HOW DOES MODIFICATION AFFECT THE PROCESSING OF FORMULAIC LANGUAGE?

EVIDENCE FROM L1 AND L2 SPEAKERS OF CHINESE

BY

SHANG JIANG

A thesis

Submitted to the Victoria University of Wellington

in fulfilment of the requirements for the degree of

Doctor of Philosophy

in Applied Linguistics

Victoria University of Wellington

2021

Abstract

It has been well documented that formulaic language (FL) – strings of language above the word level in most cases, such as binomials, collocations, n-grams, idioms – enjoys a processing advantage over novel language (e.g., Arnon & Snider, 2010; Carrol & Conklin, 2020; Hallin & Van Lancker Sidtis, 2017; Siyanova-Chanturia, Conklin, & Schmitt, 2011a; Siyanova-Chanturia, Conklin, & van Heuven, 2011b; Tremblay, Derwing, Libben, & Westbury, 2011). The majority of these studies focused on formulaic sequences (FSs) in their original form. In natural language use, however, many FSs are modified with words intervening in-between the individual constituents (e.g., *provide information* → *provide some of the information*, see Vilkaitė, 2016a). Whether or not the processing advantage can be extended to modified FSs remains poorly investigated. In this thesis, three reading studies were conducted to address this gap from the perspective of Chinese FL processing, in which the influence of phrase frequency, modification degree, age, and L2 proficiency, were explored and discussed in detail.

Study 1 recorded the eye movements of L1 Chinese adults when reading sentences embedded with frequent collocations and infrequent controls. The results suggested a significant processing advantage for collocations over relative controls. Critically, the processing advantage for collocations in their original form extended to their short- (with two Chinese characters inserted in the middle of the phrase) and long-insertion forms (with four Chinese characters inserted in the middle of the phrase). The processing advantage was largely observed in the whole phrase in the late processing stage.

Study 2 and Study 3 borrowed the materials from Study 1 and applied them in a self-paced reading experiment with L2 Chinese learners from Japan and Thailand of different language proficiencies, and L1 Chinese children at Grades Three and Six, respectively. The results of Study 2 revealed a significant processing advantage for collocations over respective controls in L2 learners. More crucially, the processing advantage for L2 collocations persisted in their short- and long-insertion forms. L2 proficiency and L1 background were also found to play a role in L2 FS processing.

The results of Study 3 suggested a significant processing advantage for collocations over respective controls in L1 children. More importantly, the processing advantage for collocations remained in their short- and long-insertion forms in sixth graders; and remained in their short-insertion form in third graders. Furthermore, age was found to play a role in FS processing in children.

Contributions to the current literature and methodologies are discussed at the end of the thesis.

Acknowledgements

There is no doubt that this research journey has been the most wonderful one I have ever had. This journey began in December 2017 and after three years and a half, it has now come to the end. This journey has been filled with anticipation, inspiration, and perspiration. Without the help and support of certain great people and things, this journey could be less enjoyable and more difficult to complete. So here, I would like to take this opportunity to thank these great people and things.

First, I wish to express my deepest gratitude to my supervisors, Dr Anya Siyanova and Associate Professor Gina Grimshaw. Anya, thank you for setting a superb example for me of what an excellent young scholar should be like. I can always learn a lot from you, not only in academic research but also in conducting oneself. Your warm words can always give me confidence and your suggestions can always inspire me. I am more than happy to have you as my supervisor, friend, and also collaborator. Gina, thank you for always being there for me, especially when I was preparing for the proposal as well as the oral defence. I really enjoy the in-depth discussion with you on my research as well as the pertinent presentation. If I can turn back the clock or hold it still, I would beg for more time for our discussion.

Second, I wish to express my sincere gratitude to Victoria University of Wellington, who has kindly provided me with such a great opportunity to study on a doctoral scholarship, which saved me from financial concern during the study. Victoria University of Wellington has also offered me helpful and adequate learning and research resources, which facilitated the progress of the study. For example, Student Learning helped me improve my academic writing and speaking skills. The Statistical Consultant, Dr Lisa Woods helped me improve my skills in data analysis and visualization. Faculty of Humanities and Social Sciences financially supported the data collection of my research, in both New Zealand and China. Without this generous help, my PhD study would not go as smoothly as it was.

Third, I am particularly grateful for the beautiful sceneries in New Zealand, especially the Queens Wharf and Botanic Garden in Wellington, where I go very often to release, recharge, and reflect. These places are like my close friends in spirit. Every time I visit them, I would feel good and get ready to face challenges on the journey.

Fourth, I would like to take this opportunity to thank all my colleagues and friends. With the companion of these cool people, my PhD journey became colourful, with cheers and laughter.

Last and most importantly, I feel enormously grateful for having the best mom in the world, Lily Cao, who is always patient, caring, supportive, positive, and understanding to me. Lily, our life is getting better and our future is becoming brighter, and I will make it better for us!

Every end is a new beginning, Shang, so, keep going!

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

1.1 Setting the scene

"*Our language does not expect us to build everything starting with lumber, nails, and blueprint. Instead, it provides us with an incredibly large number of prefabs, which have the magical property of persisting, even when we knock some of them apart and put them together in unpredictable ways.*"

(Dwight Bolinger, 1979, p. 96)

We humans can be quite indolent when using language. We are reluctant to comprehend and produce an expression element by element, and unwilling to check the grammaticality of the expression every time we encounter it, even though we have gained sound knowledge for relevant vocabularies and grammar rules. For example, when greeting people around Christmas time, we would say '*Merry Christmas*' without hesitation or thinking; we seldom generate and utter '*Cheerful Christmas*', although this phrase conveys the idea we want to express. And when your friend is going to give a public performance at the local theatre, you would go and cheer him/her up, saying '*break a leg*' rather than '*hope you succeed*', without considering the exact meaning of '*leg*' in this phrase or whether it is grammatical to go after the verb '*break*'. It seems that we are not always creative language users.

However, this indolence or uncreativity is strongly and widely recommended by the academic community (e.g., Becker, 1975; Conklin & Schmitt, 2008; Ellis, 1996; Fillmore, 1979; Firth, 1957; Jespersen, 1924; Lamb, 1999; Nattinger & DeCarrico, 1992; Peters, 1983; Schmitt, 2004; Siyanova-Chanturia & Pellicer-Sánchez, 2019; Tomasello, 2003; Wray, 2002). In fact, as Jespersen (1924) put it, "a language would be a difficult thing to handle if its speakers had the burden imposed on them of remembering every little item separately" (p. 21). This suggests that the indolence or uncreativity in language processing, learning, and use is, in fact, necessary and expected.

Indeed, there is no need for any shame or qualms about the indolence or uncreativity in language use. This is because we have an excellent helper in our mental lexicon – formulaic language, which always provides us with "ready-made frameworks" (Becker, 1975, p. 17), with which we can cope with the (ever-changing) situations confidently and comfortably, without further efforts such as creating or analysing.

With formulaic language in our mental lexicon, we can always know not only a linguistic element but also (and more crucially) the company it keeps (e.g., Firth, 1957), even though the company is not immediately adjacent to the element (e.g., Bolinger, 1979). For example, when seeing or hearing '*seek*', we are expecting '*help*' in the meantime, even if some words intervene between (e.g., '*seek your older brother's help*'). This phenomenon suggests that formulaic language exists in our mental lexicon, and even when it is modified, we are still able to recognize it. The current thesis is thus going to explore the psychological reality of modified formulaic language in our mental lexicon from the perspective of online processing.

Online processing in the current thesis refers to the moment-to-moment cognitive process during reading, listening, or speaking. No matter what the modality is, the key tenet is that the cognitive activity – comprehension or production – should be finished under time pressure.

The online processing of formulaic language has been researched for a relatively long time since the 1970s (for a review, see Siyanova-Chanturia & van Lancker Sidtis, 2019). However, the early investigations (e.g., Cacciari & Tabossi, 1988; Estill & Kemper, 1982; Gibbs, 1980; Gibbs & Gonzales, 1985; Glass, 1983; Kemper, 1986; Ortony, Schallert, Reynolds, & Antos, 1978; Swinney & Cutler, 1979; Van Lancker, Canter, & Terbeek, 1981) have largely focused on figurative (non-compositional) formulaic language, such as idioms (e.g., *break the ice*). Literal (compositional) formulaic language, such as collocations (e.g., *look forward to*) had not been investigated until the 2000s (e.g., Jiang & Nekrasova, 2007; Kapatsinski & Radicke, 2009; Schmitt, Grandage, & Adolphs, 2004; Schmitt & Underwood, 2004; Siyanova-Chanturia & Schmitt, 2008; Sosa & MacFarlane, 2002; Underwood, Schmitt, & Galpin, 2004). More importantly, the majority of existing studies focused on FL in its original form, with the processing of FL in its modified form being unexplored to a great extent.

Besides, although studies on formulaic language processing have proliferated recently, most of them have focused on adults, with children being largely disregarded. Given that formulaic language plays an important role in first language (L1) acquisition and development (e.g., Clark, 1974; Lieven, Behrens, Speares, & Tomasello, 2003; Nelson, 1973; Peters, 1977), a better understanding of how it is processed by children is urgently needed (e.g., Jiang, Jiang, & Siyanova-Chanturia, 2020; Wray, 2002).

Furthermore and in a similar vein, although formulaic language has been acknowledged as the key in second language (L2) learning (e.g., Ellis, 2001; Tomasello, 2003; Wong-Fillmore, 1979), the processing of L2 formulaic language has been less studied than L1 formulaic language. Among the existing studies on L2 formulaic language processing, researchers have found a number of factors that may influence the processing. These potential factors include language proficiency (e.g., Shantz, 2017; Siyanova-Chanturia, Conklin, & van Heuvan, 2011b; Vilkaitė & Schmitt, 2019) and L1 backgrounds (e.g., Carrol, Conklin, & Gyllstad, 2016; Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2011). However, how these potential factors affect L2 formulaic language processing remains extremely unclear.

Finally, existing studies have by and large focused on the processing of formulaic language in alphabetic languages, such as English, with little attention paid to other languages, such as Chinese. As one of the most spoken languages in the world (Statista, 2020), Chinese boasts some unique linguistic properties that are distinct from English and other alphabetic languages. Thus, extending the findings pertinent to English and other alphabetic languages to Chinese without verification is problematic at best. Among the existing studies on Chinese formulaic language processing, the majority have looked at figurative formulaic sequences, such as idioms (e.g., Liu, Li, Shu, Zhang, & Chen, 2010; Shen, Wang, & Liu, 2017; Yu et al., 2016; Zhang, Yang, Gu, & Ji, 2013; Zhou, Zhou, & Chen, 2004). While idioms are perhaps the most salient and quintessential example of formulaic language, many idioms are admittedly rather infrequent. How more frequent exemplars, such as collocations, are processed in L1 and L2 Chinese remains largely unstudied.

To address the above gaps, I carried out three studies to investigate the online processing of modified Chinese literal formulaic language in L1 adults and children, as well as L2 learners from distinct L1 backgrounds and of different language proficiencies.

1.2 Aims of the thesis

The current thesis is an empirical investigation on the topic of the online processing of modified Chinese formulaic language, with several aims as follows:

The primary aim is to examine the psychological reality of modified formulaic language in the mental lexicon of L1 users and L2 learners. The psychological reality in the current thesis refers to the sensitivity to frequency information that is above the word level. This definition is in line with existing research (e.g., Durrant & Doherty, 2010; Ellis et al., 2008, 2009; Kapatsinski & Radicke, 2009; Kong et al., 2016). For example, as suggested by

Durrant and Doherty (2010), phrases of different frequencies are not stored in the same way in our mind, with higher-frequency phrases more entrenched relative to lower-frequency ones. The different degrees of entrenchment enable us to distinguish phrases of different frequencies. This ability, or in other words, sensitivity to phrase frequencies can be indicated by the processing speed (e.g., Durrant & Doherty, 2010). That is, if a higher-frequency phrase is processed faster than a lower-frequency one, this may be indicative of our sensitivity to phrase frequencies.

Specifically, I am interested in whether or not L1 adults, children, and L2 learners are able to identify formulaic sequences in their modified form, and whether these populations are able to distinguish, in cognitive terms, between modified formulaic sequences (e.g., *包虾仁饺子 (**make** prawn **dumplings**)*) and novel phrases (e.g., *卖虾仁饺子 (**sell** prawn **dumplings**)*) during reading.

The current thesis further aims to investigate the processing of modified formulaic language versus novel language. For instance, if modified formulaic language is processed faster than novel language, this processing advantage could be due to the fact that formulaic language occurs frequently in comparison with novel language. Exposure makes language users and learners more familiar with formulaic language relative to novel language, and familiarity further facilitates the processing of formulaic language. Even when formulaic language is modified, facilitation (e.g., faster reading) remains.

In addition, this thesis aims to explore how modification degree, age, L2 proficiency, and L1 background might affect the online processing of modified formulaic language. To be more specific, is the processing of formulaic language with major modification similar to formulaic language with minor modification? Is formulaic language processing in younger children similar to that in older children? Is formulaic language processing in lower-proficiency L2 learners similar to that in higher-proficiency learners? Are there any differences in formulaic language processing between L2 learners from distinct L1 backgrounds, such as Japanese and Thai? It is important to explore the role of these factors. This is because the majority of existing studies focused on only one level of modification degree (e.g., Geeraert, Baayen, & Newman, 2017; Haeuser, Baum, & Titone, 2020; Kyriacou, Conklin, & Thompson, 2020; McGlone, Glucksberg, & Cacciari, 1994; Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019), one age group of children (e.g., Hallin & Van Lancker Sidtis, 2017; Skarabela, Ota, O'Connor, & Arnon, 2021), one proficiency group (e.g., Hernández, Costa, & Arnon, 2016; Jiang &

Nekrasova, 2007; Siyanova-Chanturia & Janssen, 2018; Yi, Lu, & Ma, 2017), and one L1 background (e.g., Carrol & Conklin, 2014a, 2017; Carrol et al., 2016). It remains largely unclear in the current literature whether and how different the processing of FL is with different modification degrees; whether and how different the processing of FL is among different age groups.; and, whether and how different the processing of FL is in L2 learners of different proficiencies and from distinct L1 backgrounds.

Furthermore, this thesis aims to test whether the methodologies adopted for the current investigation are suitable measures for the examination of the psychological reality of formulaic language in its modified form, from the perspective of online processing.

1.3 Structure of the thesis

The current thesis attempts to explore and answer the questions proposed above in eight chapters. The first four chapters introduce and review what has been done in the processing of (modified) formulaic language, followed by three experimental studies that explored this topic with different populations. The last chapter discusses and interprets the main findings obtained from the three experimental studies in light of pertinent research and theories.

