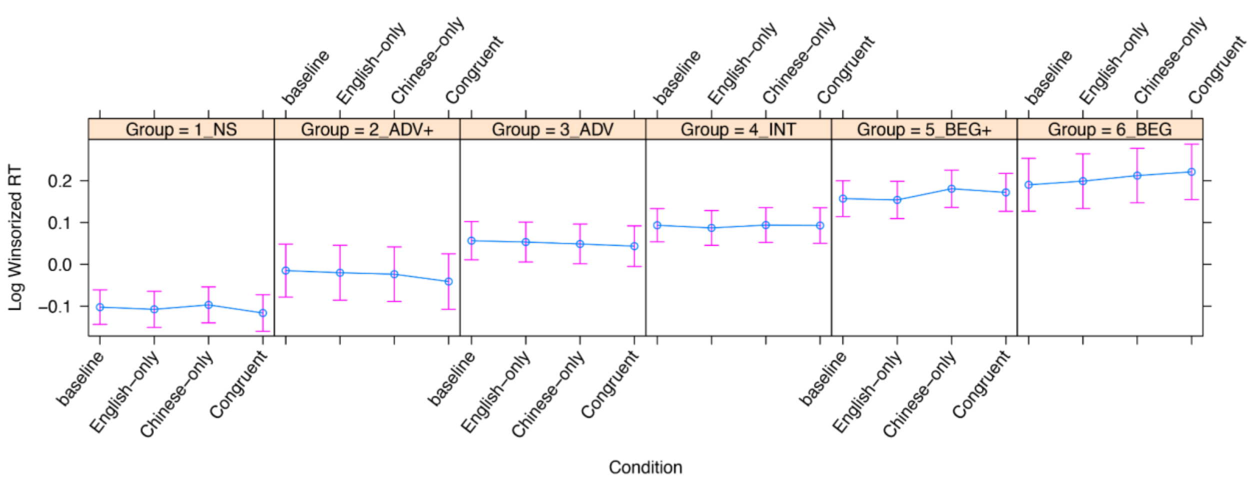
i.) Chinese MI : Group



j.) Chinese log word 2 frequency : Group



k.) Condition : Group



4.3 Discussion

Exp. 3 investigated L2 collocation processing of English NNSs across different levels of English dominance. We found that similar to NSs, NNSs were sensitive to the same distributional information at different grain sizes of language and that these effects were modulated by dominance. Unlike the English NSs in Exp.1, the NNSs' performance on the English LDT was significantly impacted by the frequencies and lengths of both word 1 and word 2, which coincides with the pattern observed in their performance in the Chinese LDT as NSs in Exp. 2. Notably, collocation association strength did not have a significant main effect across all participants; however, this collocation-level effect emerges as dominance increases. This pattern is consistent with Wolter and Yamashita's (2018) finding that collocation-level effects are more pronounced for NSs and high proficiency NNSs than for low proficiency NNSs. In addition, the interactions between dominance and the word-level and collocation-level variables suggest a shift-away from reliance on word-level variables to collocation-level variables which occurs with progressions in dominance. With repeated use and exposure as reflected in language proficiency and dominance, the facilitatory effect of collocation association strength becomes larger, while the inhibitory effect of word lengths and the facilitatory effect of word frequencies become smaller. While Wolter and Yamashita did not find the same interaction between word frequency and proficiency in their NNSs, it does not contradict our findings since both proficiency groups in their study were relatively high in proficiency (graduate vs. undergraduate International students in the same English-speaking institution), and that the two groups might not have differed enough in proficiency and other factors subsumed under dominance for the detection of these graded word-level effects.

Notably, we found evidence of L1 influence as indicated by the effects of Chinese language variables, namely, Chinese MI and word 2 frequency, on the processing of the English collocations by the bilingual participants. The interactions between these Chinese language variables and language groups suggest that the beginner NNSs likely have activated more L1 knowledge while processing the collocations in their L2 than did more advanced NNSs. This finding on the effect of L1 language variables on L2 collocation processing provides further empirical support to previous findings on the similar patterns of collocation use observed in L2 speech and writing (e.g., Wray and Perkins, 2000; Paquot, 2017). In fact, effects of L1 frequency on on-line processing of L2 have been found in other aspects of language, such as in the case of intralingual homographs (i.e., words that have two distinct meanings for one graphemic form common to two languages, e.g., “coin” meaning “corner” in French) whereby the order in which the two lexical entries of intralingual homographs is accessed depends more on the frequency of the entries in either language than the language mode contexts embedding the entries (Beauvillain & Grainger, 1987). In the research on formulaic language, L1 frequency has been found to play a role in the processing of L2 metaphoric expression (Türker, 2016). For metaphors shared across L1 and L2, the ease of L2 metaphor comprehension, given limited context, was positively predicted by the frequency of the metaphor in L1. Our study is the first to demonstrate this carry-over effect of L1 distributional statistics on on-line processing of L2 collocations, indicating a universal and non-selective nature of L1 access in L2 language processing (Dijkstra & van Heuven, 2002, see also Thierry & Wu, 2007; Martin et al., 2009; Zhou et al., 2010).

We also found an intriguing interaction between congruence and dominance: while the main effect of word-pair type did not reach significance, the NNSs in the two beginner groups had more difficulties processing congruent items compared to items in the other conditions. This lack of main effect and the direction of the interaction regarding congruence was unpredicted by existing studies (e.g., Yamashita & Jiang, 2020). The main contributing factor that could explain this discrepancy is how the task environment and participation incentive for the two beginner groups differ from the rest of the NNS groups in our study, which might have led to differences in task strategy. Specifically, while the advanced and intermediate groups participated in the experiment for monetary compensation without any prior knowledge of what the experiment would be about and thus were likely to process the collocation implicitly, the beginner groups completed it as an extra-credit assignment for an English class, from which they might have inferred that the task had something to do with assessing their explicit knowledge of English and thus adopted a more explicit processing strategy around the items with grammatical Chinese equivalents. Therefore, for the two beginner groups, there might have been an over-activation of the metalinguistic knowledge about the two languages, amplifying the interference effect due to active inhibition of the L1 knowledge (Kroll, 2008; see Kroll et al., 2009, page 390-392 for a discussion on L1 inhibition in bilingual language processing). This pattern of inhibition of L1 activation in early L2 learning has been previously observed in native English-speaking Spanish learners (Bice & Kroll, 2015). This explanation is consistent with the fact that many previous studies using a phrasal acceptability judgment paradigm (e.g., Yamashita & Jiang, 2010) only found congruence effect in the accuracy measure, which is underpinned by explicit processing, and not in the RT measure, which is underpinned by implicit processing[22]. This difference in task strategy and the elicited dissociation between implicit and explicit processes can also explain the patterns of results on the L1-only false-friend trials. While the RT measure did not show any false-friend effects, the accuracy on the false-friend trials was significantly lower than baseline for participants in the two beginner NNS groups compared to NSs, who did not show this difference. The explicit processing resulting from awareness of the task demand in the two beginner groups and the increased inhibition of Chinese knowledge might have introduced an additional source of difficulty and distraction, leading those participants to perform worse on items involving Chinese knowledge, i.e., the congruent collocations and false-friends.

More recently, Otwinowska et al. (2021) studied the false-friends phenomenon in the case of L2-to-L1 back transfer and found a similar dissociation between explicit off-line judgment and indices of implicit on-line processing. Highly proficient L2 English speakers with an L1 Polish background read Polish word pairs which were either correctly-formed Polish collocation or English-only collocations translated verbatim into Polish. While the false-friends were judged to be less acceptable than the Polish-only items in an off-line acceptability judgment task, the two

[22] see Yi, 2018 for a discussion on the different latent factors reflected by RT and accuracy measures

types of word pairs evoked the same pattern of brain response in the event-related potentials (ERPs) in a subsequent on-line reading task. Admittedly, these results on back-transfer might not be generalizable to the context of forward-transfer, they nonetheless reinforce the importance of distinguishing the implicit and explicit components when interpreting different sources of empirical data on language processing. Future research on forward-transfer with better control of explicit and implicit measures is needed to confirm the plausibility of our account of the pattern observed in our data.

Chapter 5 General Discussion

This study investigated the processing of collocations in L1 and L2 in double LDTs in two languages - English and Mandarin Chinese - and the underlying factors impacting the performance of English monolingual and Chinese-English bilinguals. Using a collocational priming paradigm, Exp. 1 and 2 replicated the L1 collocation advantage in the two languages and found that after controlling for word-level factors, the lexical decision RTs for words in collocations were faster than those for words in non-collocations. Exp. 3 further extended these findings to NSs. Both NSs and NNSs were affected by characteristics and distribution information at different grain-sizes of languages. Notably, the effects of word-level factors, e.g., word frequency, diminished as collocation strength increased, while the information at the word level is still accessed by both NSs and NNSs in varying degrees.

The first two experiments demonstrated the processing advantage of L1 collocations in both English NSs and Chinese NSs, who also showed sensitivity to the strength of association of the collocations, confirming the psychological reality of collocations in an on-line processing setting. On first review, this seems to reinforce the notion that collocations, once acquired, are stored and retrieved as a whole unit almost like "big words" (e.g., Ellis, 1996, p. 111; Wray, 2002). This is partially true but might not represent the full picture. The fact that the NSs were still highly sensitive to word-level variables, namely word 1 frequency for the English NSs and the lengths and the frequencies for both words for the Chinese NSs, suggests that the information associated with the component-words was still accessed and might have played a role in L1 collocation processing. Experiment 3 found the same collocational priming effects in Chinese L1 English L2 NNSs, extending the finding to NNSs varying in language dominance. Compared to NSs, NNSs were overall more affected by word-level factors, which seems to suggest that NNSs adopted a more analytical approach in line with Wray's (2002) prediction. However, we also see that as dominance increases, the effect of collocation strength becomes more prominent while that of the length and frequency of individual words becomes less prominent for the NNSs. In other words, throughout the L2 learning trajectory, the processing of L2 collocations becomes more and more "native-like" in a graded fashion, not only in terms of the faster RT and higher accuracy, but also in terms of the underlying factors that drive the observed collocation advantage. Unsurprisingly, evidence of L1 influence, primarily reflected by the L1 carry-over effect, also diminished as L2 dominance grew. Taken together, the distinction between NSs and NNSs might lie on a continuum reflected by dominance, rather than being strictly categorical. Given that both NSs and NNSs of varying dominance are affected by the same word-level and collocation-level factors to varying degrees, it is not unreasonable to posit an alternative unified approach to L1 and L2 collocation processing with dominance – i.e., language use and exposure – being the primary variable to account for the quantitative differences between the speaker groups, reconciling Wray's hypotheses and the current findings. It could be argued that both top-down "holistic" processing and bottom-up "analytical" processing happens simultaneously for all groups of speakers, while previous language experience serves as a "slider bar" to tune up or down the degree of the impact from each.

This idea of parallel processing of representations at multiple levels echoes a well-established computational model in word and letter recognition literature, namely, the interactive activation model (McClelland & Rumelhart, 1981), which explains the phenomenon that a letter is recognized better when it is part of a meaningful word versus when it is presented alone, aka., the Word-Superiority Effect (WSE, Reicher, 1969; Wheeler, 1970). In WSE experiments, a string of letters are flashed on screen for a few milliseconds and participants are then tasked with selecting the flashed letter out of two choices in the next screen. For example, if "WORD" had been flashed, participants might have to decide whether "K" or "D" had been in the final letter position. Results show that participants choose the correct letter more consistently when the letter strings are words rather than non-words (e.g., "OWRD") or if they were single letters presented alone (e.g., "___D") (Wheeler, 1970). According to the interactive activation model, when a reader sees the letter D, there are activations at the feature-level (e.g., a vertical bar and a curve in D), letter-level, and word-level (e.g., if D is a part of WORD) happening simultaneously and interactively (i.e., propagated across levels), all contributing to the recognition of the letter. The reason why D is better recognized when presented as part of a meaningful word is because of the increased activation at the word-level, exerting stronger top-down influence, compared to when the letter is presented alone or as part of a non-words, in which only bottom-up feature-level information is available. Importantly, each activated connection in this large parallel network of connections carries a different weight, which is dependent on various factors such as subjective familiarity and inherent characteristics of the words or the letters, such as frequency. For example, if the letter D is presented in the low-frequency word RAND, the top-down word-level activation might carry more weight for a reader who is a shoemaker than for the general population. Likewise, the word WORD, although highly frequent, might only exert a minimal top-down effect to a beginner L2 English learner who is not yet familiar with this word.