To be more specific, Chapter 2 gives a general introduction of the concept of formulaic language on its definition, pervasiveness, diversity, characteristics, and role in language learning and processing.

In Chapter 3, I introduce the topic and review what has been found in existing studies on formulaic language processing in L1 adults, L2 learners, as well as children. Despite the scarcity, research into the processing of modified formulaic language and the processing of Chinese formulaic language, is also reported. The rationale behind the processing differences between formulaic language and novel language is discussed at the end of the chapter.

In Chapter 4, I introduce two of the most highly recommended and widely used research methods for formulaic language processing: self-paced reading and eye-tracking, which are adopted in the current thesis. The introduction covers the key concepts, measurement, and advantages of using these methods.

Based on the eye-tracking methodology, Chapter 5 presents an investigation of the processing of modified Chinese formulaic sequences (Verb+Noun collocations and Adjective+Noun collocations) versus novel phrases in L1 adults. This investigation further explores how modification degree affects formulaic language processing.

Based on the self-paced reading methodology, the subsequent chapters report two studies on the processing of modified Chinese formulaic sequences versus novel phrases in

L2 learners (Chapter 6) and children (Chapter 7), respectively. The two studies further explore how modification degree, age, L2 proficiency, and L1 background affect formulaic language processing, respectively.

Finally, I summarize the findings that emerged from the three experiments and discuss what these findings mean in language processing, learning, and use in the real world. Then, I point out what and how this thesis contributes to the current literature and relevant theories, models, and approaches in language learning, processing, and use. Before the end of the thesis, I draw conclusions for the current investigation and provide possible directions for future research (Chapter 8).

Chapter 2: Formulaic Language as a Linguistic Phenomenon

2.1 Introduction

Before investigating formulaic language processing, we first need to have a general understanding of what formulaic language is. Thus, in this chapter, formulaic language will be introduced in detail. First, I will provide its definition and explain the relationship between formulaic language and other relevant terms, such as formulaic sequence, multi-word expression. Following this, I will introduce its principal and fundamental characteristics in frequency, familiarity, predictability, fixedness, and phonology. At the end of this chapter, I will emphasize its role in first (L1) and second language (L2) learning and processing.

2.2 Terminology and definition

Formulaic language (FL) refers to strings or sequences of "letters, words, sounds, or other elements, contiguous or non-contiguous, of any length, size, frequency, degree of compositionality, literality/figurativeness, abstractness and complexity … that enjoy a degree of conventionality or familiarity among (typical) speakers of a language community or group, and that hold a strong relationship in communicating meaning" (Siyanova-Chanturia & Pellicer-Sánchez, 2019, p. 5). It is a "single choice, even though it might appear to be analyzable into segments" (Sinclair, 1991, p.110), and can circumvent the entire syntactic analyses (e.g., Bateson, 1975; Jespersen, 1976; Wray, 2002).

The phenomenon of FL has been noticed and investigated for a long time, although miscellaneous terms were used in these investigations. For example, more than a century ago, Saussure (1916) observed a few successions of very common words and phrases that language users always refuse to analyse. He proposed "a simple unit" (p. 177) to describe this interesting phenomenon. Firth (1937) argued that the unit of our speech is not a word but a "holophrase" (p. 83). A few years later, Hymes (1968) found that a vast portion of "recurrent sequences" (p. 126) exist in our verbal behaviour. Jespersen (1976) advocated that our verbal behaviour involves not only free combinations but also "formulas" (p. 88). Bolinger (1979) maintained that "our language does not expect us to build everything starting with lumber, nails and blueprint. Instead, it provides us with an incredibly large number of prefabs" (p. 96). Similarly, Peters (1983) argued that language users often build up convenient constructions and then file them away as "chunks" for future use (p. 80). With sufficient reserves of "formulaic utterances" (Fillmore, 1979, p. 92), people are able to get along (well) with the target language; with the mastery of “lexical phrases” (Nattinger & DeCarrico, 1992, p. 32), we can speak the target language with more fluency. Apart from these, researchers used other terms, such as “big words” (Ellis, 1996, p. 111), “formulaic sequences” (e.g., Wray, 2002), “multiword units” (e.g., Wray, 2002), “multiword expressions” (e.g., Wray, 2002), and “morpheme equivalent units” (e.g., Wray, 2008) to refer to the same linguistic phenomenon in their investigations.

Despite the variety of terms and definitions that have been used, *formulaic language* (FL) is chosen for the current thesis to describe the linguistic phenomenon mentioned above. First, according to Buerki (2020), the concept of this phenomenon “cannot be fixed at any particular structural level and should incorporate elements specified at varying levels of schematicity” (p. 103). All the terms mentioned above (except for FL), however, cannot embody sequences of letters, such as expletives (e.g., *damn!*), or exclamations (e.g., *hooray!*). Second, as suggested in Siyanova-Chanturia and Pellicer-Sánchez (2019), FL is considered as “the most inclusive term” (p. 2), which not only includes sequences of words but also sequences of letters. In this case, FL is considered to be the most appropriate umbrella term to embrace all cases of this linguistic phenomenon.

I will also borrow the term *formulaic sequence* (FS) from Wray (2002) to refer to various instances of FL, which are above the word level.

2.3 Pervasiveness and diversity

FL is a ubiquitous linguistic phenomenon. Many studies have identified a great number of FSs that exist in the spoken (e.g., Altenberg, 1998; Biber, Johansson, Leech, Conrad, & Finegan, 1999; Erman & Warren, 2000; Foster, 2001; Glucksberg, 1989; Pawley & Syder, 1983; Pollio, Barlow, Fine, & Pollio, 1977) and written discourse of adult L1 users (e.g., Biber et al., 1999; Erman & Warren, 2000; Howarth, 1998). For example, Altenberg (1998) estimated that over 80% of the utterances in the London-Lund corpus (Svartvik, 1990) are made up of FSs. And, according to Erman and Warren (2000), FL accounts for “58.60%” in spoken language and “52.30%” in written language (p. 37). These studies suggest that L1 users frequently use FL in their speaking and writing.

A number of studies have also identified the use of FSs in the spoken (e.g., Foster, 2001; Paquot & Granger, 2012) and written discourse of L2 learners (e.g., Granger, 1998; Howarth, 1998; Paquot & Granger, 2012). For example, Howarth (1998) estimated that about one-quarter of L2 writing consists of conventional combinations. Foster (2001) found that approximately one-fifth of L2 speech (in task-based production) was made up of FSs. These studies suggest that although FL is less pervasive in L2 than in L1 context, L2 learners still frequently use it in their speaking and writing.

FSs also abound in the early production of L1 children (e.g., Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Pine, & Barnes, 1992; Lieven, Salomo, & Tomasello, 2009; Nelson, 1973; Pine & Lieven, 1993; Pollio & Pollio, 1974; Reuterskiöld & Van Lancker Sidtis, 2012). For example, Lieven et al. (1992) found that up to "44%" of the first 100 words are FSs (p. 302). Reuterskiöld and Van Lancker Sidtis (2012) estimated that approximately 25% of child talk is made up of FSs. These studies suggest that children could also frequently use FL.

FL abounds not only in English discourse but also in Chinese. For example, Wong (2012) estimated that, on average, L1 and L2 speakers articulate "28.60" and "20.80" FSs per 50 words, respectively (p. 114). All estimates mentioned above suggest that FL abounds in the spoken and written discourse of adult and child L1 and L2 speakers.

FL is a common linguistic phenomenon comprised of various categories. It embraces exclamations (sequences of letters or words that express sudden pain, surprise, anger, excitement, happiness, or other emotion, e.g., *oh my gosh!*), collocations (co-occurring word pairs, e.g., *make a wish*), binomials (sequences of X and Y where a specific word order is highly preferred, e.g., *bride and groom*), n-grams (co-occurring word strings retrieved from corpora, e.g., *in the middle of*), phrasal verbs (Verb+Particle combinations with more than one meaning, e.g., *chew out*), idioms (non-compositional sequences with a figurative meaning, e.g., *a piece of cake*), similes (non-compositional sequences involving direct comparison of one thing with another of a different kind indicated by the words 'as' or 'like', e.g., *clear as crystal*), metaphors (non-compositional sequences involving direct comparison of one thing with another of a different kind without indicator words, e.g., *rock the boat*), discourse organizers (sequences that help connect ideas in a text, e.g., *not only … but also…*), proverbs and clichés (concrete and traditional sayings stating a general truth or piece of advice, e.g., *a cat has nine lives*; *actions speak louder than words*), social routine formulae (highly frequent expressions that are closely tied to types of recurrent social situations, e.g., *Nice to meet you!*), and so forth.

2.4 Characteristics

Miscellaneous as FSs are, they still share a number of key characteristics in some ways, such as frequency, familiarity, predictability, fixedness, modification, and phonology. Note that these basic features apply to most but not all FSs.

2.4.1 Frequency

As suggested in the previous section (2.3 pervasiveness and diversity), FL enjoys wide use in written and spoken discourse (e.g., Erman & Warren, 2000; Foster, 2001; Lieven et al., 1992). When describing the use of a language element, Sinclair and Renouf (1988) contended that we should not circumvent the (relative) frequency information of the element in corpora. According to corpus studies, compared with novel language (expressions that are constructed anew), FL boasts a higher frequency of occurrences. For example, in the British National Corpus,[1] the formulaic sequence *up and down* is used 2077 times, while the novel phrase *down and up* is used only 17 times. It is noteworthy that the relatively high-frequency feature is deemed to be the most adopted criterion to detect FL in psycholinguistic studies (e.g., Siyanova-Chanturia & Pellicer-Sánchez, 2019; Wood, 2015).

Akin to single word processing (e.g., Balota & Chumbley, 1984; Bod, Hay, & Jannedy, 2003; Ellis, 2002a, 2002b; Forster, 1976; Jescheniak & Levelt, 1994; Rayner & Duffy, 1986), numerous studies show that the processing of FL is greatly modulated by the frequency of occurrence (e.g., Arnon & Snider, 2010; Bell, Jurafsky, Fosler-Lussier, Girand, Gregory, & Gildea, 2003; Hernández, Costa, & Arnon, 2016; Siyanova-Chanturia, Conklin, & Schmitt, 2011a; Siyanova-Chanturia, Conklin, & van Heuven, 2011b). That is, FL is processed faster than novel language (e.g., Siyanova et al., 2011a, 2011b); and, higher-frequency FL is processed faster than lower-frequency FL (e.g., Arnon & Snider, 2010; Bell et al., 2003).

In FL processing, phrase frequency is operationalized in different ways. Some researchers treat it as a dichotomy (low frequency vs high frequency, e.g., Arnon & Cohen Priva, 2013; Isobe, 2011; Jiang & Nerasova, 2007; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019) or various frequency bins (low frequency vs mid-low frequency vs mid-high frequency vs high frequency, e.g., Kim & Kim, 2012; Sosa & MacFarlane, 2002) in their experimental design and data analysis. These researchers found that highly frequent phrases (i.e., FL) enjoy a privilege in processing over infrequent ones (i.e., novel language).

---

[1] Retrieved from: https://www.english-corpora.org/bnc/.

Some researchers treat phrase frequency as a categorical variable (i.e., dichotomy or various frequency bins) in the experimental design but a continuum in the data analysis (e.g., Bannard & Matthews, 2008; Ellis, Frey, & Jalkanen, 2009; Ellis, Simpson-Vlach, & Maynard, 2008; Hernández et al., 2016; Snider & Arnon, 2012, Experiments 1-2; Sonbul, 2014; Sonbul & El-Dakhs, 2020, Experiment 2; Yi, 2018; Yi, Lu, & Ma, 2017). These researchers found that higher frequency phrases (i.e., FL) were processed faster with more accuracy than lower frequency ones (i.e., novel language); with the increase of frequency, the processing facilitation became more considerable and significant.

Other researchers examined phrase frequency both as a categorical and a continuous variable in the data analysis (e.g., Arnon & Snider, 2010; Jiang, Jiang, & Siyanova-Chanturia, 2020; Northbrook & Conklin, 2019; Shantz, 2017; Siyanova-Chanturia et al., 2011b; Siyanova-Chanturia & Janssen, 2018; Wolter & Gyllstad, 2013). Some studies found the effect of phrase frequency was significant when it was treated as a continuous, not categorical, variable (e.g., Arnon & Snider, 2010; Shantz, 2017; Siyanova-Chanturia & Janssen, 2018). Others found the phrase frequency effect was significant both as a categorical and a continuous variable (e.g., Jiang et al., 2020; Northbrook & Conklin, 2019; Siyanova-Chanturia et al., 2011b; Wolter & Gyllstad, 2013); however, the magnitude of the continuous measure was always found to be greater than the categorical measure (e.g., Jiang et al., 2020; Northbrook & Conklin, 2019; Siyanova-Chanturia et al., 2011b; Wolter & Gyllstad, 2013). This finding suggests that, compared with phrase frequency as a categorical predictor, the continuous phrase frequency might be a more precise and sensitive measure for FL processing (see Jiang et al., 2020 for more details).

The effect of phrase frequency is not consistently strong across all language proficiency levels. In most cases, while phrase frequency effect is present in older L1 learners and higher-proficiency L2 learners, it is marginal or negligible in younger L1 learners (e.g., Jiang et al., 2020) and lower-proficiency L2 learners (e.g., Siyanova-Chanturia et al., 2011b; Siyanova-Chanturia & Janssen, 2018; Shantz, 2017). Besides, the sensitivity to the phrase frequency effect is different across language proficiency levels. For example, in a timed-rating experiment, Siyanova and Schmitt (2008, Study 3) found that while L1 users (more proficient) successfully differentiated high-frequency collocations from medium- and low-frequency collocations, L2 learners (less proficient) could only discriminate between high- and low-frequency phrases. This study suggests that higher-proficiency learners could be more sensitive to phrase frequencies than lower-proficiency learners.

It is also suggested that phrase frequency might not be the only driving force for FL processing. In most cases, phrase frequency often interacts with familiarity (subjective frequency; e.g., Cronk, Lima, & Schweigert, 1993; Isobe, 2011; Tabossi, Fanari, & Wolf, 2009) and other distributional information, such as mutual information (e.g., Ellis et al., 2008; Yi, 2018; Yi et al., 2017) and transitional probability (e.g., Tremblay & Tucker, 2011). However, in some cases, it is suggested that the role of phrase frequency may become less important (e.g., Durrant & Doherty, 2010, Experiment 2), especially when FSs are figurative and metaphorical (e.g., Haeuser, Baum, & Titone, 2020; Hallin & Van Lancker Sidtis, 2017). Besides, it is found that meaningfulness (e.g., Jolsvai, McCauley, & Christiansen, 2013) and age of acquisition (e.g., Arnon, McCauley, & Christiansen, 2017) could also override phrase frequency to dominate FL processing.

2.4.2 Familiarity

While frequency is regarded as the defining feature of FL in psycholinguistic studies (e.g., Siyanova-Chanturia & Pellicer-Sánchez, 2019; Wood, 2015), it may still not be the most essential characteristic. Some have proposed that familiarity may be the key feature of FL (e.g., Siyanova-Chanturia & Pellicer-Sánchez, 2019). This is because those FSs that are infrequent (figurative FSs in particular) are still highly familiar to language users. As noted in Foster (2001), corpora always "fail to show even a single example of many phrases that would be considered a normal part of any native speaker's repertoire" (p. 81).

Low frequency in corpora does not necessarily mean that an expression is seldom used (e.g., Foster, 2001; Moon, 1998) or unfamiliar to language users (e.g., Carrol & Conklin, 2020; Hallin & Van Lancker Sidtis, 2017; Rammell, Van Lancker Sidtis, & Pisoni, 2017). The asynchrony between the frequency and familiarity of (some) FSs may be due to the fact that, unlike single words, FL is not always acquired according to frequency (e.g., Fernández & Schmitt, 2015; Garnier & Schmitt, 2016; Macis & Schmitt, 2017; Siyanova-Chanturia & Pellicer-Sánchez, 2019). For some FSs that are infrequent in corpora (figurative FSs in particular), they could abound in the daily exposure of language learners (especially young children). The constant exposure to these FSs makes them as conventional as other FSs to language learners. Consequently, these relatively infrequent yet familiar FSs could still be learned, processed, and used as well as other highly frequent FSs. For example, in an eye-tracking study, Carrol and Conklin (2020) found that less frequent yet highly familiar idioms were processed similarly to frequent binomials and collocations.

2.4.3 Predictability

Once a FS is known, it should be much easier for language users to process its final constituent(s), compared with the final constituent(s) in a novel expression. For instance, when seeing or hearing the word '*sweep*', we can think of '*the floor*' effortlessly, instead of '*the door*'. It seems that the former constituent(s) of a FS could automatically activate its latter/final parts. Indeed, the final constituent(s) of a FS is often comprehended and produced faster, or even skipped, than the same word(s) in a non-formulaic phrase (e.g., Conklin, Pellicer-Sánchez, & Carrol, 2018; Siyanova-Chanturia et al., 2017; Siyanova-Chanturia & Pellicer-Sánchez, 2019). This is because the constituents (at least for content words) of FL are always strongly associated with each other, and the "high frequency of occurrence and familiarity of a phrasal configuration are likely to render its final constituents close to be redundant" (Siyanova-Chanturia & Pellicer-Sánchez, 2019, p. 4). This suggests that the high frequency and/or high familiarity of FL could be the prerequisites for the high predictability for its final word.

2.4.4 Fixedness and modification

Compared with novel language that is completely compositional and flexible, FL occurs on a continuum of fixedness/cohesion (e.g., Gibbs, Nayak, & Cutting, 1989; Siyanova-Chanturia & Pellicer-Sánchez, 2019) and a relatively narrow spectrum of modifiability (e.g., Wray, 2002).

The degree of fixedness differs greatly among FL in various types (e.g., Schmitt, 2005). Some FSs (e.g., idioms) can be extremely fixed (frozen, fossilized), bound, and stable in form and meaning. The components of such FSs almost always combine exclusively with each other (e.g., Sinclair, 1991), with no or very limited modification(s) allowed. For example, we often say '*hit the nail on the head*', but never say '*hit the **polished** nail on the head*'.

On the contrary, some FSs can be relatively flexible and fluid in form and meaning (e.g., Wray, 2002), which permits more modification(s). For example, we can say '*save energy*' as well as '*save <u>all forms of</u> energy*'. As suggested in processing studies (e.g., Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013; Vilkaitė, 2016a), the modification to its form and meaning, unlike in some figurative FSs, will not entirely jeopardize its formulaicity but enrich its original meaning.

The modifications of FL can be by and large categorized into five types: inflections, negation, transpositions, truncations (omissions), and insertions (e.g., Moon, 1998; Schmitt, 2005).

Inflection is a type of modification of the form rather than meaning and should always be subject to grammatical rules. This kind of modification often happens in FSs that consist of verbs or nouns. The inflection could be a change of the end of a constituent verb (e.g., *make it clear → made it clear*), change of passivization (e.g., *lay down the law → the law is laid down*), and change of pluralization (e.g., *take the wind out of someone's sail → take the wind out of someone's sails*).

Negation is a type of modification of both form and meaning, with a negative word being added or replacing the article in a FS, such that the modified FS will carry an opposite meaning from its original version. For example, we usually use '*spill the beans*' and also '*spill no beans*'.

Transposition is another type of modification of the form with the order of the constituent words reversed. Of note is that the reversing of word order may also change the original meaning of the FS in some cases. For example, we use '*from rags to riches*' and also '*from riches to rags*'.

Truncation is also a type of modification of the form and often appears in classical proverbs and sayings. The truncation in form can never affect the completeness and preciseness of the meaning of the FS. As long as the FS is familiar and conventional to the language user, it is often used with some of the constituents omitted. For example, we usually use '*a silver lining*' to substitute for '*every cloud has a silver lining*'. Of note is that only constituent(s) that do not carry the key meaning(s) can be omitted.

On the contrary, insertion is a type of modification that amplifies both the form and meaning(s) of a FS. For some FSs, we can add one or more words to their original form, although the insertion position (before, former, middle, latter, after) may vary among different FSs (e.g., Krishnamurthy, Sinclair, Jones, & Daley, 2004). The added word(s) can be adjectives (e.g., *lend a hand → lend a* ***helping*** *hand*), quantifiers (e.g., *provide information → provide* ***some*** *information*), adverbials (e.g., *concentrate the mind → concentrate the mind* ***wonderfully***), and so forth. It is suggested that the inserted version of FL is even more common in some languages, such as Chinese (in particular). According to Siewierska, Xu, and Xiao (2010), over 50% of Chinese splittable compounds (e.g., *帮忙* '*offer help*') are often written and spoken in their inserted form (e.g., *帮了一个大忙* '*offered* *great help*'), instead of the original version. It is also evidenced that inserted FL can be acquired as well as its original version (e.g., Vilkaitė, 2017).

2.4.5 Phonology

FL is markedly different from novel language in phonology in several ways. First, according to online production studies (e.g., Hallin & Van Lancker Sidtis, 2017; Siyanova-Chanturia & Janssen, 2018; Siyanova-Chanturia & Lin, 2018), FL is often articulated with a faster speech rate than novel language. Second, FL is more resistant to internal pauses or hesitations than novel language (e.g., Wray, 2004). For example, '*Every cloud has a silver lining*' in the speech of a proficient (and healthy) language learner cannot be '*Every cloud … has a s-silver um … lining*'. Third, compared with novel language, FL requires less stress in the tonal pattern (e.g., Hallin & Van Lancker Sidtis, 2017). Finally and crucially, compared with novel language, FL is often less precisely articulated (e.g., Bell et al., 2003; Bybee, 2000, 2002; Gregory, Raymond, Bell, Fosler-Lussier, & Jurafsky, 1999; Jurafsky, Bell, Gregory, & Raymond, 2001). The constituent(s) of a FS is often phonologically reduced (e.g., Bybee, 2002; Bybee & Scheibman, 1999; Fowler, 1988). For example, a proficient (and healthy) language speaker would say '*kinda*', '*wanna*', '*gonna*', instead of '*kind of*', '*want to*', '*going to*'. This is because FL is more frequent, familiar, and predictable relative to novel language, which could activate the exact meaning of the FS without full articulation. This suggests that the high frequency, familiarity, and predictability of FL could be the prerequisites for its phonological reduction.

To recap the characteristics of FL briefly, most FL is frequent, highly familiar, and predictable for its final constituent(s). The structure of FL is relatively fixed, but this does not mean it cannot be modified. Compared with novel language, the articulation of FL requires fewer, or no, pauses and stress, and the constituent(s) of FL is more often phonologically reduced. These characteristics of FL suggest it is a unique phenomenon in linguistics that can never be neglected.

2.5 Role of formulaic language in language processing and language learning

In addition to the characteristics, researchers attach great importance to the role of FL in language processing (comprehension and production) and language learning. In the aspect of processing, FL can save cognitive efforts. This is because "it gives us access to 'ready-made frameworks' on which to hang the expression of our ideas thus we do not have to go through the labour of generating an utterance every time we want to say something" (Becker, 1975, p. 17). With these ready-made frameworks stored in our brain, we can spend less time searching for individual elements one by one to build up the expression every time we intend to produce. With these ready-made frameworks stored in our brain, we can spend less time focusing on comprehending the meaning of individual constituents of a recurrent expression. This is because the (relatively) high frequency, familiarity, predictability, and formulaicity of the framework can help us with easy access to the meaning(s) of the whole expression, with only certain, but not all, its constituents activated. With these ready-made frameworks stored in our brain, we do not need to spend time considering whether the combination of individual elements of a recurrent expression is grammatically and semantically correct every time we encounter or use it. Thus, FL is cognitively economic.

In L1 acquisition, FL is regarded as an "acquisitional aid" (Wray, 2002, p.119). In early development, it provides infants with starter utterances in specific contexts and enables them to have social interactions with others (e.g., Locke, 1997). In later development, FL "allows children to say more and more completely what they mean than they would if they had to construct an utterance from scratch" (Nelson, 1981, p.181-182), despite the lack of grammar or vocabulary knowledge. It has been argued that FL acts as a template for children to acquire new L1 forms (e.g., Arnon, 2011; Arnon & Clark, 2011; Lieven, Pine, & Barnes, 1992; Tomasello, 2003; Wray, 2002).

FL is also widely acknowledged in L2 acquisition. Akin to it in L1 acquisition, FL provides "a kind of frame expressing stance, discourse organization, or referential status, associated with a slot for the expression of new information relative to that frame" (Biber, Conrad, & Cortes, 2004, p. 400). Besides, it is suggested that FL contributes to the fluency with which language is used, rendering it more native-like (e.g., Boers, Eyckmans, Kappel, Stengers, & Demecheleer, 2006; Lewis, 2000; Nattinger & DeCarrico, 1992; Schmitt, 2004; Wray, 2002, 2008). Some researchers (e.g., Ellis, 2001; Tomasello, 2003; Wong-Fillmore, 1979) even propose that FL learning is the key to the general process of L2 learning.

Therefore, FL is indispensable in L1 and L2 learning (e.g., Arnon & Christiansen, 2017; Christiansen & Arnon, 2017). Without FL, language processing would be laborious; language learning would be less efficient; and language use would be non-native-like.

2.6 Summary

To sum up, FL is a universal linguistic phenomenon, characterized by relatively high frequency, high familiarity, relatively high predictability, relative fixedness, and phonological reduction.

This phenomenon plays a vital role in language processing and language learning. It provides templates for comprehension and production, such that the processing could be effortless. It is an essential aid in first and second language acquisition, and helps learners use the target language in a fluent and native-like way.

In the next chapter, I will review how FL is processed by different populations of language users and learners (L1 and L2 adults as well as L1 children) in detail. I will also introduce how modified FL and Chinese FL are processed in detail.

Chapter 3: The Processing of Formulaic Language

3.1 Introduction

My previous chapter helped forge a general understanding of what formulaic language (FL) is. In this chapter, I will take a closer look at how this linguistic unit is processed by healthy individuals, young adults and children. Note that the 'processing' in the current thesis refers to online processing. Unlike offline methods (such as untimed rating, pen and paper tests), online processing refers to real-time comprehension and production under time pressure, which directly reflects the cognitive process during reading, listening, and speaking (repetition or articulation), and requires no secondary tasks.

I will first look at FL processing in first language (L1) users. Importantly, I will discuss whether and how L1 FL is processed differently from novel language (relatively infrequent expressions that need to be analysed and constructed anew). Next, I will look at FL processing in second language (L2) learners. Again, I will discuss whether and how L2 FL is processed differently from novel language. After that, I will introduce the processing of FL in children and make comparisons between FL processing and novel language processing in this population.

Following the review of FL processing in these three populations, I will look at modified FL, specifically, to see whether or not it is processed differently from modified novel language.

In addition, I will cover the research into FL processing in Chinese, and discuss whether and how it is processed differently from novel language.

At the end of the chapter, I will draw on pertinent theories, such as usage-based models, exemplar-based models, connectionist models, and statistical learning theory, that are able to account for the mechanism for FL processing.

3.2 The processing of formulaic language in L1 users

In this section, I will introduce FL processing in L1 adults, in most cases university students. The introduction to relevant studies will be broadly divided into two parts: the processing of figurative (non-compositional) FL and the processing of literal (compositional) FL. Admittedly, it is never easy to draw a clear-cut distinction between these two broad categories. This is because FL tends to manifest itself on a continuum of frequency, familiarity, predictability, fixedness, cohesion, decomposability, nuanced semantics, non-literal meaning, and so forth (e.g., Siyanova-Chanturia & Van Lancker Sidtis, 2019; Wray, 2002). Nonetheless, in order to have a general understanding of what has been found in the processing of various formulaic sequences (FSs), a distinction for FL is needed in any case. I choose to broadly divide FL into two categories - figurative FL versus literal FL - for the current literature review. This is because, in most cases, figurative FL is distinctively different from literal FL in many ways, such as nuanced semantics, non-literal meanings, fixedness, and compositionality, which could result in a (slight) difference in semantic processing between these two categories. Please note that the division of FL is never the aim of the current thesis, and the division of FL in the current literature review is made for convenience (such that this section could be marked with signposts and become more reader-friendly). Further, I will acknowledge and follow the terms (such as binomials, collocations, idioms, phrasal verbs, n-grams) used in the original studies when reviewing them; however, the definitions or conditions of the same term used by individual studies might be (slightly) different.

The literature review of L1 FL processing will start with the processing of figurative FL followed by literal FL processing. This is because the earliest investigations (of the 1970s) were almost entirely on figurative FL; investigations into literal FL processing did not start until the 2000s. Although figurative FL is not used in the current thesis on Chinese FL processing, as existing research into Chinese FL processing has largely pertained to figurative FSs (e.g., Liu, Li, Shu, Zhang, & Chen, 2010; Shen, Wang, & Liu, 2017; Yu et al., 2016; Zhang, Yang, Gu, & Ji, 2013; Zhou, Zhou, & Chen, 2004), it is still relevant to the topic of the current thesis.