The WSE experiments demonstrate how frequency and activation at the word and feature levels may affect the processing of letters, but since linguistic constructions are inherently nested across various overlapping levels, it is likely that the distributional statistics across different levels of representation can be activated interactively and simultaneously (Gries and Ellis, 2015). The pattern of results across our three experiments is consistent with an extended version of the interactive activation model that includes a phrase-level above the word-level, and a parallel representation of all the levels in another language for bilingual speakers. The collocation advantage observed in the LDT indicated a phrasal equivalent of the WSE, a.k.a., Phrase-Superiority Effects (Guo & Ellis, 2020), whereby the processing of words was facilitated when they were part of a meaningful or formulaic phrase, compared to when they were presented as unrelated individual words. Representations at the phrase-level, word-level, and the letter- and feature-levels are activated in parallel, but the activations at the specific levels carry different weight depending on 1) objective characteristics (e.g., frequency, length, association strength) and 2) subjective familiarity of the representations at each level. Word pairs that are highly collocational (i.e., highly frequent and/or strongly associated), given that they are familiar to the speaker, will likely elicit a stronger and faster activation at the phrase-level, over-shadowing the word-level activation; by contrast, words in less collocational or unconnected pairs will likely elicit more activation on the word-level, making the effects of word-level distributional factors more prominent. We see this pattern manifested in our NS data as an interaction between MI and word 1 frequency, and in Öksüz et al.'s (2021) finding that individual word frequency plays a lesser role in the processing of highly frequent collocation items. What about language proficiency or dominance? Low proficiency/dominance NNSs, given limited exposure and familiarity to English collocations, will likely experience weaker activation at the English phrase-level, making the English word-level activation more salient. At the same time, the overall lower exposure and familiarity to English likely leads them to rely more on knowledge of their L1 Chinese, eliciting strong activation at both the word-level and phrase-level representations in Chinese. This explains the carry-over effects of Chinese MI and word frequency on the processing of English collocations by the beginner NNSs in Exp. 3. As language exposure and dominance increases, the English phrase-level representations become more psychologically prominent, eventually overshadowing the English word-level activations, while the English representations at all levels become more activated than their Chinese counterparts, leading to a more “native-like” processing pattern with diminishing L1 influence.

5.1 Limitations and Future Directions

The current study has several limitations that inform future research. First, there were several language variables that were beyond the scope of our study. These include idiomaticity or transparency of meaning which are found to be related to collocation processing (e.g., Gyllstad & Wolter, 2016; Yamashita, 2018). There are also several variables involved in LDTs that we did not focus on. These include polysemy (Hino & Lupker, 1996), valence (Kissler & Koessler, 2011), neighborhood density, imageability/concreteness, etc., many of which have been generalized to the visual word recognition in Chinese (Sze et al., 2015). Future studies could investigate these additional language variables in L2 collocation processing to see if they interact with language proficiency/dominance and with language transfer. Second, although the bilingual participants in our study spanned a wide range of English learning backgrounds, they were all highly Chinese-dominant, even those in the advanced groups. Future research could recruit additional English-Chinese bilingual speakers who are more dominant in English than in Chinese or equally dominant in both languages to further understand the transfer effect, including L2-to-L1 back transfer (e.g., Otwinowska et al., 2021). Another suggestion for further research on collocation processing is to use paradigms where collocations are presented in meaningful contexts (e.g., in a self-paced reading task) rather than in isolation to increase external validity.

Chapter 6 Conclusions

This study investigated collocation processing in L1 and L2 in three behavioral experiments. In a series of LDTs, we successfully replicated the collocation advantage in NSs of two typologically distinct languages as well as in NNSs. We observed that both NSs and NNSs were sensitive to distributional statistics of linguistics representation at both the word-level and the collocation-level, and the differing degrees of sensitivity to each level can be explained by quantitative differences in language exposure. These findings informed a unified processing model of L1 and L2 collocation processing using a framework adapted from the interactive activation model (McClelland and Rumelhart, 1981) which explains the intricate L2 proficiency/dominance effects and its interactions with the sensitivity of distributional statistics of linguistics representations at different levels and in different languages. Our findings are also in line with the predictions of the usage-based approaches that the cumulative experience speakers have with a target language can similarly impact both L1 and L2 speakers (Ellis, 2002).

References

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. https://doi.org/10.1080/15475441.2010.505489

Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82. https://doi.org/10.1016/j.jml.2009.09.005

Bahns, J., & Eldaw, M. (1993). Should we teach EFL students collocations? *System*, *21*(1), 101–114. https://doi.org/10.1016/0346-251X(93)90010-E

Balota, D. A., & Chumbley, J. I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(3), 340–357. https://doi.org/10.1037/0096-1523.10.3.340

Bannard, C., & Matthews, D. (2008). Stored Word Sequences in Language Learning: The Effect of Familiarity on Children's Repetition of Four-Word Combinations. *Psychological Science*, *19*(3), 241–248. https://doi.org/10.1111/j.1467-9280.2008.02075.x

Barry, C., & Seymour, P. H. K. (1988). Lexical priming and sound-to-spelling contingency effects in nonword spelling. *The Quarterly Journal of Experimental Psychology Section A*, *40*(1), 5–40. https://doi.org/10.1080/14640748808402280

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using **lme4**. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Beauvillain, C., & Grainger, J. (1987). Accessing interlexical homographs: Some limitations of a language-selective access. *Journal of Memory and Language*, *26*(6), 658–672. https://doi.org/10.1016/0749-596X(87)90108-2

Bice, K., & Kroll, J. F. (2015). Native language change during early stages of second language learning. *NeuroReport*, *26*(16), 966–971. https://doi.org/10.1097/WNR.0000000000000453

Birdsong, D., Gertken, L.M., & Amengual, M. (2012). Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. Available online: https://sites.la.utexas.edu/bilingual/.

Cañizares-Álvarez, C., & Gathercole, V. C. M. (2020). The influence of first language polysemy and first language and second language lexical frequencies on second language learners' use of false cognates. *International Journal of Bilingualism*, *24*(3), 530–541. https://doi.org/10.1177/1367006918814380

Carrol, G., & Conklin, K. (2014). Getting your wires crossed: Evidence for fast processing of L1 idioms in an L2. *Bilingualism: Language and Cognition*, *17*(04), 784–797. https://doi.org/10.1017/S1366728913000795

Carrol, G., & Conklin, K. (2017). Cross language lexical priming extends to formulaic units: Evidence from eye-tracking suggests that this idea 'has legs.' *Bilingualism: Language and Cognition*, *20*(02), 299–317. https://doi.org/10.1017/S1366728915000103

Carrol, G., Conklin, K., & Gyllstad, H. (2016). Found in translation: The influence of the L1 on the reading of idioms in a L2. *Studies in Second Language Acquisition*, *38*(3), 403–443. https://doi.org/10.1017/S0272263115000492
Childers, J. B., & Tomasello, M. (2001). The role of pronouns in young children's acquisition of the English transitive construction. *Developmental Psychology*, *37*(6), 739–748. https://doi.org/10.1037/0012-1649.37.6.739
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
Crossley, S. A., Salsbury, T., & Mcnamara, D. S. (2014). Assessing Lexical Proficiency Using Analytic Ratings: A Case for Collocation Accuracy. *Applied Linguistics*, amt056. https://doi.org/10.1093/applin/amt056
Davies, M. (2008-). *The Corpus of Contemporary American English (COCA).* Available online at: https://www.english-corpora.org/coca/ (accessed May, 2021).
De Cock, S., Granger, S., Leech, G., & McEnery, T. (2014). An automated approach to the phrasicon of EFL learners. In S. Granger, *Learner English on Computer* (1st ed., pp. 67–79). Routledge. https://doi.org/10.4324/9781315841342-5
Dijkstra, T., & van Heuven, W. J. B. (2002). The architecture of the bilingual word recognition system: From identification to decision. *Bilingualism: Language and Cognition*, *5*(3), 175–197. https://doi.org/10.1017/S1366728902003012
Ding, C., & Reynolds, B. L. (2019). The effects of L1 congruency, L2 proficiency, and the collocate-node relationship on the processing of L2 English collocations by L1-Chinese EFL learners. *Review of Cognitive Linguistics*, *17*(2), 331–357. https://doi.org/10.1075/rcl.00038.din
Du, L., Elgort, I., & Siyanova-Chanturia, A. (2021). Cross-Language Influences in the Processing of Multiword Expressions: From a First Language to Second and Back. *Frontiers in Psychology*, *12*, 666520. https://doi.org/10.3389/fpsyg.2021.666520
Duanmu, S. (2012). Word-length preferences in Chinese: A corpus study. *Journal of East Asian Linguistics*, *21*(1), 89–114. https://doi.org/10.1007/s10831-011-9087-y
Duanmu, S., Feng, S., Dong, Y., & Zhang, Y. (2017). A Judgment Study of Length Patterns in Chinese: Prosody, Last Resort, and Other Factors. *Journal of Chinese Linguistics*. https://doi.org/10.1353/jcl.2017.0026
Dunning, T. (1993). Accurate Methods for the Statistics of Surprise and Coincidence. *Computational Linguistics*, *19*, 61–74.
Durrant, P., & Doherty, A. (2010). Are high-frequency collocations psychologically real? Investigating the thesis of collocational priming. *Corpus Linguistics and Linguistic Theory*, *6*(2). https://doi.org/10.1515/cllt.2010.006
Ellis, N. C. (1996). Sequencing in SLA: Phonological Memory, Chunking, and Points of Order. *Studies in Second Language Acquisition*, *18*(1), 91–126. https://doi.org/10.1017/S0272263100014698
Ellis, N. C. (2002). Frequency effects in language processing. *Studies in Second Language Acquisition*, *24*(02). https://doi.org/10.1017/S0272263102002024
Ellis, N. C. (2006). Language Acquisition as Rational Contingency Learning. *Applied Linguistics*, *27*(1), 1–24. https://doi.org/10.1093/applin/ami038
Fang, N., & Zhang, P. (2021). L1 congruency, word frequency, collocational frequency, L2 proficiency, and their combined effects on Chinese–English bilinguals' L2 collocational

processing. *International Journal of Bilingualism*, *25*(5), 1429–1445. https://doi.org/10.1177/13670069211024747
Frauenfelder, U. H., Baayen, R. H., & Hellwig, F. M. (1993). Neighborhood Density and Frequency Across Languages and Modalities. *Journal of Memory and Language, 32*(6), 781–804. https://doi.org/10.1006/jmla.1993.1039
Gablasova, D., Brezina, V., & McEnery, T. (2017). Collocations in Corpus-Based Language Learning Research: Identifying, Comparing, and Interpreting the Evidence: Collocations in Corpus-Based Language Learning Research. *Language Learning*, *67*(S1), 155–179. https://doi.org/10.1111/lang.12225
Gries, S. Th. (2021). What do (some of) our association measures measure (most)? Association? *Journal of Second Language Studies*. https://doi.org/10.1075/jsls.21028.gri
Gries, S. Th., & Ellis, N. C. (2015). Statistical Measures for Usage-Based Linguistics: Statistical Measures for Usage-Based Linguistics. *Language Learning*, *65*(S1), 228–255. https://doi.org/10.1111/lang.12119
Guo, R., & Ellis, N. C. (2021). Language Usage and Second Language Morphosyntax: Effects of Availability, Reliability, and Formulaicity. *Frontiers in Psychology*, *12*, 582259. https://doi.org/10.3389/fpsyg.2021.582259
Gyllstad, H., & Wolter, B. (2016). Collocational Processing in Light of the Phraseological Continuum Model: Does Semantic Transparency Matter?: Collocational Processing and Semantic Transparency. *Language Learning*, *66*(2), 296–323. https://doi.org/10.1111/lang.12143
Hino, Y., & Lupker, S. J. (1996). Effects of polysemy in lexical decision and naming: An alternative to lexical access accounts. *Journal of Experimental Psychology: Human Perception and Performance*, *22*(6), 1331–1356. https://doi.org/10.1037/0096-1523.22.6.1331
Howarth, P. (1998). Phraseology and Second Language Proficiency. *Applied Linguistics*, *19*(1), 24–44. https://doi.org/10.1093/applin/19.1.24
Jakubíček, M., Kilgarriff, A., Kovář, V., Rychlý, P., & Suchomel, V. (2013). The TenTen corpus family. In 7th International Corpus Linguistics Conference CL (pp. 125-127). Available online at: https://www.sketchengine.eu/zhtenten-chinese-corpus/ (accessed May, 2021).
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
Kellerman, E. (1986). An eye for an eye: Crosslinguistic constraints on the development of the L2 lexicon. In Kellerman, E. and Smith, M. Sharwood (eds.) *Crosslinguistic influence in second language acquisition.* Oxford: Pergamon Press. 35–48.
Kilgarriff, A, Baisa, V, Bušta, J, Jakubíček, M, Kovář, V, Michelfeit, J, Rychlý, P, & Suchomel, V. (2014). The Sketch Engine: ten years on. *Lexicography, 1*, 7-36.
Kissler, J., & Koessler, S. (2011). Emotionally positive stimuli facilitate lexical decisions—An ERP study. *Biological Psychology*, *86*(3), 254–264. https://doi.org/10.1016/j.biopsycho.2010.12.006
Kroll, J. F., Bobb, S. C., Misra, M., & Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, *128*(3), 416–430. https://doi.org/10.1016/j.actpsy.2008.02.001
Kroll, J. F., & De, G. A. M. B. (Eds.). (2009). *Handbook of bilingualism: Psycholinguistic approaches.* Oxford University Press, Incorporated.

Laufer, B., & Waldman, T. (2011). Verb-Noun Collocations in Second Language Writing: A Corpus Analysis of Learners' English: Verb-Noun Collocations in L2 Writing. *Language Learning*, *61*(2), 647–672. https://doi.org/10.1111/j.1467-9922.2010.00621.x

Martin, C. D., Dering, B., Thomas, E. M., & Thierry, G. (2009). Brain potentials reveal semantic priming in both the 'active' and the 'non-attended' language of early bilinguals. *NeuroImage*, *47*(1), 326–333. https://doi.org/10.1016/j.neuroimage.2009.04.025

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review*, *88*(5), 375–407. https://doi.org/10.1037/0033-295X.88.5.375

Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, *90*(2), 227–234. https://doi.org/10.1037/h0031564

McDonald, S. A., & Shillcock, R. C. (2003). Eye Movements Reveal the On-Line Computation of Lexical Probabilities During Reading. *Psychological Science*, *14*(6), 648–652. https://doi.org/10.1046/j.0956-7976.2003.psci_1480.x

McNamara, T. P. (2005). *Semantic Priming: Perspectives from Memory and Word Recognition* (0 ed.). Psychology Press. https://doi.org/10.4324/9780203338001

Nguyen, T. M. H., & Webb, S. (2017). Examining second language receptive knowledge of collocation and factors that affect learning. *Language Teaching Research*, *21*(3), 298–320. https://doi.org/10.1177/1362168816639619

Nesselhauf, N. (2003). The Use of Collocations by Advanced Learners of English and Some Implications for Teaching. *Applied Linguistics*, *24*(2), 223–242. https://doi.org/10.1093/applin/24.2.223

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. https://doi.org/10.1111/lang.12427

Paquot, M. (2017). L1 frequency in foreign language acquisition: Recurrent word combinations in French and Spanish EFL learner writing. *Second Language Research*, *33*(1), 13–32. https://doi.org/10.1177/0267658315620265

Pawley, A., & Syder, F. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*.

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Qin, Z., & Duanmu, S. (2017). A judgment study of word-length preferences in Chinese NN compounds. *Lingua*, *198*, 1–21. https://doi.org/10.1016/j.lingua.2017.06.012

R Core Team (2017). R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing.

Reicher, G. M. (1969). Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, *81*(2), 275–280. https://doi.org/10.1037/h0027768

Schwanenflugel, P. J., Harnishfeger, K. K., & Stowe, R. W. (1988). Context availability and lexical decisions for abstract and concrete words. *Journal of Memory and Language*, *27*(5), 499–520. https://doi.org/10.1016/0749-596X(88)90022-8

Siyanova, A., & Schmitt, N. (2008). L2 Learner Production and Processing of Collocation: A Multi-study Perspective. *The Canadian Modern Language Review*, *64*(3), 429–458. https://doi.org/10.3138/cmlr.64.3.429

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

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (Eds.). (2018). *Understanding Formulaic Language: A Second Language Acquisition Perspective* (1st ed.). Routledge. https://doi.org/10.4324/9781315206615

Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2018). What online processing tells us about formulaic language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language* (1st ed., pp. 38–61). Routledge. https://doi.org/10.4324/9781315206615-3

Skarabela, B., Ota, M., O'Connor, R., & Arnon, I. (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, *211*, 104612. https://doi.org/10.1016/j.cognition.2021.104612

Sze, W. P., Yap, M. J., & Rickard Liow, S. J. (2015). The role of lexical variables in the visual recognition of Chinese characters: A megastudy analysis. *Quarterly Journal of Experimental Psychology*, *68*(8), 1541–1570. https://doi.org/10.1080/17470218.2014.985234

Thierry, G., & Wu, Y. J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, *104*(30), 12530–12535. https://doi.org/10.1073/pnas.0609927104

Tremblay, A., Derwing, B., Libben, G., & Westbury, C. (2011). Processing Advantages of Lexical Bundles: Evidence From Self-Paced Reading and Sentence Recall Tasks: Lexical Bundle Processing. *Language Learning*, *61*(2), 569–613. https://doi.org/10.1111/j.1467-9922.2010.00622.x

Türker, E. (2016). The role of L1 conceptual and linguistic knowledge and frequency in the acquisition of L2 metaphorical expressions. *Second Language Research*, *32*(1), 25–48. https://doi.org/10.1177/0267658315593336

Ueno, T. (2009). *An investigation of the relationship between the development of bilingual semantic organisation and interactive connectivity across languages* (Unpublished doctoral dissertation). Trinity College, Dublin.

Vilkaitė, L. (2016). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(10), 1632–1642. https://doi.org/10.1037/xlm0000259

Wheeler, D. D. (1970). Processes in word recognition. *Cognitive Psychology*, *1*(1), 59–85. https://doi.org/10.1016/0010-0285(70)90005-8

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

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics*, *36*(5), 1193–1221. https://doi.org/10.1017/S0142716414000113

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. https://doi.org/10.1017/S0272263117000237

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

Wray, A. (2002). *Formulaic Language and the Lexicon* (1st ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511519772

Wulff, S., & Ellis, N. C. (2018). Usage-based approaches to second language acquisition. In D. Miller, F. Bayram, J. Rothman, & L. Serratrice (Eds.), *Studies in Bilingualism* (Vol. 54, pp. 37–56). John Benjamins Publishing Company. https://doi.org/10.1075/sibil.54.03wul

Xu, J. (2015). *Predicting ESL learners' oral proficiency by measuring the collocations in their spontaneous speech* (Order No. 10009317). Available from Linguistics and Language Behavior Abstracts (LLBA); ProQuest Dissertations & Theses Global. (1762721972).

Yamashita, J. (2018). Possibility of semantic involvement in the L1-L2 congruency effect in the processing of L2 collocations. *Journal of Second Language Studies*, *1*(1), 60–78. https://doi.org/10.1075/jsls.17024.yam

Yamashita, J., & Jiang, N. (2010). L1 Influence on the Acquisition of L2 Collocations: Japanese ESL Users and EFL Learners Acquiring English Collocations. *TESOL Quarterly*, *44*(4), 647–668. https://doi.org/10.5054/tq.2010.235998

Yi, W. (2018). Statistical sensitivity, cognitive aptitudes, and processing of collocations. *Studies in Second Language Acquisition*, *40*(4), 831–856. https://doi.org/10.1017/S0272263118000141

Zhou, H., Chen, B., Yang, M., & Dunlap, S. (2010). Language Nonselective Access to Phonological Representations: Evidence from Chinese–English Bilinguals. *Quarterly Journal of Experimental Psychology*, *63*(10), 2051–2066. https://doi.org/10.1080/17470211003718705

Appendices

Appendix 1 – Additional notes on item development

For the items marked as “collocation” in either language, i.e., English-only, Chinese-only, and congruent items, there were 18 verb-noun pairs, where the noun is always the direct object of the verb, hereafter, VN (e.g., “answer phone”) and 18 adjective-noun or noun-noun pairs, where the adjective or the first noun is always the descriptor of the second noun, hereafter, AN (e.g., “short notice”; “leather jacket”). Since the first nouns of the noun-noun items usually fulfill the syntactic slot of an adjective, we categorized them as adjective-noun items and will use the label AN for both true adjective-noun items and noun-noun items. The 108 non-collocational baseline items were created by pseudo-randomly combining the component words of the collocations so that every collocation word appeared once as part of a baseline item. Additionally, we created nonce trials that contained one nonword each (half on as the first word and half as the second word), with the other word always being one of the collocation component words, all of which appeared once in the nonce trials (n=216). Table 4 includes sample items illustrating this process. So this means that in the 432 total trials in the task, each participant sees every real word 3 times, once as part of a collocation (once of the three types), once as part of a baseline item, and once as part of a nonce trial. To offset any repetition priming effects leading to facilitated processing of words that were already presented earlier, the trial order is individually randomized for each participant.

Table 4. Sample experimental items in each condition.

Type/Condition	English-only	Chinese-only	Congruent	Baseline1	Baseline2	Baseline3
VN	answer phone	drink soup	break record	answer soup	drink record	break phone
VN NONCE	answer knags	drink cleld	break gnoat	spouch phone	tighs soup	cadd record
AN	jet lag	dark cuisine	front door	jet cuisine	dark door	front lag
AN NONCE	jet ferm	dark skotch	front slade	rher lag	leese cuisine	fres door

This paragraph will further describe our criteria for the corpus searches and how collocation status was verified in the corpus. For the English items used in Exp 1 and 3, we used the "collocates" function in COCA. For both VN and AN items (collocations and baselines), the second word was set as the node word and the first word as the collocate. To decide the span of the collocation, we consulted native English-speaking research assistants and agreed on the span of [-2, +2] for VN items and [-1, 2] for AN items. This means that for VN items such as "answer phone", the search returns the frequency of all instances where the lemma "answer" occurs one or two words before or after "phone", so instances like "answer the phone" (-2) and "the phone was answered" (+2) were all included in the frequency count; for AN items such as "fair skin", the search returns the frequency where the lemma "fair" occurs one word before or one or two words after the lemma "skin", so instances like "fair skin" (-1) and "the skin is fair" (+2) were all included in the frequency count. For the Chinese items used in Exp 2, we used the "[meet]" command in Sketch Engine. Same as the English items, the second word was set as the node word. We consulted a native Chinese-speaking colleague with expertise in Chinese linguistics to decide the span for the Chinese collocations and set the span of [-2, 1] for VN items and [-2, 4] for AN items. Due to the problem of indefinite word boundaries in Chinese, many Chinese collocations register occurrences both as a collocation and as a single word entry in the corpus (e.g., "watch movie" occurred 113,049 times as a collocation and 3,049 times as a single word. After manually examining the contexts in which an item was tagged as a collocation versus the context in which the same item was tagged as a word, we noticed that the distinction was arbitrary with no systematicity. This was confirmed by the Sketch Engine Support Team. Following their advice, we combined the number of occurrences as a collocation and the number of occurrences as a word to get the final collocation frequency for each item. Note that all corpus searches were conducted on lemma forms.