3.2.1 The processing of figurative formulaic language in L1

Investigations into L1 figurative FL processing can be traced back to the 1970s (for a review, see Siyanova-Chanturia & Van Lancker Sidtis, 2019) and have been developed well to this day. By and large, two topics have been focused on: (1) the processing of figurative FSs, such as idioms (non-compositional sequences with a figurative meaning, e.g., *a piece of cake*), proverbs (concrete and traditional sayings stating a general truth or piece of advice, e.g., *a cat has nine lives*), phrasal verbs (Verb+Particle combinations with more than one meaning, e.g., *chew out*), etc. versus novel expressions; and (2) the processing of figurative versus literal meanings in ambiguous idioms (idioms that allow both figurative and literal interpretations).

Table 1:

Summary of Studies on the Processing of Figurative Formulaic Language versus Novel Language in L1 Adults

Study	Journal/Book/Conference	Language of FL	Type of FL	Research method
Swinney and Culter (1979)	*Journal of Verbal Learning and Verbal Behavior*	English	Idioms	Phrasal decision
Van Lancker, Canter, and Terbeek (1981)	*Journal of Speech and Hearing Research*	English	Idioms	Phrasal production
Gibbs and Gonzales (1985)	*Cognition*	English	Idioms	Phrasal decision
Laurent, Denhières, Passerieux, Iakimova, and Hardy-Baylé (2006)	*Brain Research*	French	Idioms	EEG
Caillies and Butcher (2007)	*Metaphor and Symbol*	French	Idioms	Lexical decision
Tabossi, Fanari, and Wolf (2008)	*Journal of Experimental Psychology: Learning, Memory, and Cognition*	Italian	Idioms	Phrasal decision
Tabossi, Fanari, and Wolf (2009)	*Memory and Cognition*	Italian	Idioms & clichés	Phrasal decision
Vespignani, Canal, Molinaro, Fonda, and Cacciari (2009)	*Journal of Cognitive Neuroscience*	Italian	Idioms	EEG
Molinaro and Carreiras (2010)	*Biological Psychology*	Spanish	Idioms	EEG
Jolsvai, McCauley, and Christiansen (2013)	*Proceedings of the 35th Annual Conference of the Cognitive Science Society.*	English	Idioms	Phrasal decision
Rommers, Dijkstra, and Bastiaansen (2013)	*Journal of Cognitive Neuroscience*	Dutch	Idioms	Lexical decision & EEG
Siyanova-Chanturia and Lin (2018)	*International Journal of Applied Linguistics*	English	Idioms	Reading-aloud

Notes. All studies in Table 1 examined figurative FL processing in L1 adults only; those studies observed both L1 and L2 adults or both L1 adults and children were not included in the table. This table is not complete and includes representative studies. All studies listed in the table observed a processing advantage for figurative FL over novel language in L1 adults.

The first topic has been explored for a long time (see Table 1 for a summary of relevant studies). Among the early investigations (pre-2000), a number of them have sought to compare the processing of idioms versus novel phrases (e.g., Gibbs & Gonzales, 1985, Experiment 2; Swinney & Cutler, 1979; Van Lancker, Canter, & Terbeek, 1981). For example, in a phrasal decision study, Swinney and Cutler (1979) asked L1 users to decide whether or not each of the presented word strings was meaningful in English. Two conditions of word strings were observed: (1) idiomatic word strings (e.g., *break the ice*) and (2) matched controls with the final word of relative idioms replaced (e.g., *break the cup*). Swinney and Cutler (1979) found that idiomatic word strings were processed significantly faster than matched controls. This result suggests a processing advantage for figurative FSs over novel phrases. Crucially, this processing advantage was observed from both the macro (whole phrase) and micro views (the first- and last-constituent words) (Swinney & Cutler, 1979).

Later investigations (post-2000) went further with the comparisons between the processing of figurative FL and novel language. Unlike the early investigations (e.g., Gibbs & Gonzales, 1985; Swinney & Cutler, 1979), later investigations started to look at figurative FL processing in sentence context rather than in isolation (e.g., Caillies & Butcher, 2007; Vespignani, Canal, Molinaro, Fonda, & Cacciari, 2009), and compared the processing of decomposable versus non-decomposable idiomatic expressions (e.g., *uscire dal seminato* '*wander off the point*' vs *Tenere a stecchetto* '*keep someone short of money or food*') relative to their novel controls (e.g., Tabossi, Fanari, & Wolf, 2008, Experiment 3; Tabossi, Fanari, & Wolf, 2009). Despite the variety of research methods being used, these investigations confirmed the processing advantage for figurative FSs over novel phrases in sentence context and in both decomposable and non-decomposable figurative FSs. The processing advantage for figurative FL over novel language was also confirmed in recent investigations (post-2010; e.g., Hallin & Van Lancker Sidtis, 2017; Jolsvai, McCauley, & Christiansen, 2013; Rommers, Dijkstra, & Bastiaansen, 2013; Siyanova-Chanturia & Lin, 2018).

The advancement of research techniques has enabled researchers to tap into figurative FL processing from the perspective of electrophysiology (e.g., Laurent, Denhières, Passerieux, Iakimova, & Hardy-Baylé, 2006; Molinaro & Carreiras, 2010; Rommers et al., 2013, Experiment 2; Vespignani et al., 2009). For example, in an electroencephalogram (EEG) study, Vespignani et al. (2009) found that the constituent word after the recognition point of idioms induced less negative electrophysiological potentials (e.g., N400, which is associated with semantic violation) than in matched novel phrases. This finding suggests that figurative FL enjoys higher predictability for its constituents than novel language, and the higher predictability further facilitates the processing of figurative FL. Furthermore, it has been evidenced that the processing facilitation took effect not only within the area of a certain

FS but also extended to the end of a sentence embedded with the FS (e.g., Vespignani et al., 2009).

In addition to the processing of FL versus novel language, researchers sought to explore the processing of figurative versus literal interpretations of ambiguous idioms (e.g., *kick the bucket*). This topic attracted the attention of researchers because of the availability of the two distinct meanings in some idioms. For example, *kick the bucket* can either mean ‘to kick a pail’ (literal) or ‘to die’ (figurative). However, results of the existing studies are mixed. Some researchers found that these two meanings were processed in parallel in L1 users (e.g., Estill & Kemper, 1982; Glass, 1983; Kemper, 1986). On the contrary, other researchers argue that these two meanings are processed distinctively in a serial process (e.g., Cacciari & Tabossi, 1988; Gibbs, 1980; Gibbes, Bogdanovich, Sykes, & Barr, 1997; Ortony, Schallert, Reynolds, & Antos, 1978; Siyanova-Chanturia & Lin, 2018; Yang, Ahn, & Van Lancker Sidtis, 2015). Some researchers claim that the figurative meaning should be processed with priority over the literal one (e.g., Gibbs, 1980, Experiment 1; Ortony et al., 1978, Experiment 2; Schweigert & Moates, 1988; Siyanova-Chanturia & Lin, 2018; Yang et al., 2015). Conversely, some researchers hold that the figurative meaning does not always enjoy a priority over the literal one (e.g., Cacciari & Tabossi, 1988; Kemper, 1986), and the prioritised meaning could alter among idioms with different predictability (e.g., Cacciari & Tabossi, 1988) or with different familiarity (e.g., Kemper, 1986).

Furthermore, existing studies explored some factors that could affect figurative FL processing. These factors include decomposability (also known as analysability or frozenness) (e.g., Caillies & Declercq, 2010; Cutting & Bock, 1997; Gibbs, Nayak, & Cutting, 1989), context (e.g., Cieślicka, Heredia, & Olivares, 2014; Colombo, 1993; Fanari, Cacciari, & Tabosssi, 2010), familiarity (e.g., Cronk, Lima, & Schweigert, 1993; Libben & Titone, 2008), predictability (e.g., Tabossi, Fanari, & Wolf, 2005; Titone & Connine, 1994), and transparency (e.g., Cacciari & Glucksberg, 1995; Carrol & Littlemore, 2020).

To recap briefly, investigations into L1 figurative FL processing have by and large focused on the comparisons between figurative FS and novel phrase processing, and the comparisons between the figurative and literal meanings in ambiguous idiom processing. Although whether the figurative meaning enjoys faster processing over literal meaning has not been affirmed yet, a number of studies have accumulated to suggest a significant processing advantage for figurative FL over novel language, with faster processing for figurative FSs relative to novel phrases.

3.2.2 The processing of literal formulaic language in L1

Compared with figurative FL processing, investigations into literal FL processing started relatively late. It was not until the 2000s that researchers started to look at the processing of literal FSs, such as binomials (sequences of X and Y where a specific word order is highly preferred, e.g., *bride and groom*), collocations (co-occurring word pairs, e.g., *make a wish*), n-grams (co-occurring word strings retrieved from corpora, e.g., *in the middle of*). Studies on literal FL processing have by and large focused on one topic, that is, the processing of literal FSs versus novel phrases, in isolation or within sentence context (see Table 2 for a summary of relevant studies).

Table 2:

Summary of Studies on the Processing of Literal Formulaic Language versus Novel Language in L1 Adults

Study	Journal/Book/Conference	Language of FL	Type of FL	Research method
Sosa and MacFarlane (2002)	*Brain and Language*	English	Collocations	Word monitor
Molinaro, Vespignani, Canal, Fonda, and Cacciai (2008)	*Psychophysiology*	Italian	N-grams	EEG
Kapatsinski and Radicke (2009)	*In Corrigan et al. (Eds.), Formulaic Language. John Benjamins*	English	Collocations	Word monitor
Ellis, Frey, and Jalkanen (2009)	*In Römer and Schulze (Eds.), Exploring the Lexis-Grammar Interface. John Benjamins*	English	Collocations	Lexical decision
Tremblay, Derwing, and Libben (2009)	*Proceeding of 23rd Northwest Linguistics Conference*	English	N-grams	Self-paced reading
Arnon and Snider (2010)	*Journal of Memory and Language*	English	N-grams	Phrasal judgement
Durrant and Doherty (2010)	*Corpus Linguistics and Linguistic Theory*	English	Collocations	Phrasal judgement
Columbus (2010)	*In Wood (Ed.), Perspectives on Formulaic Language: Acquisition and Communication. Continuum*	English	Collocations N-grams Idioms	Eye-tracking

Study	Journal/Book/Conference	Language of FL	Type of FL	Research method
Tremblay and Baayen (2010)	*In Wood (Ed.), Perspectives on Formulaic Language: Acquisition and Communication. Continuum*	English	N-grams	EEG
Tremblay and Tucker (2011)	*The Mental Lexicon*	English	N-grams	Phrasal production
Millar (2011)	*Applied Linguistics*	English	Collocations	Self-paced reading
Tremblay, Derwing, Libben, and Westbury (2011)	*Language Learning*	English	N-grams	Self-paced reading
Janssen and Barber (2012)	*PLOS ONE*	Spanish	Collocations	Phrasal production
Arnon and Cohen Priva (2013)	*Language and Speech*	English	N-grams	Phrasal production
Siyanova-Chanturia, Conklin, Caffarra, Kaan, and van Heuven (2017)	*Brain and Language*	English	Binomials	EEG
Carrol and Conklin (2020)	*Language and Speech*	English	Binomials Collocations Idioms	Eye-tracking

Notes. All studies in Table 2 examined literal FL processing in L1 adults only; those studies observed both L1 and L2 adults or both L1 adults and children were not included in the table. This table is not complete and includes representative studies. All studies listed in the table, except for Sosa and MacFarlane (2002) and Kapatsinski and Radicke (2009) who explored the fixedness of FL versus novel language, observed a processing advantage for literal FL over novel language in L1 adults.

Early studies sought to explore the relationship between the parts and the whole of FL versus novel language during processing (e.g., Kapatsinski & Radicke, 2009; Sosa & MacFarlane, 2002). For example, in a word monitoring study, Sosa and MacFarlane (2002) asked L1 users to listen and monitor the English function word '*of*' in phrases that varied in frequency (e.g., low frequency: *each of*; mid-low frequency: *much of*; mid-high frequency: *most of*; high frequency: *some of*). Sosa and MacFarlane (2002) found that it was more difficult (in terms of speed and accuracy) for L1 users to detect the target particle when it was in a higher frequency phrase than in a lower frequency phrase. However, this finding is questioned by Kapatsinski and Radicke (2009), in that the particle '*of*' is considered to be always phonologically reduced in phrases. With this in mind, Kapatsinski and Radicke (2009) used Verb+*up* combinations, which are less phonologically reduced. These experimenters had

L1 users monitor the particle '*up*' when listening to sentences. Although distinct sets of materials were used, Kapatsinski and Radicke (2009) found a similar result as in Sosa and MacFarlane (2002). These findings suggest that frequent literal FSs are more fixed and fused than infrequent novel phrases; and, frequency above the word level (i.e., phrase frequency) plays a role in phrasal processing.

The supposition that frequency could play a role in phrasal processing is evidenced in later studies on the processing of literal FL versus novel language (e.g., Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Carrol & Conklin, 2020; Columbus, 2010; Durrant & Doherty, 2010; Ellis, Frey, & Jalkanen, 2009; Millar, 2011; Tremblay, Derwing, & Libben, 2009; Tremblay, Derwing, Libben, & Westbury, 2011; Tremblay & Tucker, 2011). For example, Ellis et al. (2009) asked L1 users to decide whether a word appearing in the given phrases was a real word or not. Ellis and collaborators (2009) found that the words in high-frequency collocations (e.g., *start again*) were decided significantly faster than in low-frequency phrases (e.g., *end again*). This finding suggests that literal FL enjoys a processing advantage over novel language, and high frequency facilitates the processing of phrases. Ellis et al. (2009) further found that the magnitude of the facilitation varied among collocations with different phrase frequencies, with the facilitation greater in higher frequency collocations than in lower frequency collocations. This finding suggests that the magnitude of processing facilitation of literal FL is dependent on its phrase frequency. Similar results were found in Arnon and Snider (2010, see also Snider & Arnon, 2012, Experiments 1 & 2), Durrant and Doherty (2010), Tremblay and Tucker (2011), Janssen and Barber (2012, Experiment 1), and Arnon and Cohen Priva (2013, Study 1), which revealed that FL advantage was modulated by the frequency of the whole phrase, instead of the frequency of its constituent.

The processing advantage for literal FL has been found not only in isolation (e.g., Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Durrant & Doherty, 2010; Ellis et al., 2009; Janssen & Barber, 2012; Tremblay & Tucker, 2011) but also in sentential context (e.g., Carrol & Conklin, 2020; Columbus, 2010; Millar, 2011; Tremblay et al., 2009, 2011). For example, in a self-paced reading study, Tremblay et al. (2009) compared the reading times of two sentence types: (1) FS sentence (a frequent FS embedded in a sentence; e.g., *His friend's got nothing to do next Friday*); (2) non-FS sentence (an infrequent novel phrase embedded in a sentence; e.g., *His friend's got one to do next Friday*). All sentences were presented in three distinct patterns – (a) word-by-word, (b) portion-by-portion, and (c) sentence-by-sentence. Tremblay and colleagues (2009) found that FSs were read faster than novel phrases, in the (b) and (c) but not in the (a) presentation patterns. This finding implies that the pattern of material presentation may affect the processing advantage of FL. As FL is relatively fixed, the word-by-word presentation might jeopardize its inner structure (formulaicity), and thus pose a threat to its processing advantage. In their follow-up study, however, Tremblay and colleagues (2011) found the processing advantage for FS sentences in all three presentation patterns. This finding should be taken with caution, given the relatively small sample size of materials (only nine pairs of target items went to final data analysis after norming). In another self-paced reading study, Millar (2011) asked L1 users to read sentences embedded with Adjective+Noun phrases. Two phrase types were observed: (1) the native-speaker collocation condition (i.e., real collocations frequently registered in the British National Corpus, e.g., *ideal partner*); (2) the learner collocation condition (i.e., pseudo collocations infrequently registered in the corpus; these phrases had their front constituent mismatched or misformed, which made the combination less accepted by L1 users, e.g., *best partner*). Millar (2011) embedded these two phrase types into identical sentence contexts and compared the final word (e.g., *partner*) in each pair of phrases. Millar (2011) found that the final word in the learner collocation condition was processed significantly slower than in the native-speaker collocation condition. This result suggests that, unlike the facilitation in FL processing, the processing of novel language, which is infrequent and non-nativelike, causes a great cognitive burden to language processers.

The processing advantage for literal FL over novel language was also observed in EEG research (e.g., Molinaro, Vespignani, Canal, Fonda, & Cacciai, 2008; Siyanova-Chanturia, Conklin, Caffarra, Kaan, & van Heuven, 2017; Tremblay & Baayen, 2010). For example, in Molinaro et al. (2008) and Tremblay and Baayen (2010), high-frequency sequences (e.g., *in the middle of*) elicited small electrophysiological deflections, whereas low-frequency sequences (e.g., *in the middle are*) elicited large electrophysiological deflections in L1 users. This finding suggests that frequent FL is predictable and expected relative to infrequent novel language. In another EEG study, Siyanova-Chanturia et al. (2017, Experiment 1) asked L1 users to read high-frequency binomials (e.g., *knife and fork*), strongly associated but infrequent controls (e.g., *spoon and fork*), and semantically violated fillers (e.g., *theme and fork*). Results of the electrophysiological components suggest an easier template (category) matching (P300) and greater semantic integration (N400) for binomials relative to infrequent controls. In their second study (Siyanova-Chanturia et al., 2017, Experiment 2), the conjunction word ‘*and*’ was removed from all phrases used in the first experiment (e.g., *knife fork* vs *spoon fork* vs *theme fork*). Siyanova-Chanturia et al.

(2017, Experiment 2) found that the processing facilitation for binomials in the first experiment disappeared. This finding suggests that structural integrity is essential for FL facilitation.

Besides, existing studies explored some additional factors that could affect literal FL processing. These factors include probability of occurrence (e.g., Tremblay & Tucker, 2011), mutual information (e.g., Carrol & Conklin, 2020; Columbus, 2010), age of acquisition (e.g., Arnon, McCauley, & Christiansen, 2017).

To recap briefly, investigations into L1 literal FL processing have accumulated to reveal a significant processing advantage for FL over novel language, with faster reading speed, shorter articulation durations, and easier semantic integration for FSs relative to novel phrases. Although several factors (such as probability, mutual information, age of acquisition) can also affect phrasal processing, phrase frequency is suggested to be the dominant factor for eliciting FL facilitation.

All in all, as suggested in existing studies, FL, figurative or literal, enjoys a significant processing advantage over novel language in L1 users. The FL advantage is, to a large extent, driven by phrase frequency, especially for literal FSs.

3.3 The processing of formulaic language in L2 learners

Compared with L1 users, L2 learners are less experienced language users. This is principally because L2 learners have limited experience with (exposure to) the target language, which can result in slower processing, more difficult learning, and less accurate use relative to L1 users (e.g., Doughty & Long, 2003; VanPattern & Williams, 2015). In fact, as VanPattern and Wiliams (2015) put it, L2 learners "will never reach a level of proficiency or competence comparable to a native speaker's" (p. 5). Furthermore, for some L2 learners, their L1 is different from the target language in some ways, such as word order. The incongruency between their L1 and L2 can pose a threat to L2 learning (e.g., Carrol, Conklin, & Gyllstad, 2016; Peters, 2016; Sonbul & El-Dakhs, 2020). The limited exposure and incongruency between their L1 and the target language could make FL processing in L2 different from it in L1. Therefore, FL processing in L2 learners warrants investigations.

Unlike investigations into L1 FL processing, which dated back to the 1970s, research into L2 FL processing started relatively late. It was not until the early 2000s that researchers started to look at the processing of FSs in L2 learners (e.g., Matlock & Heredia, 2002; Underwood, Schmitt, & Galpin, 2004; Van Lancker-Sidtis, 2003). Most researchers sought to look at L2 FL processing in learners at intermediate and/or proficient level (university students in most cases), and compare the performance of L2 learners with L1 users (e.g., Carrol & Conklin, 2014a, 2017; Ellis, Simpson-Vlach, & Maynard, 2008; Jiang & Nekrasova, 2007; Matlock & Heredia, 2002; Siyanova-Chanturia, Conklin, & Schmitt, 2011a; Siyanova-Chanturia, Conklin, & van Heuven, 2011b; Sonbul, 2014; Underwood, Schmitt, & Galpin, 2004). Some researchers have found a comparable processing pattern for FL in L1 users and L2 learners, while others have not. In this section, I will talk about FL processing in L2 and its similarities and differences with L1 FL processing in detail.

I will begin with the review of investigations into L2 figurative FL processing, followed by L2 literal FL processing. This is because: (1) the investigations into L2 figurative FL processing (e.g., Matlock & Heredia, 2002; Underwood et al., 2004) pre-date those on L2 literal FL processing (e.g., Ellis et al., 2008; Jiang & Nekrasova, 2007); and (2) to keep the structure of literature review consistent with it of L1 FL processing.

3.3.1 The processing of figurative formulaic language in L2

Akin to figurative FL processing in L1, investigations into the processing of L2 figurative FSs have by and large focused on two topics: (1) the processing of FL versus novel language; (2) the processing of the figurative versus literal meanings of ambiguous idioms.

Table 3:

Summary of Studies on the Processing of Figurative Formulaic Language versus Novel Language in (L1 users and) L2 Learners

Study	Journal/Book/Conference	Language proficiency	Measurement of proficiency	L1 background	Language of FL	Type of FL	Research method	Main finding
Matlock and Heredia (2002)	*Advances in Psychology*	L1 users & Higher- & Lower-proficiency L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	Phrasal verbs	Self-paced reading	Processing advantage for FL over novel language in proficient L2 learners
Underwood, Schmitt, and Galpin (2004)	*In Schmitt (Ed.), Formulaic Sequences: Acquisition, Processing, and Use. John Benjamins*	L1 users & High-proficiency L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	Idioms Proverbs N-grams	Eye-tracking	Processing advantage for FL over novel language in proficient L2 learners
Cieślicka (2006)	*Second Language Research*	High-proficiency L2 learners	Length of L2 learning	Polish	English	Idioms	Lexical decision	No processing advantage for L2 FL

Study	Journal/Book/Conference	Language proficiency	Measurement of proficiency	L1 background	Language of FL	Type of FL	Research method	Main finding
Conklin and Schmitt (2008)	*Applied Linguistics*	L1 users & High-proficiency L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	Idioms	Self-paced reading	Processing advantage for FL over novel language in proficient L2 learners
Siyanova-Chanturia, Conklin, and Schmitt (2011a)	*Second Language Research*	L1 users & High-proficiency L2 learners	Length of L2 learning & Self-rating	Mixed L1 backgrounds	English	Idioms	Eye-tracking	No processing advantage for L2 FL
Yeganehjoo and Thai (2012)	*The Southeast Asian Journal of English Language Studies*	High-proficiency L2 learners	Length of L2 learning	Iranian	English	Idioms	Elicited production	Processing advantage for FL over novel language in proficient L2 learners
Carrol and Conklin (2014a)	*Bilingualism: Language and Cognition*	L1 users & High-proficiency L2 learners	Length of L2 learning	Chinese	English	Idioms	Lexical decision	No processing advantage for L2 FL
Titone, Columbus, Whitford, Mecier, and Libben (2015)	*In Heredia and Cieślicka (Eds.), Bilingual Figurative Language Processing. Cambridge University Press.*	High-proficiency L2 learners	Length of L2 learning	French	English	Idioms	Semantic judgement	Processing advantage for FL over novel language in proficient L2 speakers
Beck and Weber (2016a)	*Proceeding of the 38th Annual Conference of the Cognitive Science Society*	High-proficiency L2 learners	Length of L2 learning	German	English	Idioms	Lexical decision	Processing advantage for FL in proficient L2 learners
Carrol, Conklin, and Gyllstad (2016)	*Studies in Second Language Acquisition*	L1 users & High-proficiency L2 learners	Length of L2 learning & Self-rating & Vocabulary test	Swedish	English	Idioms	Eye-tracking	Processing advantage for FL over novel language in proficient L2 learners
Carrol and Conklin (2017)	*Bilingualism: Language and Cognition*	L1 users & High-proficiency L2 learners	Length of L2 learning	Chinese	English	Idioms	Eye-tracking	No processing advantage for L2 FL

Note. This table is not complete and includes representative studies.

The first topic has been extensively investigated (see Table 3 for a summary of relevant studies). Some researchers found the processing advantage for figurative FL over novel language in (both L1) and proficient L2 speakers during reading (e.g., Beck & Weber, 2016a; Carrol, Conklin, & Gyllstad, 2016; Conklin & Schmitt, 2008; Titone, Columbus, Whitford, Mecier, & Libben, 2015; Yeganehjoo & Thai, 2012; Underwood et al., 2004).

Some researchers found the processing advantage for FL in L1 and L2 speakers with higher, but not lower, proficiency (e.g., Matlock & Heredia, 2002). For example, in a self-paced reading study, Matlock and Heredia (2002, Experiment 2) had participants read sentences embedded with frequent figurative phrasal-verbs and infrequent phrases for comprehension. The participants consisted of three populations: L1 English speakers, higher-proficiency L2 learners (who began to learn English before the age of 12), and lower-proficiency L2 learners (who began to learn English after the age of 12). Matlock and Heredia (2002, Experiment 2) found that L1 and higher-proficiency L2 speakers read sentences embedded with phrasal-verbs faster than sentences embedded with control phrases; on the contrary, lower-proficiency L2 learners read these two types of sentences at the same rate. This finding suggests that the emergence of FL facilitation in L2 requires considerable experience with the target language and L2 FL processing is modulated by language proficiency. Higher-proficiency learners are experienced relative to lower-proficiency learners, and the great experience enables proficient learners to be more sensitive to FL and phrase frequencies, which facilitates the processing of FL.

However, some researchers did not find the processing advantage for FL in L2, even though their participants were proficient learners (e.g., Carrol & Conklin, 2014a, 2017; Cieślicka, 2006; Siyanova-Chanturia, Conklin, & Schmitt, 2011a). For example, in an eye-tracking study, Siyanova-Chanturia et al. (2011a) asked L1 and highly proficient L2 speakers to read ambiguous idioms (with either figurative meaning or literal meaning) and respective control phrases embedded in (biased) sentences. Siyanova-Chanturia et al. (2011a) found that only L1, but not L2 speakers, exhibited a significant processing advantage for idioms over control phrases. Similar results were also found in Carrol and Conklin (2014a), Carrol and Conklin (2017, Experiment 1), and Cieślicka (2006).

In addition to the processing of L2 figurative FL versus novel language, existing studies compared the processing of figurative versus literal meanings of ambiguous idioms (e.g., *kick the bucket*: ‘to die’ vs ‘to kick a pail’) in proficient L2 learners (as well as L1 users as a reference group). A number of researchers support the literal facilitation in L2 ambiguous idiom processing (e.g., Cieślicka, 2006; Beck & Weber, 2016a, 2016b; Carrol & Conklin, 2017, Experiment 2; Cieślicka, 2013; Cieślicka & Heredia, 2011; Cieślicka, Heredia, & Olivares, 2014; López, Vaid, Tosun, & Rao, 2017; Siyanova-Chanturia et al., 2011a), while others advocate the figurative facilitation (e.g., Paulmann, Ghareeb-Ali, &

Felser, 2015). On the contrary, some researchers hold that these two meanings are processed in parallel (e.g., Conklin & Schmitt, 2008; Van Lancker-Sidtis, 2003).

To recap briefly, investigations into L2 figurative FL processing have by and large focused on the comparisons between FL versus novel language processing, and the comparisons between the figurative versus literal meanings in ambiguous idiom processing. Akin to the studies on L1 figurative FL processing, existing studies have accumulated to suggest a significant processing advantage for figurative FL over novel language in L2 learners, those with high proficiency in most cases.

Results of the comparisons between the figurative and literal meanings in L2 ambiguous idiom processing are mixed. Despite this, a number of studies have accumulated to suggest that the literal, rather than figurative, meaning is more often to be processed with priority.

3.3.2 The processing of literal formulaic language in L2

The investigations into L2 literal FL processing can be dated back to the late 2000s (e.g., Ellis, Simpson-Vlach, & Maynard, 2008; Jiang & Nekrasova, 2007). Akin to literal FL processing in L1, investigations into L2 literal FL processing have principally focused on the comparisons between FL versus novel language in (L1 and) L2 speakers, in isolation or within sentence context (see Table 4 for a summary of relevant studies).