To measure collocation association strength, we calculated Mutual Information (MI) scores for each item using the following formula: MI = log ((AB * sizeCorpus) / (A * B * span)) / log (2) where AB stands for the frequency of the collocation, sizeCorpus stands for the total number of words in the corpus, A and B stand for the frequency of the node word and the collocate, respectively, span stands for the total span of the words, and log(2) stands for the log10 of the number 2, i.e., 0.30103. For example, to calculate the English MI score for the collocate “phone” occurring near “answer” in the [-2,+2] (i.e., 4-gram) span, MI = log ((1,537 * 1,001,610,938) / (171,844 * 65734 * 4)) / .30103 = 5.09. Chinese MI scores were calculated in the same way using the frequencies gathered from the zhTenTen17 Chinese corpus.

Appendix 2 – Procedures for Removing Outliers

Because all the experiments were run remotely with no supervision, we took steps to ensure the quality of the data by reducing the influence of outliner participants and responses. To begin with, we identified and removed low-effort participants whose average accuracy rate was less than 80% and whose average RT (of all trials) was extreme according to the interquartile rule, i.e., either less than Q1-1.5IQR or higher than Q3+1.5IQR, where Q1 and Q3 stand for the first and third quartile, and IQR stands for interquartile range, i.e., the difference between Q3 and Q1. In addition, we identified one participant who did the experiment twice using different email addresses and we removed the responses from their second attempt. After eliminating low-effort and abnormal participants, Exp 1 (English NSs) had 41 remaining participants (93% out of 44 total), Exp 2 (Chinese NSs) had 163 remaining participants (94% out of 173 total), and Exp 3 had 149 remaining participants (86% out of 173 total)

Appendix 3 - Language background of Exp 3 participants in different proficiency groups

Table 5. Mean BLP scores for each group across four components in Chinese (CH) and English (EN): language history, use, proficiency, and attitudes, along with the dominance scores calculated based on these four components.[23]*.*

NNS Groups	# of Participants		History_CH	History_EN	Use_CH	Use_EN	Proficiency_CH	Proficiency_EN	Attitudes_CH	Attitudes_EN	Dominance_CH	Dominance_EN
ADV+	17	Mean	38.59	21.61	27.89	26.67	47.14	47.94	49.14	41.13	162.76	137.34
		SD	7.41	9.53	8.65	9.53	8.44	4.98	6.23	7.70	23.09	22.96
ADV	33	Mean	44.39	11.90	38.25	15.83	53.62	39.35	52.59	34.86	188.85	101.94
		SD	5.68	4.10	7.76	4.10	1.31	8.42	3.44	10.08	13.05	26.02
INT	44	Mean	50.53	11.01	50.04	4.54	49.91	21.46	50.79	23.97	201.28	60.98
		SD	2.86	5.32	5.05	5.32	5.26	10.45	4.90	13.08	11.38	26.99
BEG+	38	Mean	46.92	14.56	49.82	4.30	51.11	17.03	50.34	22.03	198.19	57.91
		SD	2.56	4.11	5.64	4.11	4.61	7.83	6.00	9.74	10.90	19.81
BEG	17	Mean	48.31	12.10	49.03	4.86	50.48	15.57	52.35	20.93	200.17	53.47
		SD	3.12	4.36	5.87	4.36	3.55	9.82	2.56	11.63	9.83	24.06
Total / Average	149		47.31	13.03	46.35	7.89	50.77	23.84	51.16	25.80	195.60	70.56

[23] In-depth descriptions of how each component is measured and how the dominance scores are calculated can be found here: https://sites.la.utexas.edu/bilingual/

Appendix 4 – Complete model output[24]

4.1 Experiment 1

Table 6. Model EN-1: English NS LDT Log Winsorized RT by English collocation status, word 1 length, word 2 length, word 1 log frequency, and word 2 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
Intercept	-0.009	0.034	223.800	-0.266	0.791		0.013	0.114	0.000	0.020
EN_col	-0.026	0.006	62.640	-4.359	0.000	***	0.001	0.024		
EN_W1len	0.000	0.002	210.000	0.231	0.817					
EN_W2len	0.002	0.001	210.000	1.449	0.149					
EN_W1freq	-0.019	0.004	210.000	-4.412	0.000	***				
EN_W2freq	-0.001	0.004	210.000	-0.305	0.761					

[24]Abbreviation of variable names in tables throughout Appendix 4: EN_col (English collocation status); EN_W1len (English word 1 length); EN_W2len (English word 2 length); EN_W1freq (English word 1 log frequency); EN_W2freq (English word 2 log frequency); EN_MI (English Mutual Information); CH_col (Chinese collocation status); CH_W1strk (Chinese word 1 stroke count); CH_W2strk (Chinese word 2 stroke count); CH_W1freq (Chinese word 1 log frequency); CH_W2freq (Chinese word 2 log frequency); CH_MI (Chinese Mutual Information); Dominance (English dominance obtained from the BLP). Interaction terms are noted with a colon ":" between two variables. Reference levels for categorical variables: for English collocation status and Chinese collocation status, the reference level is non-collocation; for group, the reference level is Group 1 (English monolinguals); for condition, the reference level is baseline.

Table 7. Model EN-1: English NS LDT Accuracy rate by English collocation status, word 1 length, word 2 length, word 1 log frequency, and word 2 log frequency.*[25]

	Fixed effects						*Random effects*			
							By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.889	0.024	221.400	36.703	0.000	***	0.001	0.026	0.000	0.014
EN_col	0.008	0.004	210.000	2.077	0.039	*				
EN_W1len	0.004	0.001	210.000	2.738	0.007	**				
EN_W2len	0.002	0.001	210.000	1.572	0.117					
EN_W1freq	0.008	0.004	210.000	2.218	0.028	*				
EN_W2freq	0.005	0.003	210.000	1.481	0.140					

[25] Throughout Appendix 4, the asterisks (*) notation is used to refer to the Accuracy analysis that corresponds to the RT analysis with the same factors.

Table 8. Model EN-1a: English NS LDT Log Winsorized RT by English collocation status, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, and collocation status : word 1 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
Intercept	0.024	0.036	241.500	0.654	0.514		0.012	0.108	0.000	0.019
EN_col	-0.130	0.046	209.000	-2.814	0.005	**				
EN_W1len	0.000	0.002	209.000	0.062	0.950					
EN_W2len	0.001	0.001	209.000	1.176	0.241					
EN_W1freq	-0.024	0.005	209.000	-4.996	0.000	***				
EN_W2freq	-0.002	0.004	209.000	-0.462	0.644					
EN_col:EN_W1freq	0.021	0.009	209.000	2.268	0.024	*				

Table 9. Model EN-2 English NS LDT Log Winsorized RT by English collocation status, MI, word 1 length, word 2 length, word 1 log frequency, and word 2 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
Intercept	-0.004	0.034	221.800	-0.129	0.898		0.013	0.114	0.000	0.019
EN_col	-0.012	0.008	143.200	-1.494	0.137		0.001	0.024		
EN_MI	-0.001	0.001	209.000	-2.464	0.015	*				
EN_W1len	0.001	0.002	209.000	0.360	0.719					
EN_W2len	0.001	0.001	209.000	0.879	0.381					
EN_W1freq	-0.019	0.004	209.000	-4.469	0.000	***				
EN_W2freq	-0.003	0.004	209.000	-0.769	0.443					

Table 10. Model EN-2 English NS LDT Accuracy Rate by English MI, word 1 length, word 2 length, word 1 log frequency, and word 2 log frequency.*

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters[26]	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.887	0.024	221.400	36.645	0.000	***	0.001	0.026	0.000	0.014
EN_MI	0.001	0.000	210.000	2.338	0.020	*				
EN_W1len	0.004	0.001	210.000	2.673	0.008	**				
EN_W2len	0.002	0.001	210.000	1.898	0.059	.				
EN_W1freq	0.008	0.004	210.000	2.207	0.028	*				
EN_W2freq	0.006	0.003	210.000	1.800	0.073	.				

[26] The EN collocation status variable was not included as it competes for effects with EN MI, causing both of them to be insignificant when included in the same model. Since MI is shown to be the overriding predictor in the previous RT analysis (Model EN-2), we keep the MI variable in this Accuracy analysis.

Table 11. Model EN-2a: English NS LDT Log Winsorized RT by English collocation status, MI, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, and MI : word 1 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
Intercept	-0.017	0.034	222.100	-0.498	0.619		0.013	0.114	0.000	0.019
EN_col	-0.012	0.008	140.700	-1.456	0.148		0.001	0.024		
EN_MI	-0.010	0.004	208.000	-2.723	0.007	**				
EN_W1len	0.000	0.002	208.000	0.300	0.764					
EN_W2len	0.001	0.001	208.000	0.605	0.546					
EN_W1freq	-0.015	0.005	208.000	-3.211	0.002	**				
EN_W2freq	-0.004	0.004	208.000	-1.080	0.281					
EN_MI:EN_W1freq	0.002	0.001	208.000	2.373	0.019	*				

4.2 Experiment 2

Table 12. Model CH-1: Chinese NS LDT Log Winsorized RT by Chinese collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, and word 2 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.102	0.026	256.600	3.885	0.000	***	0.012	0.108	0.001	0.025
CH_col	-0.047	0.004	210.000	-11.502	0.000	***				
CH_W1strk	0.001	0.000	210.000	3.066	0.002	**				
CH_W2strk	0.002	0.000	210.000	4.250	0.000	***				
CH_W1freq	-0.012	0.003	210.000	-3.997	0.000	***				
CH_W2freq	-0.013	0.003	210.000	-4.949	0.000	***				

Table 13. Model CH-1 Chinese NS LDT Accuracy rate by Chinese collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, and word 2 log frequency.*

	Fixed effects						*Random effects*			
							By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.926	0.015	216.800	61.957	0.000	***	0.001	0.024	0.000	0.012
CH_col	0.011	0.002	210.000	4.439	0.000	***				
CH_W1strk	0.000	0.000	210.100	0.258	0.797					
CH_W2strk	0.000	0.000	210.100	-0.868	0.386					
CH_W1freq	0.004	0.002	210.100	2.432	0.016	*				
CH_W2freq	0.004	0.002	210.100	2.784	0.006	**				

Table 14. Model CH-1a: Chinese NS LDT Log Winsorized RT by Chinese collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, word 2 log frequency, and collocation status : word 1 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
Intercept	0.141	0.030	246.000	4.777	0.000	***	0.012	0.110	0.001	0.024
CH_col	-0.143	0.035	210.000	-4.081	0.000	***	0.000	0.016		
CH_W1strk	0.001	0.000	209.000	3.011	0.003	**				
CH_W2strk	0.002	0.000	209.000	4.361	0.000	***				
CH_W1freq	-0.018	0.004	209.000	-4.911	0.000	***				
CH_W2freq	-0.015	0.003	209.000	-5.395	0.000	***				
CH_col:CH_W1freq	0.017	0.006	209.000	2.763	0.006	**				

Table 15. Model CH-1a: Chinese NS LDT Accuracy rate by Chinese collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, word 2 log frequency, and collocation status : word 1 log frequency.*

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.901	0.017	216.000	53.166	0.000	***	0.001	0.028	0.000	0.012
CH_col	0.071	0.021	210.200	3.407	0.001	***	0.000	0.013		
CH_W1strk	0.000	0.000	209.100	0.370	0.712					
CH_W2strk	0.000	0.000	209.100	-0.931	0.353					
CH_W1freq	0.008	0.002	209.100	3.698	0.000	***				
CH_W2freq	0.005	0.002	209.100	3.254	0.001	**				
CH_col:CH_W1freq	-0.010	0.004	209.100	-2.925	0.004	**				

Table 16. Model CH-2: Chinese NS LDT Log Winsorized RT by Chinese MI, collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, and word 2 log frequency.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.126	0.026	258.000	4.838	0.000	***	0.012	0.110	0.001	0.023
CH_MI	-0.002	0.001	209.000	-4.338	0.000	***				
CH_col	-0.024	0.007	221.000	-3.616	0.000	***	0.000	0.016		
CH_W1strk	0.001	0.000	209.000	3.283	0.001	**				
CH_W2strk	0.001	0.000	209.000	3.678	0.000	***				
CH_W1freq	-0.014	0.003	209.000	-4.847	0.000	***				
CH_W2freq	-0.016	0.003	209.000	-5.963	0.000	***				

Table 17. Model CH-2: Chinese NS LDT Accuracy rate by Chinese MI, word 1 length, word 2 length, word 1 log frequency, and word 2 log frequency.*

	Fixed effects						*Random effects*			
							By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.917	0.015	215.600	60.741	0.000	***	0.001	0.024	0.000	0.012
CH_MI	0.001	0.000	209.000	2.670	0.008	**				
CH_col	0.002	0.004	209.000	0.543	0.587					
CH_W1strk	0.000	0.000	209.000	0.206	0.837					
CH_W2strk	0.000	0.000	209.100	-0.444	0.657					
CH_W1freq	0.005	0.002	209.100	2.890	0.004	**				
CH_W2freq	0.005	0.002	209.000	3.330	0.001	**				

Table 18. Model CH-2a: Chinese NS LDT Log Winsorized RT by Chinese MI, collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, word 2 log frequency, and MI : word 1 log frequency.