Table 4:

Summary of Studies on the Processing of Literal Formulaic Language versus Novel Language in (L1 users and) L2 Learners

Study	Journal/Book/Conference	Language proficiency	Measurement of proficiency	L1 background	Language of FL	Type of FL	Research method	Main finding
Jiang and Nekrasova (2007)	*The Modern Language Journal*	L1 users & High-proficiency L2 learners	Length of L2 learning & Self-rating	Mixed L1 backgrounds	English	N-grams	Grammatical judgement	Processing advantage for FL over novel language in proficient L2 learners
Ellis, Simpson-Vlach, and Maynard (2008)	*TESOL QUARTERLY*	L1 users & High-proficiency L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	N-grams	Grammatical judgement & Elicited production & Priming	Processing advantage for FL over novel language in proficient L2 learners

Study	Journal/Book/Conference	Language proficiency	Measurement of proficiency	L1 background	Language of FL	Type of FL	Research method	Main finding
Siyanova and Schmitt (2008)	*The Canadian Modern Language Review*	L1 users & High-proficiency L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	Collocations	Phrasal judgement	Processing advantage for FL over novel language in proficient L2 learners
Isobe (2011)	*JACET Kansai Journal*	L1 users & less-proficient L2 learners	Length of L2 learning	Japanese	English	N-grams	Grammatical judgement	Processing advantage for FL over novel language in less-proficient L2 learners
Siyanova-Chanturia, Conklin, and van Heuven (2011b)	*Journal of Experimental Psychology: Learning, Memory, and Cognition*	L1 users & Higher- & Lower-proficiency L2 learners	Self-rating & Length of L2 learning	Mixed L1 backgrounds	English	Binomials	Eye-tracking	Processing advantage for FL over novel language in proficient L2 learners
Kim and Kim (2012)	*TESOL QUARTERLY*	L1 users & High-proficiency L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	Collocations	Self-paced reading	Processing advantage for FL over novel language in proficient L2 learners
Valsecchi, Künstler, Saage, White, Mukherjee, and Gegenfurtner (2013)	*Journal of Eye Movement Research*	L1 users & proficient L2 learners	Length of L2 learning	German	English	N-grams	Eye-tracking	No processing advantage for L2 FL
Wolter and Gyllstad (2013)	*Studies in Second Language Acquisition*	L1 users & proficient L2 learners	Length of L2 learning & Vocabulary size test	Swedish	English	Collocations	Grammatical judgement	Processing advantage for FL over novel language in proficient L2 learners
Edmond (2014)	*Studies in Second Language Acquisition*	L1 users & proficient L2 learners	Length of L2 learning	English	French	N-grams	Naturalness judgement	Processing advantage for FL over novel language in proficient L2 learners

Study	Journal/Book/Conference	Language proficiency	Measurement of proficiency	L1 background	Language of FL	Type of FL	Research method	Main finding
Sonbul (2014)	*Bilingualism: Language and Cognition*	L1 users & proficient L2 learners	Length of L2 learning & Self-rating & Vocabulary level test	Mixed L1 backgrounds	English	Collocations	Eye-tracking	Processing advantage for FL over novel language in proficient L2 learners
Babaei, Najafabadi, and Fotovatnia (2015)	*The Journal of Teaching Language Skills*	Intermediate L2 learners	Length of L2 learning	Iranian	English	N-grams	Self-paced reading	No processing advantage for L2 FL
Wolter and Yamashita (2015)	*Applied Psycholinguistics*	L1 users & intermediate & advanced L2 learners	Length of L2 learning & Self-rating & Vocabulary size test	Japanese	English	Collocations	Lexical decision	No processing advantage for L2 FL
Gyllstad and Wolter (2016)	*Language Learning*	L1 users & proficient L2 learners	Length of L2 learning & Self-rating	Swedish	English	Collocations	Semantic judgement	Processing advantage for FL over novel language in proficient L2 learners
Hernández, Costa, and Arnon (2016)	*Language, Cognition, and Neuroscience*	L1 users & less-proficient L2 learners	Length of L2 learning & Vocabulary level test	Mixed L1 backgrounds	English	N-grams	Phrasal decision	Processing advantage for FL over novel language in less-proficient L2 learners
Shantz (2017)	*Second Language Research*	L1 users & Low-, mid-, and high-proficiency L2 learners	Length of L2 learning & Vocabulary level test	Chinese	English	N-grams	Self-paced reading	Processing advantage for FL over novel language in mid- and high-proficiency L2 learners
Siyanova-Chanturia and Janssen (2018)	*Journal of Experimental Psychology: Learning, Memory, and Cognition*	L1 users & proficient L2 learners	Length of L2 learning	Mixed L1 backgrounds	English	Binomials	Elicited production	No processing advantage for L2 FL
Wolter and Yamashita (2018)	*Studies in Second Language Acquisition*	L1 users & intermediate & advanced L2 learners	Length of L2 learning & Self-rating	Japanese	English	Collocations	Phrasal judgement	Processing advantage for FL over novel language in less- and highly proficient L2 learners

Study	Journal/Book/Conference	Language proficiency	Measurement of proficiency	L1 background	Language of FL	Type of FL	Research method	Main finding
Yi (2018)	*Studies in Second Language Acquisition*	L1 users & proficient L2 learners	Length of L2 learning	Chinese	English	Collocations	Phrasal judgement	Processing advantage for FL over novel language in proficient L2 learners
Northbrook and Conklin (2019)	*Applied Linguistics*	L1 users & less-proficient L2 learners	Length of L2 learning & Self-rating	Japanese	English	N-grams	Phrasal judgement	Processing advantage for FL over novel language in less-proficient L2 learners
Zhao, Yasunaga, and Kojima (2019)	*Journal of Psycholinguistic Research*	L1 users & proficient L2 learners	Length of L2 learning	Chinese	Japanese	Collocations	Functional Near-Infrared Spectroscopy	Processing advantage for FL over novel language in proficient L2 learners
Öksüz, Brezina, and Rebuschat (2020)	*Language Learning*	L1 users & proficient L2 learners	Length of L2 learning & Self-rating & Vocabulary size test	Turkish	English	Collocations	Phrasal judgement	Processing advantage for FL over novel language in proficient L2 learners
Sonbul and El-Dakhs (2020)	*Applied Psycholinguistics*	L1 users & lower- & higher-proficiency L2 learners	Vocabulary size test	Saudi	English	Collocations	Phrasal judgment	Processing advantage for FL over novel language in less- and proficient L2 learners
Kerz, Wiechman, Frinsel, and Christiansen (2020)	*Ongoing research*	L1 users & less-proficient L2 learners	Vocabulary size test & Self-rating & Length of L2 learning	German	English	N-grams	Phrasal judgment	Processing advantage for FL over novel language in less-proficient L2 learners

Note. This table is not complete and includes representative studies.

A number of existing studies have reported the processing advantage for literal FL over novel language in (L1 users and) proficient L2 learners (e.g., Ellis, Simpson-Vlach, & Maynard, 2008; Gyllstad & Wolter, 2016; Jiang & Nekrasova, 2007; Öksüz, Brezina, &

Rebuschat, 2020; Siyanova & Schmitt, 2008, Study 3; Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018; Yi, 2018; Zhao, Yasunaga, & Kojima, 2019). For example, in one of the earliest studies, Jiang and Nekrasova (2007) found in their grammatical judgment study that both L1 and proficient L2 speakers judged frequent n-grams significantly faster and more accurately than infrequent control sequences. Jiang and Nekrasova (2007) further found that the processing advantage for frequent n-grams existed in both uppercase and lowercase. Similar results were found in Wolter and Gyllstad (2013), Gyllstad and Wolter (2016), Yi (2018), and Sonbul and El-Dakhs (2020, Experiment 2).

The processing advantage for L2 literal FL over novel language has been found not only in isolation but also in sentence context (e.g., Edmonds, 2014; Kim & Kim, 2012; Siyanova-Chanturia, Conklin, & van Heuven, 2011b; Sonbul, 2014). For example, in a self-paced reading study, Kim and Kim (2012) had L1 and proficient L2 speakers read sentences embedded with Verb+*out* combinations portion-by-portion. Four levels of combination frequency were observed: (1) low frequency (frequency range: 20 – 100; e.g., *test out*), (2) mid-low frequency (frequency range: 200 – 400; e.g., *wear out*), (3) mid-high frequency (frequency range: 500 – 1,000; e.g., *sort out*), and (4) high frequency (frequency range: 1,200 – 12,000; e.g., *find out*). Kim and Kim (2012) found that both participant groups read higher frequency combinations significantly faster than lower frequency ones. Kim and Kim (2012) further found that while L1 users reliably distinguished between high-frequency, mid-frequency, and low-frequency combinations during reading, L2 learners were only able to distinguish between high- versus low-frequency combinations. This finding suggests that FL facilitation and sensitivity to phrase frequencies are weaker in L2 learners than in L1 users. Similar results were found in eye-tracking research (e.g., Siyanova-Chanturia et al., 2011b; Sonbul, 2014). Sonbul (2014) further found the FL facilitation emerged in the early stage (as indicated by early measure: First Pass Reading Time) and disappeared in the late stage of processing (as indicated by late measures: Total Reading Time and Fixation Count). In another eye-tracking study, Siyanova-Chanturia et al. (2011b) asked L1 and L2 speakers who varied in language proficiency, to read binomials (e.g., *bride and groom*) and relatively infrequent controls (e.g., *groom and bride*) that were embedded in sentences. Siyanova-Chanturia et al. (2011b) found the processing advantage for binomials in L1 and higher-, but not lower-, proficiency L2 speakers both in the early and, in particular, late processing stages.

The processing advantage for L2 literal FL over novel language has been found not only in highly proficient learners but also in less proficient L2 learners (e.g., Hernández, Costa, & Arnon, 2016; Isobe, 2011; Kerz, Wiechmann, Frinsel, & Christiansen, 2020;

Northbrook & Conklin, 2019; Shantz, 2017; Sonbul & El-Dakhs, 2020; Wolter & Yamashita, 2018). In a recent self-paced reading study (sentences presented by portions), Shantz (2017) observed a reversed U-shape developmental trajectory of language proficiency in the processing of frequent n-grams versus infrequent sequences with sentential context. To be more specific, although L2 learners at low-proficiency level did not exhibit processing advantage for FL, mid-proficiency learners showed a processing advantage for FL over novel sequences. Interestingly, although high-proficiency learners also exhibited a processing advantage for FL, the magnitude of the facilitation was found smaller than it in mid-proficiency learners. These findings demonstrate that language proficiency plays an important role in L2 literal FL processing.

All of the studies mentioned above indicate that it is not the type of exposure (e.g., Hernández et al., 2016) but the *amount* of exposure that contributes to the FL processing advantage and the sensitivity to phrase frequencies in L2, with the greater the exposure, the more sensitive to FL and phrase frequencies. It is noteworthy that although L2 learners have been found to exhibit a processing advantage for FL and sensitivity to phrase frequencies, the magnitude of the FL facilitation and their sensitivity to frequencies is smaller than it in L1 users (e.g., Edmonds, 2014; Ellis et al., 2008; Gyllstad & Wolter, 2016; Hernández et al., 2016; Isobe, 2011; Jiang & Nekrasova, 2007; Kerz et al., 2020; Kim & Kim, 2012; Northbrook & Conklin, 2019; Öksüz et al., 2020; Shantz, 2017; Siyanova & Schmitt, 2008; Siyanova-Chanturia et al., 2011b; Sonbul, 2014; Sonbul & El-Dakhs, 2020; Wolter & Gyllstad, 2013; Wolter & Yamashita, 2018).

However, not all studies report on the processing advantage for L2 literal FL over novel language, even when learners are proficient (e.g., Babaei, Najafabadi, & Fotovatnia, 2015; Siyanova-Chanturia & Janssen, 2018; Valsecchi, Künstler, Saage, White, Mukherjee, & Gegenfurtner, 2013; Wolter & Yamashita, 2015). For example, in an eye-tracking study, Valsecchi et al. (2013) found that while L1 users exhibited a processing advantage for literal FSs over novel phrases, this processing advantage did not appear in L2 learners. Similar results were observed in Babaei et al. (2015), Wolter and Yamashita (2015), and Siyanova-Chanturia and Janssen (2018).

Although FL facilitation has been largely found in both L1 and L2 speakers, we cannot deduce from this finding that the processing of L1 and L2 literal FL is driven by the same factor(s). In fact, it has been suggested that L1 and L2 speakers are always sensitive to different distributional information during literal FL processing (e.g., Ellis et al, 2008). For example, using a variety of research methods regarding comprehension (grammatical judgment) and production (naming, priming), Ellis et al. (2008) found that L1 users were more sensitive to mutual information, whereas (proficient) L2 learners were more sensitive to phrase frequencies during literal FL processing.

In addition to phrase frequency and L2 proficiency, existing investigations have explored other factors that could affect L2 literal FL processing. These factors include mutual information (e.g., Yi, 2018), semantic transparency (e.g., Gyllstad & Wolter, 2016), and so forth.

To recap briefly, akin to literal FL in L1, existing investigations into L2 literal FL processing have largely pointed to a significant processing advantage for FL over novel language, with and without sentence context. Furthermore, it has been largely suggested that phrase frequency and language proficiency can affect L2 literal FL processing.

All in all, as shown in most existing studies, both figurative and literal FL enjoy a significant processing advantage over novel language in L2 learners. The processing advantage for L2 FL is primarily driven by phrase frequency and language proficiency, especially for those FSs that are literal.

3.4 The processing of formulaic language in children

Compared with adults, L1 children are less proficient language users. This is mainly because compared with adults, children have less experience with target FSs (e.g., Hallin & Van Lancker, 2017; Jiang, Jiang, & Siyanova-Chanturia, 2020). The limited experience or exposure can slow down the processing of FL and affect the accuracy of FL use (e.g., Hallin & Van Lancker, 2017; Jiang et al., 2020). In addition, as noted by Wray (2002), L1 children process (holistic versus analytic) and learn FL (daily life-based versus classroom-based) differently from L2 learners, and the developmental trajectory of FL processing and learning in these two populations are also dissimilar. Therefore, it can be problematic to use the results obtained from FL processing in adults and directly generalise them to children without verification. The different processing patterns and developmental trajectories of FL in L1 and L2 further warrant investigations into FL processing in children.

In comparison with FL processing in adults, research into the processing of FL in children is extremely scarce. To the best of my knowledge, only four published studies to date have explored this topic (see Table 5 for a summary of relevant studies). These studies have sought to compare children's processing of FSs, such as frequent n-grams, collocations, and proverbs, versus infrequent novel phrases (e.g., Jiang, Jiang, & Siyanova-Chanturia, 2020; Skarabela, Ota, O'Connor, & Arnon, 2021), between children and adults (e.g., Hallin

& Van Lancker, 2017; Jiang et al., 2020), and between children at different ages (e.g., Bannard & Matthews, 2008; Jiang et al., 2020). In this section, I will review these studies in detail.

Before the introduction of existing FL processing investigations in children, I will first report some important findings in previous natural observations/acquisition studies. This is because (1) investigations into FL in children have been almost entirely in natural observations; and (2) findings of these natural observations could help us forge a good understanding of the sensitivity and use of FL in children.