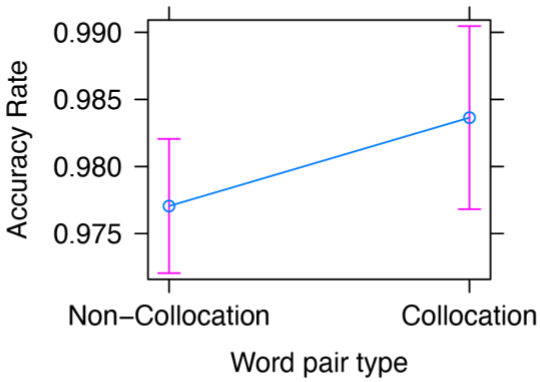
							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
Intercept	0.126	0.026	255.900	4.883	0.000	***	0.012	0.108	0.001	0.023
CH_MI	-0.007	0.003	208.000	-2.416	0.017	*				
CH_col	-0.027	0.007	208.000	-4.009	0.000	***				
CH_W1strk	-0.014	0.003	208.100	-4.841	0.000	***				
CH_W2strk	-0.016	0.003	208.000	-6.035	0.000	***				
CH_W1freq	0.001	0.000	208.000	3.367	0.001	***				
CH_W2freq	0.001	0.000	208.000	3.767	0.000	***				
CH_MI:CH_W1freq	0.001	0.001	208.000	1.663	0.098	.				

Table 19. Model CH-2a: Chinese NS LDT Accuracy rate by Chinese MI, collocation status, word 1 stroke count, word 2 stroke count, word 1 log frequency, word 2 log frequency, MI : word 1 log frequency, and MI: word 2 stroke count.*

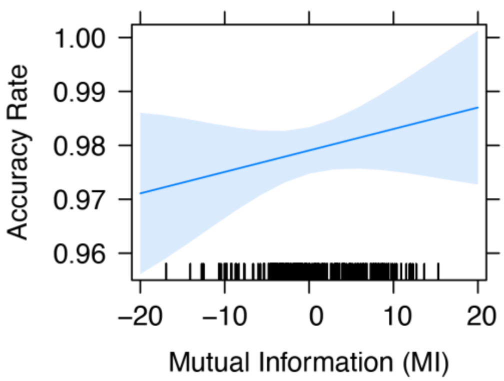
							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.923	0.015	213.800	62.633	0.000	***	0.001	0.024	0.000	0.011
CH_MI	0.004	0.002	207.000	2.331	0.021	*				
CH_col	0.007	0.004	206.900	1.624	0.106					
CH_W1strk	0.005	0.002	207.100	2.742	0.007	**				
CH_W2strk	0.005	0.002	207.000	3.287	0.001	**				
CH_W1freq	0.000	0.000	207.000	-0.165	0.869					
CH_W2freq	0.000	0.000	207.000	-1.198	0.232					
CH_MI:CH_W1freq	-0.001	0.000	207.000	-2.835	0.005	**				
CH_MI:CH_W2strk	0.000	0.000	207.000	2.960	0.003	**				

Figure 7. Effects of Chinese language variables on Accuracy rate for Chinese NSs (Model CH-2a).*

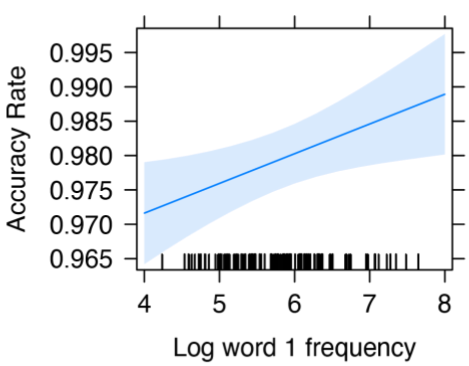
a.) Word pair type



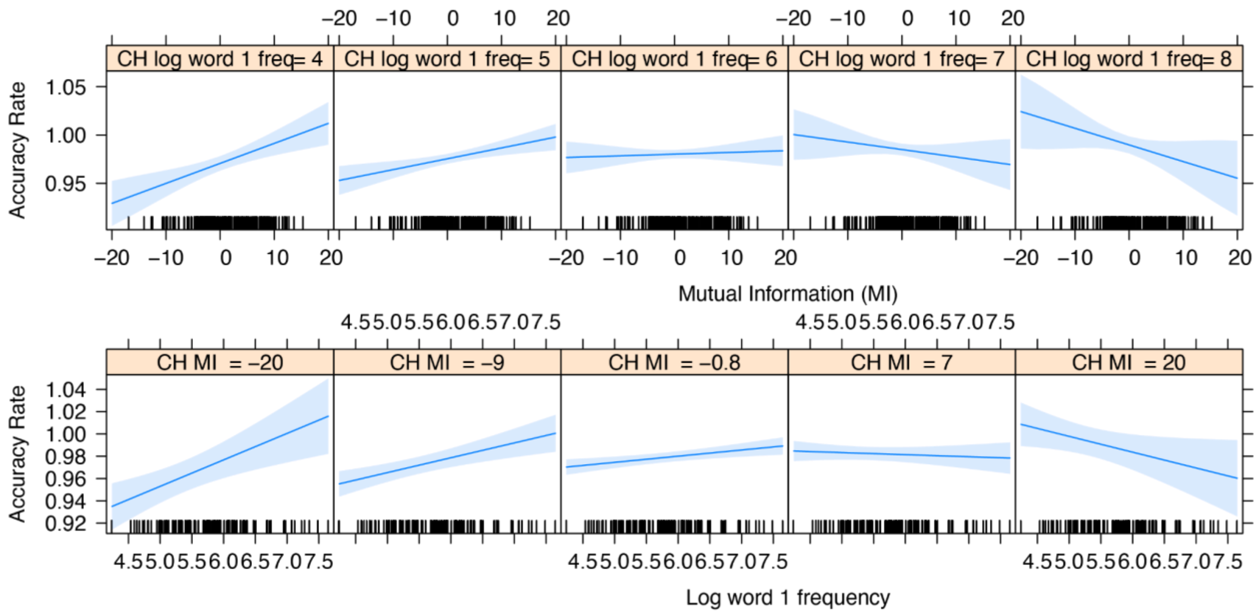
b.) MI



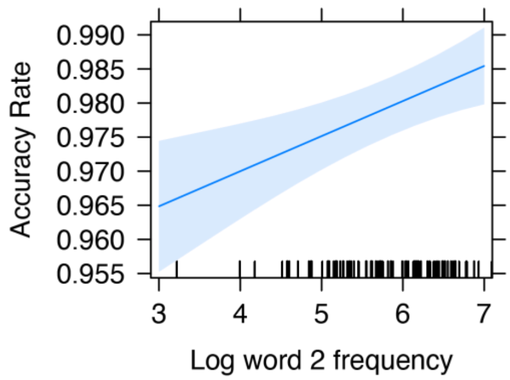
c.) Log word 1 frequency



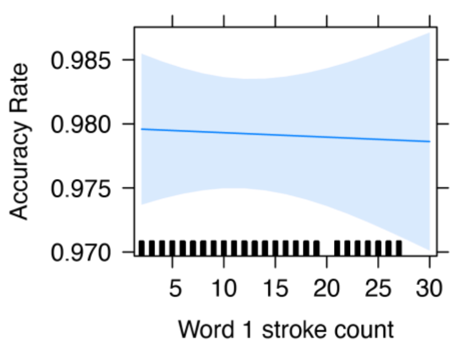
d.) MI : log word 1 frequency



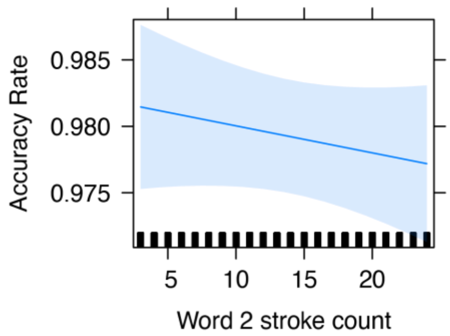
e.) Log word 2 frequency



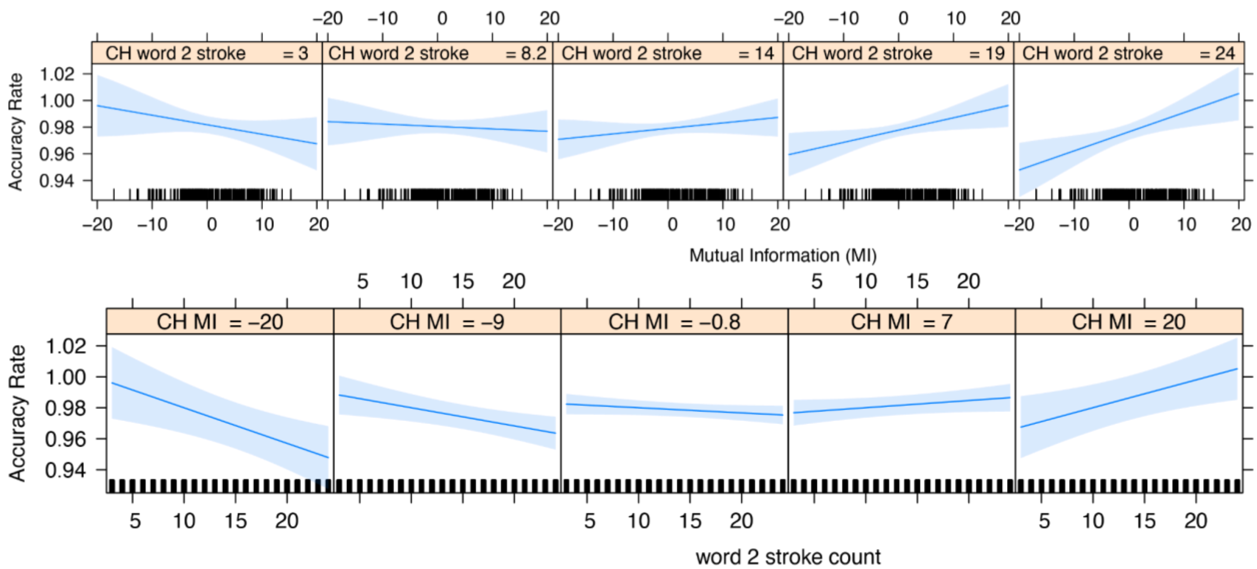
f.) Word 1 stroke count



g.) Word 2 stroke count



h.) MI : word 2 stroke count



4.3 Experiment 3

Table 20. Model ESL-1: NNS LDT Log Winsorized RT by English collocation status, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, and English dominance.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.506	0.041	355.600	12.334	0.000	***	0.019	0.136	0.019	0.031
EN_col	-0.013	0.005	210.000	-2.672	0.008	**				
EN_W1len	0.004	0.002	209.900	2.398	0.017	*				
EN_W2len	0.008	0.001	210.000	5.757	0.000	***				
EN_W1freq	-0.043	0.005	210.000	-8.946	0.000	***				
EN_W2freq	-0.027	0.004	210.000	-6.508	0.000	***				
Dominance	-0.002	0.000	147.000	-5.365	0.000	***				

Table 21. Model ESL-2: NNS LDT Log Winsorized RT by English MI, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, and language dominance.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.510	0.041	355.700	12.449	0.000	***	0.019	0.136	0.001	0.031
EN_MI	-0.001	0.000	210.000	-3.072	0.002	**				
EN_W1len	0.004	0.002	209.900	2.501	0.013	*				
EN_W2len	0.007	0.001	210.000	5.297	0.000	***				
EN_W1freq	-0.043	0.005	210.000	-8.984	0.000	***				
EN_W2freq	-0.029	0.004	210.000	-6.969	0.000	***				
Dominance	-0.002	0.000	147.000	-5.365	0.000	***				

Table 22. Model ESL-2a: NNS LDT Log Winsorized RT by English MI, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, language dominance, and two-way interactions involving English MI.