Observations of preschoolers focused on the comprehension and production of spoken language, which is frequently addressed to or repeated by young children (e.g., Bates, Bretherton, & Snyder, 1988; Clark, 1974; Foster-Cohen, 1999; Goldberg, 2006; Hakuta, 1974; Lieven, Behrens, Speares, & Tomasello, 2003; Lieven, Pine, & Barnes, 1992; Lieven, Salomo, & Tomasello, 2009; Nelson, 1973; Peters, 1977; Pine & Lieven, 1993), Researchers found that infants at the age of 20 months had been able to identify and understand multiword clusters (e.g., *give the book to Mommy*) (e.g., Bates, Bretherton, & Snyder, 1988); at the age of two, researchers found infants had started producing meaningful FSs (e.g., *let me down, ride my bike*) (e.g., Clark, 1974; Foster-Cohen, 1999; Lieven et al., 1992; Lieven et al., 2003; Lieven et al., 2009; Nelson, 1973; Peters, 1977; Pine & Lieven, 1993). These meaningful utterances are acquired in children's earlier (and repeated) exposure to rhymes, routines, stories, songs, and interactions with their caregivers, although children are not yet able to analyse these expressions consciously at this stage. After the age of three, children are found to be capable of constructing more complete and complex sentences with grammatical rules as well as a larger number of FSs (e.g., Goldberg, 2006; Hakuta, 1974).

Unlike the language experience in their infancy, children at school-ages begin to receive formal instruction on grammatical knowledge and world knowledge. As a consequence, they tend to recognize and process linguistic items as single words or morphemes, instead of multi-words (e.g., Wray, 2002). Despite this, the interplay between grammatical and formulaic knowledge enables school-agers to analyse, understand, and learn more FSs (that are transparent and compositional) (e.g., Polio & Polio, 1974; Reuterskiöld & Van Lancker Sidtis, 2012). However, analytic knowledge is not always helpful, especially when children are trying to learn and comprehend FSs that are non-compositional and metaphorical (idioms in particular). For example, Caillies and Sourn-Bissaoui (2005) found that after the age of five, children are able to gradually understand literal (compositional) FSs; whereas for those figurative (non-compositional) FSs, children are not able to understand these FSs until the age of seven (see also Levorato & Cacciari, 1999). Kempler, Van Lancker, Marchman, and Bates (1999; see also Anglin, Miller, & Wakefield, 1993; Vulchanova, Vulchanov, & Stankova, 2011) held that children acquire an adult-like understanding of figurative (non-compositional) FSs no earlier than the age of 10. However, some researchers are still not optimistic about this speculation. For example, Cacciari and Levorato (1998) found in their idiom comprehension study that even fifth graders (10.3–11.2 years old) were not able to understand idioms as well as adults.

From the existing natural observations, we can see that L1 children are aware of and able to produce FL at a very young age (in their infancy). However, the awareness and knowledge of figurative FSs develop relatively late in comparison with literal FSs. These findings might shed some light on later investigations into FL processing in children (see Table 5 for a summary of relevant studies). For example, the processing of FL could be different from the novel language in children (e.g., Bannard & Matthews, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020; Skarabela et al., 2021); FL processing could not be invariant with age (e.g., Bannard & Matthews, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020); and the sensitivity to figurative FSs might be established later than the sensitivity to literal FSs (Bannard & Matthews, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020; Skarabela et al., 2021).

Table 5:

Summary of Studies on Formulaic Language Processing in Children

Study	Journal/Book/Conference	Age	Language of FL	Type of FL	Research method	Main finding
Bannard and Matthews (2008)	*Psychological Science*	Two- & Three-year-olds & University students	English	N-grams	Phrase repetition	Processing advantage for FL over novel language in three-year-olds and somewhat in two-year-olds
Hallin and Van Lancker (2017)	*Applied Linguistics*	11-13 year-olds & University students	Swedish	Proverbs	Elicited production	Processing advantage for FL over novel language in 11-13 year-olds
Jiang, Jiang, and Siyanova-Chanturia (2020)	*Applied Psycholinguistics*	Third & Fourth graders & University students	Chinese	Collocations	Eye-tracking	Processing advantage for FL over novel language in fourth graders
Skarabela, Ota, O'Connor, and Arnon (2021)	*Unpublished research, preprint available on Psyarxiv*	11- & 12-month-olds	English	N-grams	Central fixation	Processing advantage for FL over novel language in 11- and 12-month olds

In the first of such investigations, Skarabela et al. (2021) asked infants between 11 and 12 months old to listen to three-word sequences varying in frequency (e.g., high frequency: '*clap your hands*' vs low frequency: '*take your hands*'). Skarabela and colleagues (2021) recorded the infant's visual fixation on the monitor that presenting the sequence. Skarabela et al. (2021) found that these infants showed more interest in and familiarity with high-frequency sequences than those low-frequency ones. This finding suggests that infants as young as 11 months old have established the awareness of FL and the sensitivity to frequencies above the word level during listening.

In a production study, Bannard and Matthews (2008) asked two and three year olds to repeat four-word sequences after hearing them. These sequences were retrieved from the Max Planck Child Language Corpus and varied in phrase frequency (high frequency: *sit in your chair* vs low frequency: *sit in your truck*). Bannard and Matthews (2008) found that children as young as two years old were able to produce frequent sequences with more accuracy than those infrequent counterparts. Apart from their remarkable performance on accuracy, the three-year-old participants were found to utter frequent sequences with markedly faster speed than those infrequent counterparts. This finding suggests that children as young as three years old have established the awareness of FL and the sensitivity to phrase frequencies during speaking.

In another production study, Hallin and Van Lancker (2017) explored the processing of Swedish proverbs (e.g., *Det som inte dödar, härdar* '*That which doesn't kill, toughens*') and relative controls (*Den som inte stannar, ramlar* '*The one who doesn't stop, falls*') with children aged between 11 and 14. Although the sample size of participants and materials was small (five children uttered seven proverbs), these children were found to produce proverbs with significantly fast speed (and more variability in tonal patterns) relative to control phrases. This finding suggests that the processing advantage for figurative FL over novel language manifests itself by the age of 11 during speaking.

More recently, Jiang et al. (2020) used eye-tracking to investigate the processing of Verb+Noun collocations and infrequent controls within sentence context in primary school and university students. The primary school students consisted of two age groups: the third graders (eight years and three months to nine years old) and fourth graders (nine years and four months to 10 years old). The recording of children's eye movements suggests that fourth graders exhibit a significant processing advantage for literal FL over novel language and significant sensitivity to phrase frequencies, although the processing advantage and sensitivity were largely confined to the late processing stage. On the contrary, third graders did not show facilitation of FL or sensitivity to phrase frequencies. These results imply that, unlike in listening and speaking contexts (e.g., Bannard & Matthews, 2008; Skarabela et al., 2021), the FL facilitation and sensitivity to phrase frequencies in reading appear relatively late in L1 children (at around the age of nine, as Jiang and collaborators (2020) speculated). Besides, Jiang et al. (2020) found that age significantly interacted with phrase frequency and phrase type: fourth graders, who were more proficient and experienced readers, read phrases faster than third graders; and more importantly, fourth graders exhibited significant sensitivity to FL and phrase frequencies, while third graders did not. This finding suggests that the processing of FL in children is affected by language proficiency and experience during reading.

To recap briefly, compared with FL processing in adults, research into FL processing in children is extremely scarce. Nonetheless, existing investigations suggest a significant processing advantage for FL over novel language in L1 children. Of note is that the exhibition of FL processing advantage and sensitivity to phrase frequencies, could be (largely) affected by language proficiency and experience, and may manifest itself at different ages for distinct processing modalities (e.g., speaking vs reading) and distinct FL types (e.g., literal FSs vs figurative FSs).

3.5 The processing of modified formulaic language

Existing investigations with both children and adults into FL processing have almost entirely focused on FL in its canonical form. However, according to corpus studies (e.g., Barlow, 2000; Krishnamurthy, Sinclair, Jones, & Daley, 2004; Moon, 1998; Stubbs, 1996), FL is also frequently used in its modified form(s), such as with inflections, negation, transpositions, truncations, and insertions (see Table 6 for examples of FL with different modification types). Moon (1998) estimates that on average, 40% of FSs can be used in their modified form(s), although this proportion may vary among different FSs. In fact, for some FSs, the modified version is even more frequently used than the original form (e.g., Moon, 1998; Schmitt, 2005; Siewierska, Xu, & Xiao, 2010). Moon (1998) also points out that FSs with higher phrase frequency may allow for more modifications.

Table 6:

Examples of Modification Type

Type of modification	Example
Inflection	make it clear → made it clear take the wind out of someone's sail → take the wind out of someone's sails lay down the law → the law is laid down
Negation	spill the beans → spill no beans
Transposition	from rags to riches → from riches to rags
Truncation	every cloud has a silver lining → a silver lining
Insertion	lend a hand → lend a helping hand provide information → provide some information concentrate the mind → concentrate the mind wonderfully

Despite the ubiquity of modified FSs, investigations into the processing of such sequences are extremely scarce. To date, there have been only a handful of published studies that explored this topic (see Table 7 for a summary of relevant studies). These studies focus by and large on L1 idioms with inflections (e.g., Geeraert, Baayen, & Newman, 2017; Kyriacou, Conklin, & Thompson, 2020; McGlone, Glucksberg, & Cacciari, 1994), L1 idioms with insertions (e.g., Haeuser, Baum, & Titone, 2020), L1 frequent n-grams with insertions (e.g., Molinaro, Canal, Vespignani, Pesciarelli, & Cacciari, 2013), and L1 collocations with insertions (e.g., Vilkaitė, 2016a). Existing research has also turned to L2 collocations with insertions (e.g., Vilkaitė & Schmitt, 2019).

Table 7:

Summary of Studies on Modified Formulaic Language Processing

Study	Journal/Book/Conference	Language of FL	Type of FL	Type and Degree of modification	Research method	Main finding
McGlone, Glucksberg, and Cacciari (1994)	*Discourse Processes*	English	Idioms	Inflection and transposition with one modification degree	Self-paced reading	The figurative meaning was not prioritised in modified idioms
Molinaro, Canal, Vespignani, Pesciarelli, and Cacciari (2013)	*Language and Cognitive Processes*	Italian	N-grams	Insertion with one modification degree	Self-paced reading & EEG	The final word in modified FL was more predictable than in its original form

Study	Journal/Book/Conference	Language of FL	Type of FL	Type and Degree of modification	Research method	Main finding
Vilkaitė (2016a)	*Journal of Experimental Psychology: Learning, Memory, and Cognition*	English	Collocations	Insertion with one modification degree	Eye-tracking	The processing advantage for FL over novel language existed in its modified form
Geeraert, Baayen, and Newman (2017)	*Proceeding of the 13th Workshop on Multiword Expressions*	English	Idioms	Inflection and truncation with one modification degree	Eye-tracking	FL in its modified form was processed equally fast as in its original form
Vilkaitė and Schmitt (2019)	*Applied Linguistics*	L2 English	Collocations	Insertion with one modification degree	Eye-tracking	No processing advantage for modified L2 FL
Haeuser, Baum, and Titone (2020)	*Applied Psycholinguistics*	English	Idioms	Insertion with one modification degree	Eye-tracking	The figurative meaning was not prioritised in modified idioms
Kyriacou, Conklin, and Thompson (2020)	*Language and Speech*	English	Idioms	Inflection with one modification degree	Eye-tracking	The processing advantage for FL over novel language existed in its modified form

In the earliest of such studies, McGlone, Glucksberg, and Cacciari (1994, Experiment 3) explored the semantic processing of modified L1 ambiguous idioms in a self-paced reading experiment (line-by-line). A group of L1 adults were asked to read sentences embedded with idioms in their original and modified forms. The modified forms were created by inflecting (e.g., *Sam spilt the beans* vs *Sam didn't spill a single bean*), replacing (e.g., *He had two left feet* vs *He had three left feet*), or transposing (e.g., *He had gone from rags to riches* vs *He had gone from riches to rags*), and so forth. McGlone and colleagues (1994) found that L1 readers processed the figurative meaning significantly faster than the literal meaning of idioms in their original form; however, for the idioms in their modified form, the faster processing for the figurative meaning disappeared. Despite the variety of modification types, this study implies that the semantic processing of figurative FL might be dissimilar in its original and modified forms. This study further suggests that at least for figurative FL, modification can jeopardise the formulaicity of FL, thus affect our sensitivity to FL to some degree. Similar results were found in Haeuser et al. (2020).

However, other investigations found something different. In an eye-tracking study, Geeraert, Baayen, and Newman (2017) found that L1 adults processed modified idioms (with truncations, inflections, etc.) equally fast as the original ones. This result indicates that modification does not entirely destroy our sensitivity to FL.

This supposition was further evidenced in a recent eye-tracking study. In this study, Kyriacou, Conklin, and Thompson (2020) had L1 users read ambiguous idioms and relative controls in their active (e.g., *he kicked the bucket* vs *he kicked the apple*) and passive forms (e.g., *the bucket was kicked* vs *the apple was kicked*). All phrases were embedded in sentence context, with a bias towards the figurative meaning of the respective idiom. The eye movement patterns of L1 users suggest a significant processing advantage for idioms over relative controls in both original (active) and modified (passivized) forms. This finding suggests that the processing advantage for figurative FL over novel language holds even when it is modified.

The processing advantage for modified FL was evidenced not only in figurative FSs (e.g., Kyriacou et al., 2020) but also in literal FSs (e.g., Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). For example, in a self-paced reading study (word-by-word presentation), Molinaro, Canal, Vespignani, Pesciarelli, and Cacciari (2013) asked L1 users to read sentences embedded with Italian collocational complex prepositions (CCPs). CCP is a kind of frequent sequence consisting of a preposition, a content word, and a second preposition (e.g., *in the hands of*). Each participant read the CCP either in its original (e.g., *in contrasto con* '*in contrast to*') or modified form (by adding an adjective before the noun, e.g., *in <u>chiaro</u> contrasto con* '*in <u>clear</u> contrast to*'). Molinaro et al. (2013, Experiment 1) found that CCPs in their modified form were processed significantly slower than in their original form. However, in their follow-up EEG experiment, Molinaro and collaborators (2013, Experiment 2) observed a smaller N400 response to the word after CCPs in their modified form than in their original form. This result implies that, at least for literal FL, although modification may decelerate the processing with more effort required for the comprehension of the enriched meaning, modification does not completely disrupt the predictability or formulaicity of FL.

In a more recent and better-designed study, Vilkaitė (2016a) compared the eye movements of L1 users when they were reading sentences embedded with frequent Verb+Noun collocations and infrequent novel phrases. Each phrase was presented either in its original form (e.g., collocation: *provide information* vs control: *compare information*) or modified form by adding three words to the middle of the phrase (e.g., collocation: *provide <u>some of the</u> information* vs control: *compare <u>some of the</u> information*). Vilkaitė (2016a) found a significant processing advantage for collocations both in their original and modified forms.

This finding suggests that the processing advantage for literal FL over novel language holds even when it is modified.