				Fixed effects			Random effects			
							By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.516	0.041	350.600	12.437	0.000	***	0.019	0.136	0.001	0.031
EN_MI	0.000	0.006	207.500	0.067	0.947					
EN_W1len	0.006	0.002	205.900	3.050	0.003	**				
EN_W2len	0.007	0.002	206.000	4.612	0.000	***				
EN_W1freq	-0.045	0.005	206.000	-8.257	0.000	***				
EN_W2freq	-0.029	0.004	206.000	-6.672	0.000	***				
Dominance	-0.002	0.000	147.200	-5.443	0.000	***				
EN_MI:EN_W1len	0.001	0.000	205.900	2.067	0.040	*				
EN_MI:EN_W2len	0.000	0.000	206.000	-0.133	0.895					
EN_MI:EN_W1freq	-0.001	0.001	205.900	-0.658	0.511					
EN_MI:EN_W2freq	0.000	0.001	206.000	-0.217	0.829					
EN_MI:Dominance	0.000	0.000	32,360.000	-3.409	0.001	***				

Table 23. Model ESL-2b: NNS LDT Log Winsorized RT by English MI, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, language dominance, and two-way interactions involving English MI and language dominance.

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.596	0.050	759.100	11.844	0.000	***	0.019	0.136	0.034	0.185
EN_MI	0.000	0.006	207.500	0.063	0.950					
EN_W1len	0.007	0.003	584.500	2.864	0.004	**				
EN_W2len	0.011	0.002	561.500	5.631	0.000	***				
EN_W1freq	-0.058	0.007	547.600	-8.368	0.000	***				
EN_W2freq	-0.039	0.006	591.400	-6.732	0.000	***				
Dominance	-0.003	0.000	953.500	-5.602	0.000	***				
EN_MI:EN_W1len	0.001	0.000	205.900	2.067	0.040	*				
EN_MI:EN_W2len	0.000	0.000	206.000	-0.133	0.895					
EN_MI:EN_W1freq	-0.001	0.001	205.900	-0.658	0.511					
EN_MI:EN_W2freq	0.000	0.001	206.000	-0.217	0.829					
EN_MI:Dominance	0.000	0.000	32,350.000	-3.301	0.001	***				
Dominance:EN_W1len	0.000	0.000	32,350.000	-0.808	0.419					
Dominance:EN_W2len	0.000	0.000	32,350.000	-3.253	0.001	**				
Dominance:EN_W1freq	0.000	0.000	32,350.000	3.062	0.002	**				
Dominance:EN_W2freq	0.000	0.000	32,350.000	2.513	0.012	*				

Table 24. Model ESL-2b: NNS LDT Accuracy rate by English MI, word 1 length, word 2 length, word 1 log frequency, word 2 log frequency, language dominance, and two-way interactions involving English MI and language dominance.*

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.100	0.065	421.400	1.529	0.127		0.002	0.044	0.003	0.054
EN_MI	0.001	0.009	206.800	0.060	0.952					
EN_W1len	0.005	0.004	384.500	1.317	0.189					
EN_W2len	-0.005	0.003	374.400	-1.757	0.080	.				
EN_W1freq	0.083	0.010	368.400	7.978	0.000	***				
EN_W2freq	0.080	0.009	387.500	9.200	0.000	***				
Dominance	0.005	0.000	22,990.000	11.189	0.000	***				
EN_MI:EN_W1len	0.000	0.001	206.000	-0.442	0.659					
EN_MI:EN_W2len	0.000	0.000	206.000	0.286	0.775					
EN_MI:EN_W1freq	0.001	0.001	206.000	0.526	0.600					
EN_MI:EN_W2freq	-0.001	0.001	206.000	-0.436	0.663					
EN_MI:Dominance	0.000	0.000	32,350.000	-0.603	0.547					
Dominance:EN_W1len	0.000	0.000	32350.000	-0.009	0.993					
Dominance:EN_W2len	0.000	0.000	32350.000	1.979	0.048	*				
Dominance:EN_W1freq	0.000	0.000	32350.000	-6.645	0.000	***				
Dominance:EN_W2freq	-0.001	0.000	32350.000	-8.852	0.000	***				

Table 25. Model ESL-3: NNS LDT Log Winsorized RT by English MI, word 1 and 2 length, word 1 and 2 log frequency, Chinese MI, word 1 and 2 length, word 1 and 2 log frequency, group, condition, and two-way interactions involving group.

	Fixed effects						*Random effects*			
							By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	-0.016	0.047	762.000	-0.346	0.730		0.017	0.130	0.001	0.026
EN_MI	-0.001	0.001	589.000	-1.836	0.067	.				
EN_W1len	0.001	0.002	589.000	0.366	0.715					
EN_W2len	0.001	0.002	589.000	0.314	0.754					
EN_W1freq	-0.017	0.007	589.000	-2.276	0.023	*				
EN_W2freq	-0.006	0.006	589.000	-1.092	0.275					
CH_MI	0.000	0.001	589.000	-0.208	0.835					
CH_W1strk	0.000	0.001	589.000	0.101	0.920					
CH_W2strk	0.000	0.001	589.000	0.139	0.889					
CH_W1freq	-0.002	0.006	589.000	-0.402	0.688					
CH_W2freq	0.005	0.005	589.000	1.047	0.295					
Condition_EN-only	-0.005	0.011	589.000	-0.459	0.646					
Condition_CH-only	0.005	0.011	589.000	0.480	0.631					
Condition_Congruent	-0.014	0.013	589.000	-1.045	0.296					
Group2_ADV+	0.158	0.068	1,880.000	2.333	0.020	*				
Group3_ADV	0.314	0.055	1,880.000	5.729	0.000	***				
Group4_INT	0.399	0.051	1,880.000	7.850	0.000	***				
Group5_BEG+	0.585	0.052	1,830.000	11.172	0.000	***				
Group6_BEG	0.714	0.067	1,780.000	10.729	0.000	***				
EN_MI:Group2_ADV+	0.000	0.001	41,100.000	-0.080	0.936					

EN_MI:Group3_ADV	0.000	0.001	41,100.000	0.209	0.835
EN_MI:Group4_INT	0.001	0.001	41,100.000	0.732	0.464
EN_MI:Group5_BEG+	0.001	0.001	41,100.000	1.702	0.089 .
EN_MI:Group6_BEG	0.001	0.001	41,100.000	1.404	0.160
EN_W1len:Group2_ADV+	0.004	0.003	41,100.000	1.265	0.206
EN_W1len:Group3_ADV	0.003	0.002	41,100.000	1.217	0.224
EN_W1len:Group4_INT	0.005	0.002	41,100.000	2.374	0.018 *
EN_W1len:Group5_BEG+	0.003	0.002	41,100.000	1.413	0.158
EN_W1len:Group6_BEG	0.006	0.003	41,100.000	2.086	0.037 *
EN_W2len:Group2_ADV+	0.003	0.002	41,100.000	1.492	0.136
EN_W2len:Group3_ADV	0.006	0.002	41,100.000	3.317	0.001 ***
EN_W2len:Group4_INT	0.008	0.002	41,100.000	4.684	0.000 ***
EN_W2len:Group5_BEG+	0.008	0.002	41,100.000	4.456	0.000 ***
EN_W2len:Group6_BEG	0.012	0.002	41,100.000	5.086	0.000 ***
EN_W1freq:Group2_ADV+	-0.010	0.010	41,100.000	-0.979	0.327
EN_W1freq:Group3_ADV	-0.012	0.008	41,100.000	-1.528	0.127
EN_W1freq:Group4_INT	-0.018	0.007	41,100.000	-2.477	0.013 *
EN_W1freq:Group5_BEG+	-0.035	0.008	41,100.000	-4.604	0.000 ***
EN_W1freq:Group6_BEG	-0.053	0.010	41,100.000	-5.498	0.000 ***
EN_W2freq:Group2_ADV+	-0.010	0.008	41,100.000	-1.320	0.187
EN_W2freq:Group3_ADV	-0.013	0.006	41,100.000	-2.153	0.031 *
EN_W2freq:Group4_INT	-0.015	0.006	41,100.000	-2.620	0.009 **
EN_W2freq:Group5_BEG+	-0.026	0.006	41,100.000	-4.424	0.000 ***
EN_W2freq:Group6_BEG	-0.025	0.008	41,100.000	-3.369	0.001 ***
CH_MI:Group2_ADV+	0.000	0.001	41,100.000	0.277	0.782
CH_MI:Group3_ADV	0.000	0.001	41,100.000	0.343	0.732

CH_MI:Group4_INT	0.000	0.001	41,100.000	-0.495	0.621	
CH_MI:Group5_BEG+	-0.002	0.001	41,100.000	-2.520	0.012	*
CH_MI:Group6_BEG	-0.002	0.001	41,100.000	-2.177	0.030	*
CH_W1strk:Group2_ADV+	-0.001	0.001	41,100.000	-1.068	0.285	
CH_W1strk:Group3_ADV	-0.001	0.001	41,100.000	-0.990	0.322	
CH_W1strk:Group4_INT	-0.001	0.001	41,100.000	-0.986	0.324	
CH_W1strk:Group5_BEG+	0.000	0.001	41,100.000	-0.879	0.379	
CH_W1strk:Group6_BEG	0.001	0.001	41,100.000	0.921	0.357	
CH_W2strk:Group2_ADV+	0.000	0.001	41,100.000	-0.670	0.503	
CH_W2strk:Group3_ADV	0.000	0.001	41,100.000	0.082	0.935	
CH_W2strk:Group4_INT	-0.001	0.001	41,100.000	-0.930	0.352	
CH_W2strk:Group5_BEG+	-0.001	0.001	41,100.000	-0.920	0.357	
CH_W2strk:Group6_BEG	-0.001	0.001	41,100.000	-1.294	0.196	
CH_W1freq:Group2_ADV+	0.003	0.008	41,100.000	0.408	0.683	
CH_W1freq:Group3_ADV	-0.003	0.006	41,100.000	-0.496	0.620	
CH_W1freq:Group4_INT	-0.002	0.006	41,100.000	-0.413	0.680	
CH_W1freq:Group5_BEG+	-0.003	0.006	41,100.000	-0.433	0.665	
CH_W1freq:Group6_BEG	0.001	0.007	41,100.000	0.115	0.909	
CH_W2freq:Group2_ADV+	-0.002	0.006	41,100.000	-0.342	0.732	
CH_W2freq:Group3_ADV	-0.009	0.005	41,100.000	-1.829	0.067	.
CH_W2freq:Group4_INT	-0.014	0.005	41,100.000	-2.959	0.003	**
CH_W2freq:Group5_BEG+	-0.009	0.005	41,100.000	-1.925	0.054	.
CH_W2freq:Group6_BEG	-0.022	0.006	41,100.000	-3.537	0.000	***
Condition_EN-only:Group2_ADV+	0.000	0.015	41,100.000	0.004	0.997	
Condition_CH-only:Group2_ADV+	-0.014	0.015	41,100.000	-0.937	0.349	

Condition_Congruent:Group2_ADV+	-0.012	0.018	41,100.000	-0.678	0.498
Condition_EN-only:Group3_ADV	0.002	0.012	41,100.000	0.165	0.869
Condition_CH-only:Group3_ADV	-0.013	0.012	41,100.000	-1.081	0.280
Condition_Congruent:Group3_ADV	0.001	0.015	41,100.000	0.068	0.946
Condition_EN-only:Group4_INT	-0.001	0.011	41,100.000	-0.098	0.922
Condition_CH-only:Group4_INT	-0.005	0.011	41,100.000	-0.437	0.662
Condition_Congruent:Group4_INT	0.013	0.014	41,100.000	0.993	0.321
Condition_EN-only:Group5_BEG+	0.002	0.012	41,100.000	0.180	0.857
Condition_CH-only:Group5_BEG+	0.018	0.012	41,100.000	1.570	0.116
Condition_Congruent:Group5_BEG+	0.029	0.014	41,100.000	2.076	0.038 *
Condition_EN-only:Group6_BEG	0.014	0.015	41,100.000	0.940	0.347
Condition_CH-only:Group6_BEG	0.017	0.015	41,100.000	1.133	0.257
Condition_Congruent:Group6_BEG	0.045	0.018	41,100.000	2.560	0.010 *

Table 26. Model ESL-3: NNS LDT Accuracy rate by English MI, word 1 and 2 length, word 1 and 2 log frequency, Chinese MI, word 1 and 2 length, word 1 and 2 log frequency, group, condition, and two-way interactions involving group.*

							Random effects			
	Fixed effects						By Subject		By Item	
Parameters	Estimate	SE	df	t	p	sig.	Variance	SD	Variance	SD
(Intercept)	0.901	0.059	410.300	15.234	0.000	***	0.001	0.035	0.002	0.043
EN_MI	0.000	0.001	403.400	0.108	0.914					
EN_W1len	0.004	0.003	403.400	1.171	0.242					
EN_W2len	0.002	0.002	403.400	0.866	0.387					
EN_W1freq	0.008	0.010	403.400	0.790	0.430					
EN_W2freq	0.006	0.008	403.400	0.717	0.474					
CH_MI	0.000	0.001	403.400	-0.335	0.738					
CH_W1strk	0.000	0.001	403.400	0.172	0.864					
CH_W2strk	-0.001	0.001	403.400	-0.818	0.414					
CH_W1freq	-0.002	0.008	403.400	-0.246	0.806					
CH_W2freq	-0.001	0.007	403.400	-0.152	0.879					
Condition_EN-only	0.010	0.016	403.400	0.615	0.539					
Condition_CH-only	0.014	0.016	403.400	0.920	0.358					
Condition_Congruent	0.013	0.019	403.400	0.717	0.474					
Group2_ADV+	-0.045	0.067	37,250.000	-0.668	0.504					
Group3_ADV	-0.158	0.054	37,250.000	-2.897	0.004	**				
Group4_INT	-0.408	0.051	37,250.000	-8.082	0.000	***				
Group5_BEG+	-0.820	0.052	37,080.000	-15.806	0.000	***				
Group6_BEG	-0.766	0.066	36,890.000	-11.637	0.000	***				
EN_MI:Group2_ADV+	0.000	0.001	41,110.000	0.262	0.793					
EN_MI:Group3_ADV	0.001	0.001	41,110.000	0.735	0.462					

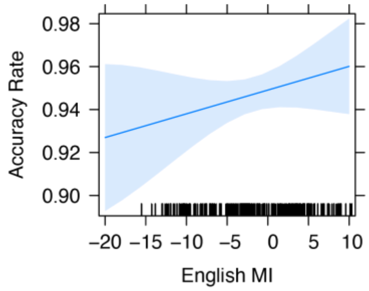
EN_MI:Group4_INT	0.001	0.001	41,110.000	0.975	0.329	
EN_MI:Group5_BEG+	0.002	0.001	41,110.000	1.824	0.068	.
EN_MI:Group6_BEG	0.003	0.001	41,110.000	2.386	0.017	*
EN_W1len:Group2_ADV+	-0.002	0.003	41,110.000	-0.468	0.640	
EN_W1len:Group3_ADV	0.003	0.003	41,110.000	0.987	0.324	
EN_W1len:Group4_INT	0.002	0.003	41,110.000	0.838	0.402	
EN_W1len:Group5_BEG+	0.004	0.003	41,110.000	1.600	0.110	
EN_W1len:Group6_BEG	-0.002	0.003	41,110.000	-0.630	0.529	
EN_W2len:Group2_ADV+	-0.002	0.003	41,110.000	-0.914	0.361	
EN_W2len:Group3_ADV	-0.003	0.002	41,110.000	-1.236	0.216	
EN_W2len:Group4_INT	-0.005	0.002	41,110.000	-2.246	0.025	*
EN_W2len:Group5_BEG+	-0.012	0.002	41,110.000	-5.922	0.000	***
EN_W2len:Group6_BEG	-0.008	0.003	41,110.000	-2.989	0.003	**
EN_W1freq:Group2_ADV+	0.005	0.012	41,110.000	0.421	0.673	
EN_W1freq:Group3_ADV	0.021	0.009	41,110.000	2.233	0.026	*
EN_W1freq:Group4_INT	0.026	0.009	41,110.000	2.946	0.003	**
EN_W1freq:Group5_BEG+	0.082	0.009	41,110.000	9.048	0.000	***
EN_W1freq:Group6_BEG	0.066	0.011	41,110.000	5.764	0.000	***
EN_W2freq:Group2_ADV+	0.010	0.009	41,110.000	1.101	0.271	
EN_W2freq:Group3_ADV	0.005	0.007	41,110.000	0.657	0.511	
EN_W2freq:Group4_INT	0.026	0.007	41,110.000	3.880	0.000	***
EN_W2freq:Group5_BEG+	0.038	0.007	41,110.000	5.485	0.000	***
EN_W2freq:Group6_BEG	0.050	0.009	41,110.000	5.619	0.000	***
CH_MI:Group2_ADV+	0.001	0.001	41,110.000	0.393	0.694	
CH_MI:Group3_ADV	0.000	0.001	41,110.000	0.465	0.642	
CH_MI:Group4_INT	0.001	0.001	41,110.000	0.618	0.536	

CH_MI:Group5_BEG+	0.000	0.001	41,110.000	0.440	0.660	
CH_MI:Group6_BEG	-0.001	0.001	41,110.000	-0.425	0.671	
CH_W1strk:Group2_ADV+	0.000	0.001	41,110.000	0.398	0.690	
CH_W1strk:Group3_ADV	0.000	0.001	41,110.000	-0.018	0.985	
CH_W1strk:Group4_INT	0.000	0.001	41,110.000	0.108	0.914	
CH_W1strk:Group5_BEG+	0.002	0.001	41,110.000	2.638	0.008	**
CH_W1strk:Group6_BEG	0.001	0.001	41,110.000	0.759	0.448	
CH_W2strk:Group2_ADV+	0.000	0.001	41,110.000	0.329	0.742	
CH_W2strk:Group3_ADV	0.002	0.001	41,110.000	2.326	0.020	*
CH_W2strk:Group4_INT	0.001	0.001	41,110.000	1.855	0.064	.
CH_W2strk:Group5_BEG+	0.004	0.001	41,110.000	5.500	0.000	***
CH_W2strk:Group6_BEG	0.001	0.001	41,110.000	1.711	0.087	.
CH_W1freq:Group2_ADV+	0.004	0.009	41,110.000	0.418	0.676	
CH_W1freq:Group3_ADV	-0.004	0.007	41,110.000	-0.524	0.600	
CH_W1freq:Group4_INT	0.008	0.007	41,110.000	1.180	0.238	
CH_W1freq:Group5_BEG+	-0.002	0.007	41,110.000	-0.239	0.811	
CH_W1freq:Group6_BEG	0.014	0.009	41,110.000	1.614	0.107	
CH_W2freq:Group2_ADV+	-0.005	0.007	41,110.000	-0.650	0.516	
CH_W2freq:Group3_ADV	0.006	0.006	41,110.000	1.068	0.285	
CH_W2freq:Group4_INT	0.015	0.006	41,110.000	2.689	0.007	**
CH_W2freq:Group5_BEG+	0.025	0.006	41,110.000	4.382	0.000	***
CH_W2freq:Group6_BEG	0.015	0.007	41,110.000	2.105	0.035	*
Condition_EN-only:Group2_ADV+	-0.006	0.018	41,110.000	-0.326	0.745	
Condition_CH-only:Group2_ADV+	-0.016	0.018	41,110.000	-0.893	0.372	
Condition_Congruent:Group2_ADV+	-0.016	0.021	41,110.000	-0.737	0.461	
Condition_EN-only:Group3_ADV	-0.020	0.015	41,110.000	-1.367	0.172	

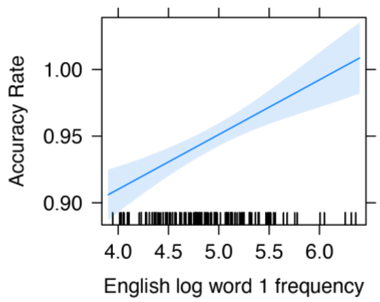
Condition_CH-only:Group3_ADV	-0.016	0.014	41,110.000	-1.129	0.259
Condition_Congruent:Group3_ADV	-0.011	0.017	41,110.000	-0.640	0.522
Condition_EN-only:Group4_INT	-0.016	0.014	41,110.000	-1.149	0.251
Condition_CH-only:Group4_INT	-0.023	0.013	41,110.000	-1.708	0.088 .
Condition_Congruent:Group4_INT	-0.018	0.016	41,110.000	-1.116	0.264
Condition_EN-only:Group5_BEG+	-0.019	0.014	41,110.000	-1.369	0.171
Condition_CH-only:Group5_BEG+	-0.030	0.014	41,110.000	-2.179	0.029 *
Condition_Congruent:Group5_BEG+	-0.020	0.016	41,110.000	-1.225	0.220
Condition_EN-only:Group6_BEG	-0.025	0.018	41,110.000	-1.437	0.151
Condition_CH-only:Group6_BEG	-0.048	0.017	41,110.000	-2.751	0.006 **
Condition_Congruent:Group6_BEG	-0.022	0.021	41,110.000	-1.037	0.300

Figure 8. Effects of English and Chinese language variables, Group, and Condition on Accuracy Rate for English NNSs. (Model ESL-3).*

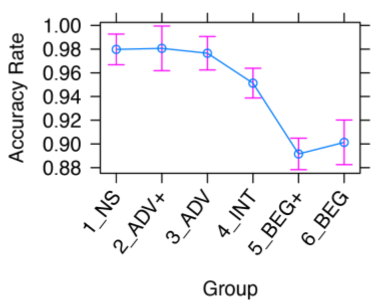
a.) English MI



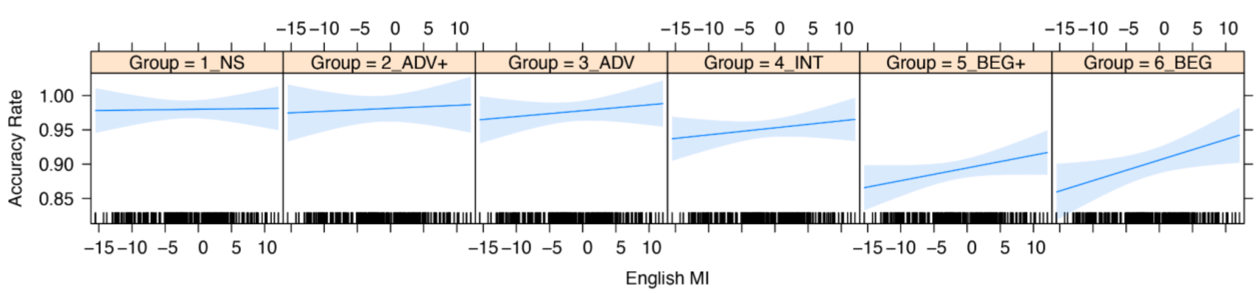
b.) English log word 1 frequency



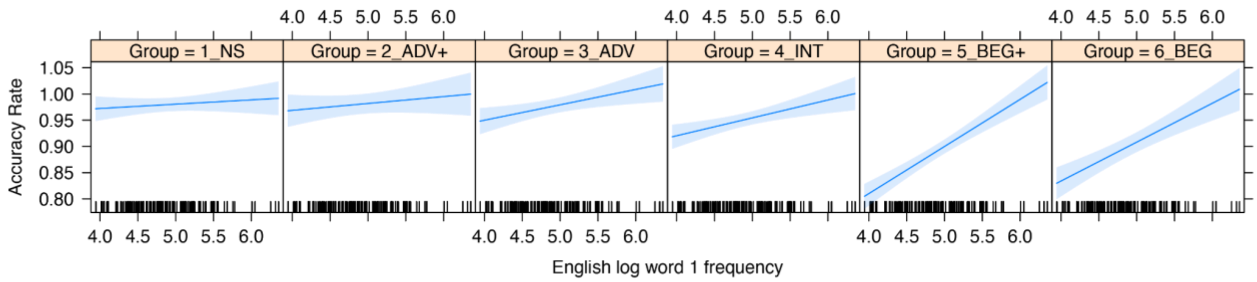
c.) Group



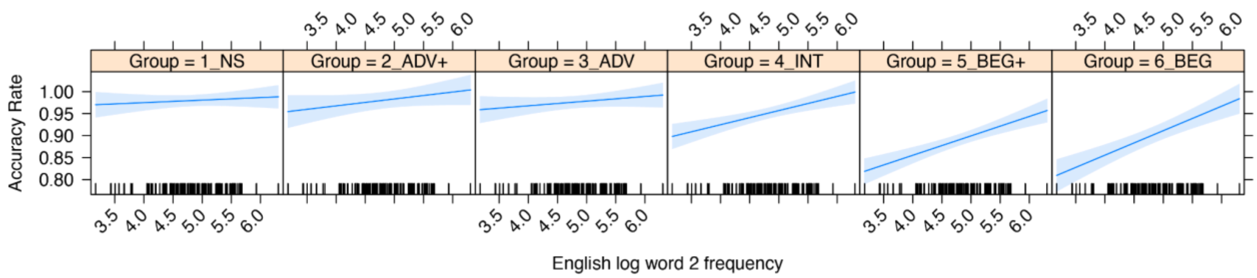
d.) English MI : Group



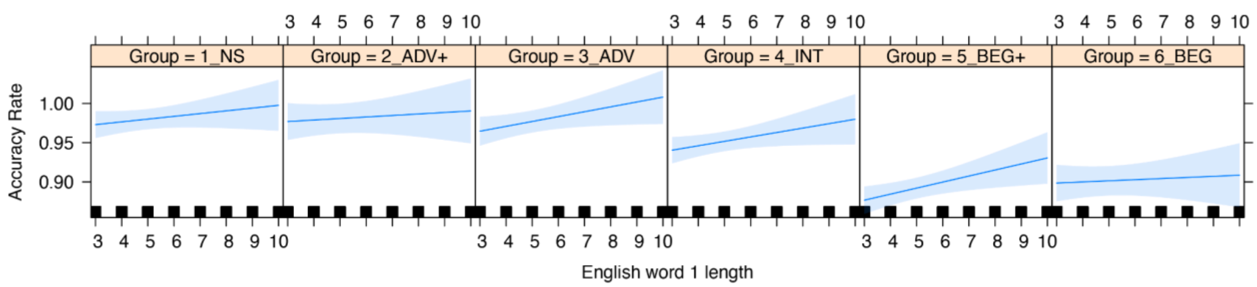
e.) English log word 1 frequency : Group



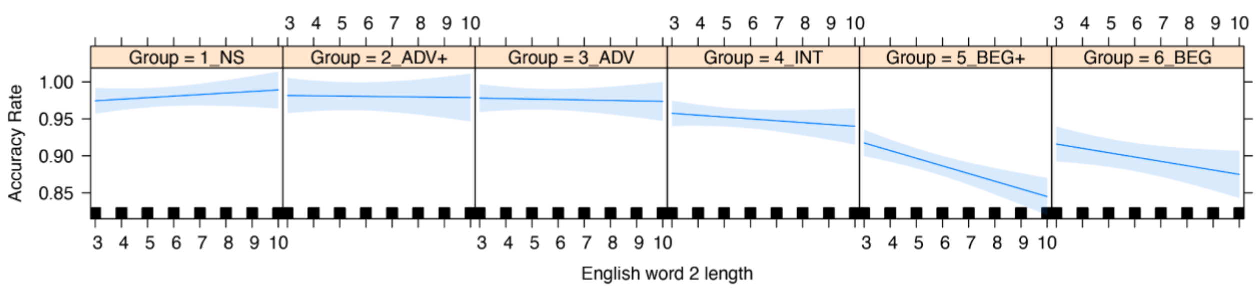
f.) English log word 2 frequency : Group



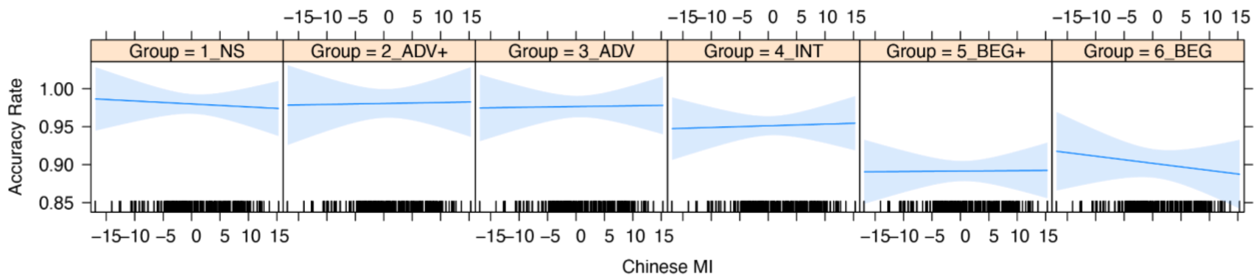
g.) English word 1 length : Group



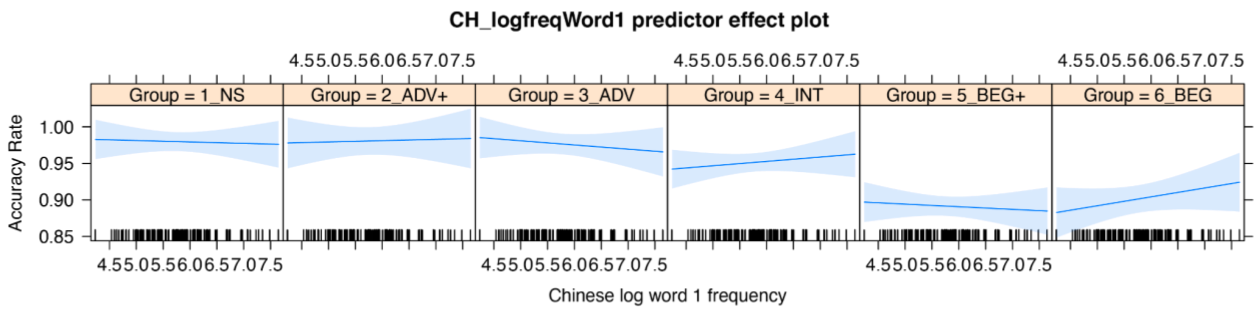
h.) English word 2 length : Group



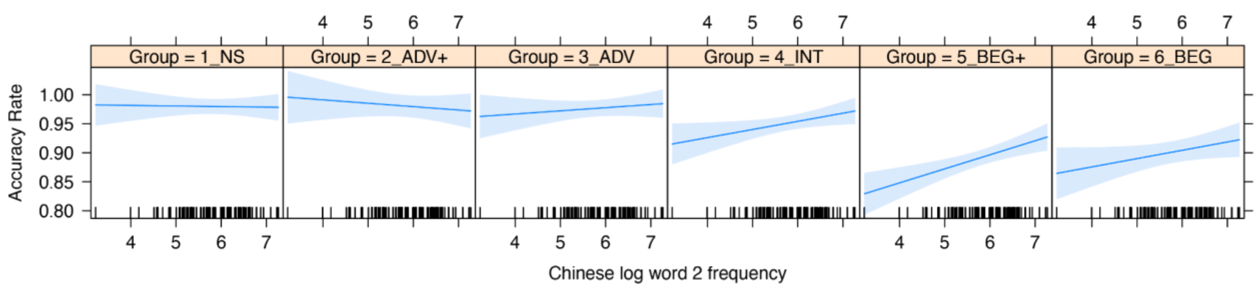
i.) Chinese MI : Group



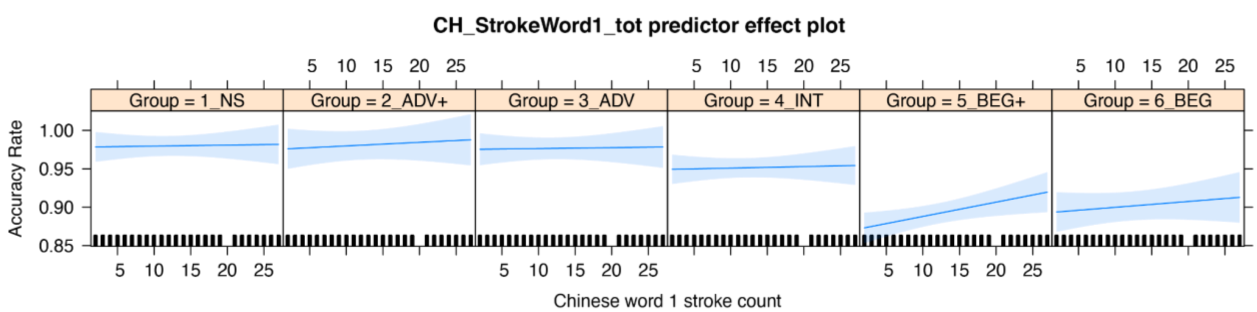
j.) Chinese log word 1 frequency : Group



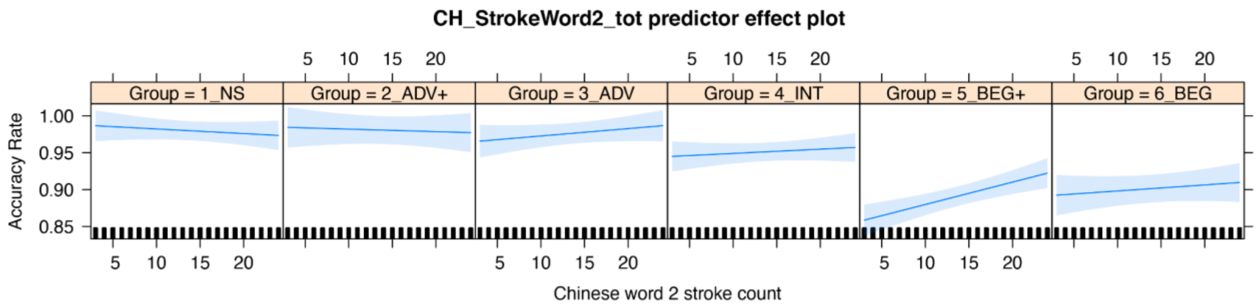
k.) Chinese log word 2 frequency : Group



l.) Chinese word 1 stroke count : Group



m.) Chinese word 2 stroke count : Group



n.) Condition : Group