All of the investigations mentioned above focus on modified FL processing in L1. To the best of my knowledge, there has been only one published study that has explored modified FL processing in L2. In an eye-tracking study, Vilkaitė and Schmitt (2019) compared the processing of frequent collocations and infrequent controls in their original and modified forms in L1 and proficient L2 speakers. Vilkaitė and Schmitt (2019) found that while L1 users exhibited a processing advantage for modified collocations over relative controls, L2 learners did not show the facilitation of modified collocations. This result implies that modification might jeopardise L2 learners' sensitivity to FL to a larger degree in comparison with L1 users.

From the existing literature of original and modified FL processing, we know that, by and large, FL enjoys a significant processing advantage over novel language in L1 adults, L2 learners, and L1 children. We also know that FL processing is largely affected by phrase frequency and language proficiency. Besides, recent investigations have accumulated to show that the processing advantage can be extended to FL in its modified form.

Apart from these important findings, there are a number of prominent gaps that need to be addressed. First, compared with the processing of FL in its original form, research into modified FL processing is still in its infancy. Whether or not modified FL (literal FSs in particular) enjoys a processing advantage over novel language, remains largely unexplored. This highlights the need for further investigations to compare FL versus novel language processing in their modified form.

Second, according to corpus studies (e.g., Barlow, 2000; Krishnamurthy, Sinclair, Jones, & Daley, 2004; Moon, 1998; Stubbs, 1996), the form of FL, literal FSs in particular, can be quite loose. The number of intervening words between the constituents of a FS (i.e., insertion length/span) varies among different FSs and languages. For example, in English collocations, the insertion can be expanded up to four words (e.g., Krishnamurthy et al., 2004; Sinclair, 1998); while in Chinese splittable compounds, the insertion can be expanded to even more than 10 words (e.g., Siewierska et al., 2010). A question thus is posed: does the insertion length affect modified FL processing? Existing investigations, however, only looked at modified FSs with one word inserted (e.g., Haeuser et al., 2020; Molinaro et al., 2013), or three words inserted (e.g., Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019); no studies to date have compared the processing of modified FL with different insertion lengths, or explored modified FL processing with the increase of insertion length. It has been suggested that the distance between nonadjacent yet mutually dependent elements could systematically affect language processing (e.g., Santelmann & Jusczyk, 1998, Experiments 3–5); and the effect of insertion length could only be active within a (very) limited processing domain/window: with the increase of insertion length, the processing could be retarded and the dependency between the elements could be weakened. Therefore, an investigation into the processing of modified FL with different insertion lengths is urgently needed.

Third, existing investigations into modified FL processing have almost entirely focused on L1 users, with language learners, such as L2 adults and L1 children, being largely disregarded. These populations, however, are worth studying because FL plays a crucial role in L1 and L2 acquisition (e.g., Tomasello, 2003; Wray, 2002). In addition, we can not directly extend the findings with L1 adults to L1 children or L2 learners, as modified FL processing in these populations may be different from that in L1 adults (e.g., Vilkaitė & Schmitt, 2019).

Fourth, for L2 learners, existing investigations have by and large focused on learners with high proficiency, and from one or mixed language background(s). As language proficiency is suggested to play an important role in L2 FL processing (e.g., Matlock & Heredia, 2002; Shantz, 2017; Sonbul & El-Dakhs, 2020), it is warranted to compare modified FL processing in learners of different proficiencies, especially those of lower proficiency. Besides, as suggested by Conklin and Carrol (2019), L1 background plays a role in L2 FL processing; thus, modified L2 FL processing in learners from distinct L1 backgrounds also warrants further exploration.

Finally and crucially, existing investigations have almost entirely pertained to alphabetic languages, English in particular. How FL is processed in other languages, such as Chinese, remains largely disregarded. Chinese, however, has a writing system that is markedly different from English and other alphabetic languages. Therefore, directly extending the findings pertinent to English or other alphabetic languages to Chinese without verification is problematic at best. This highlights the need for further investigations into FL processing in Chinese.

3.6 The processing of Chinese formulaic language

3.6.1 Chinese vs alphabetic languages

Chinese is different from English and other alphabetic languages in a number of ways. First, unlike the writing system of alphabetic languages (e.g., English, French, Italian, Spanish, Swedish, Russian, etc.), Chinese is logographic. The reading unit in Chinese text is one character (different from the unit – one word – in English). A Chinese character could be a pictograph (e.g., 月 '*moon*': ), an ideograph (e.g., 好 '*good*': 女 '*woman*' + 子 '*child*' = 好 '*good*'), or most likely (around 90% of the totality, see Rayner, Pollatsek, Ashby, & Clifton, 2012), an abstract symbol combined with sound and meaning radicals (e.g., 妈 '*mother*' = 女 '*woman*' (semantic radical) + 马 '*ma*' (phonetic radical)). It is noteworthy that the visual abstractness and complexity of Chinese text require a special set of metalinguistic skills, which could make Chinese harder to learn or understand than alphabetic languages (e.g., Chung & McBride-Chang, 2011; Ho, Chan, Lee, Tsang, & Luan, 2004; Ma & Li, 2015; Rayner et al., 2012).

Second, there is a large number of homophones and homographs in the Chinese language. In fact, among the top 7,000 most frequently used characters in modern Mandarin, approximately five distinct characters share the same pronunciation (Li, Wang, Tong, & McBride, 2017). Besides, according to Lee and Chen (1997), more than 1,000 characters have more than one pronunciation (could be in either the same syllable but with a different tone, or another entirely different syllable). A character with different pronunciations may carry distinct meanings, even though they are in identical form. These two properties of Chinese, homophony and homography, may pose great challenges to learners when they are trying to identify and distinguish the similar characters/words that are encountered in language learning, processing, and use.

Third, unlike alphabetic and syllabic languages, Chinese does not mark word boundaries with spaces in the text. The space between two characters in most cases, however, is not the essential boundary for Chinese words. This is because a single character (which is the reading unit) per se is not regarded as a word in most instances. In fact, according to the authoritative statistics (现代汉语常用词表课题组 'Research Group for the Lexicon of Common Words in Contemporary Chinese', 2008, p. 2), among the common 56,008 Chinese words, only 5.68% of them are single characters (3,181); most words (40,351; 72.04%) are made up of two characters; others are made up of three characters (6,459; 11.53%), four characters (5,855; 10.45%), or five and more characters (162; 0.29%). It is suggested that the variability of word length may obscure the recognition of the information around word boundaries (e.g., Ma, Li, & Rayner, 2014; Wang, Angele, Ma, & Li, 2020).

Finally, modifications of FL, especially insertions, may be more common and acceptable in Chinese than in English and other alphabetic and syllabic languages. Apart from common FSs such as binomials, collocations, and idioms, there is a special FL type in the Chinese language – 离合词 ('splittable compound', e.g., *帮忙* '*offer help*'). According to the statistics by Siewierska, Xu, and Xiao (2010), there are more than 3,000 of this type in the Modern Chinese Dictionary (Institute of Linguistics at the Chinese Academy of Social Sciences, 2002). A splittable compound consists of two words, normally a verb and an object. Compared with other FSs, this structure is more elastic and allows for (1) longer insertions (around 10 words in the middle of the compound; the inserted words could be particles, quantifiers, classifiers, modifiers, etc.) (Siewierska et al., 2010); and (2) transpositions in the form, without jeopardising the original meaning. However, the flexibility of splittable compounds may not help with Chinese language learning. For example, Wang (2005) found that L2 Chinese learners always treat the structure as a non-insertable word, misuse the structure, or even avoid using it.

To recap briefly, compared with English and other alphabetic languages, Chinese has a distinct writing system; Chinese text might be more abstract and complex; Chinese has a large number of homographs and homophones; Chinese does not mark word boundaries; and some Chinese FSs could be less fixed and may allow for greater modifications. All of these characteristics are likely to pose challenges to learning and processing Chinese as L2.

3.6.2 Formulaic language processing in Chinese

Compared with the boom of FL processing in alphabetic languages (reviewed in previous sections), research into Chinese FL processing is still in its infancy, with less than a handful of published studies available (see Table 8 for a summary of relevant studies). The majority of these studies have by and large focused on L1 figurative FSs (e.g., Liu, Li, Shu, Zhang, & Chen, 2010; Shen, Wang, & Liu, 2017; Yu et al., 2016; Zhang, Yang, Gu, & Ji, 2013; Zhou, Zhou, & Chen, 2004), with only a few studies looking at the processing of L1 literal FSs (e.g., Kong, Zhang, & Zhang, 2016; Jiang et al., 2020), and one study exploring Chinese FSs in L2 (Yi, Lu, & Ma, 2017).

Table 8:

Summary of Studies on Chinese Formulaic Language Processing

Study	Journal/Book/Conference	Participant	Type of FL	Research method	Main finding
Zhou, Zhou, and Chen (2004)	*Brain Topography*	L1 users	Idioms	EEG	Processing advantage for FL over novel language in Chinese
Liu, Li, Shu, Zhang, and Chen (2010)	*Journal of Neurolinguistics*	L1 users	Idioms	EEG	Processing advantage for FL over novel language in Chinese
Zhang, Yang, Gu, and Ji (2013)	*Journal of Neurolinguistics*	L1 users	Idioms	EEG	Decomposability affected Chinese idiom processing
Kong, Zhang, and Zhang (2016)	*Psychological Reports*	L1 users	N-grams	Grammatical judgement	Processing advantage for FL over novel language in Chinese
Yu et al. (2016)	*Language, Cognition, and Neuroscience*	L1 users	Idioms	Eye-tracking	Processing advantage for FL over novel language in Chinese
Shen, Wang, and Liu (2017)	*Chinese Journal of Applied Linguistics*	L1 users	Idioms	EEG	Familiarity affected Chinese idiom processing
Yi, Lu, and Ma (2017)	*Second Language Research*	L1 users & proficient L2 learners	N-grams	Eye-tracking	Processing advantage for FL over novel language in L1 and L2 Chinese
Jiang, Jiang, and Siyanova-Chanturia (2020)	*Applied Psycholinguistics*	L1 users (adults & children)	Collocations	Eye-tracking	Processing advantage for FL over novel language in Chinese

Despite the scarcity of relevant research, existing studies point to a significant processing advantage for FL over novel language in Chinese idioms (e.g., Liu et al., 2010; Yu et al., 2016; Zhou et al., 2004), n-grams (e.g., Kong et al., 2016), and collocations in L1 users (e.g., Jiang et al., 2020), as well as n-grams in proficient L2 learners (Yi et al., 2017).

In addition, existing studies have identified potential factors that can affect Chinese FL processing. One such factor is decomposability, with higher decomposability idioms being processed faster than lower decomposability ones (e.g., Zhang et al., 2013). Another factor is familiarity, with more familiar idioms processed faster than less familiar ones (e.g., Shen et al., 2017). In addition, mutual information can also come into play (e.g., Yi et al., 2017).

To recap briefly, existing investigations into Chinese FL processing largely suggest a significant processing advantage for FL over novel language. The processing advantage has been evidenced in figurative as well as literal FSs, not only in L1 users but also in L2 learners.

Despite these findings, investigations into the processing of Chinese FL, literal FSs in particular, are still scarce. Crucially, to the best of my knowledge, no investigation has so far explored the processing of Chinese FL in its modified form, let alone comparing FL processing with different modification degrees, such as different insertion lengths. This highlights further explorations of modified Chinese FL processing in L1 and L2.

3.7 What theory can predict the processing advantage for formulaic language over novel language?

The processing advantage for FL over novel language can be predicted by a number of theories and models. These theories and models account for the FL advantage from different perspectives, such as language representation (exemplar-based theory, e.g., Abbot-Smith & Tomasello, 2006; Ambridge, 2020a, 2020b; Bod, 1998, 2006; Goldinger, 1996; Pierrehumbert, 2001; connectionist theory, e.g., Christiansen & Chater, 1999; Ellis, 2001, 2003; Elman, 1990, 1991; MacWhinney, 1998; Rumelhart & McClelland, 1986; Seidenberg, 1994), language learning and input (statistical learning theory, e.g., Arnon, 2019; Christiansen, 2019; Saffran, 2003; probabilistic theory, e.g., Gregory et al., 1999; Jurafsky, 1996, 2003; McDonald & Shillcock, 2003a, 2003b; Seidenberg & MacDonald, 1999), and formulaicity of language (idiom principle, e.g., Sinclair, 1991).

Among the proposed theoretical accounts, the *Usage-based Theory* is deemed to be the most appropriate theoretical framework to accommodate the FL advantage and frequency effects in the literature (e.g., Carrol & Conklin, 2020; Conklin & Carrol, 2020). First, this theory highlights the crucial role of frequency or exposure (experience) in the processing of linguistic units of all shapes and sizes (e.g., Ellis, 2002a, 2017; Ellis & Ogden, 2017; Ellis et al., 2016). Second, this theory incorporates language acquisition, representation, processing, formulaicity, and the role of input information (e.g., Bybee, 1998, 2006, 2011; Bybee & Beckner, 2012; Bybee & Hopper, 2001; Goldberg, 2006; Langacker, 1987; Tomasello, 2003). In this case, we can draw on this theory to discuss frequency effects from not one but different perspectives, such as frequency effects in language learning, representation, processing, and use. Furthermore, it is noteworthy that this theory is not in conflict with other emergentist theories (e.g., Conklin, 2020; Van Lancker Sidtis, 2015; Wiechmann et al., 2013), such as exemplar-based theory (e.g., Abbot-Smith & Tomasello, 2006; Ambridge, 2020a, 2020b; Bod, 2006; Goldinger, 1996; Pierrehumbert, 2001), connectionist theory (e.g., Christiansen & Chater, 1999; Elman, 1991; MacWhinney, 1998; Rumelhart & McClelland, 1986; Seidenberg, 1994), and statistical learning theory (e.g., Arnon, 2019; Christiansen, 2019; Saffran, 2003). In fact, the usage-based theory borrows and embraces the fundamental concepts from all of these theories. For example, the usage-based theory borrows the concept of *exemplar* from the exemplar-based theory, the concept of *network* from the connectionist theory, and the concept of *cumulative* effect from the statistical learning theory.

According to the usage-based theory, frequency has an ongoing and cumulative effect on the acquisition, representation, processing, and use of language (e.g., Bybee & Beckner, 2012). Our brain keeps track of any elements that occur regularly, as well as multiple elements that frequently co-occur with each other (e.g., Gregory et al., 1999; Jurafsky, 1996, 2003; McDonald & Shillcock, 2003a, 2003b; Seidenberg & MacDonald, 1999). The tracking makes frequent elements entrenched in our memory and renders co-occurring elements into chunks, such that our processing of these elements becomes more efficient (e.g., Haiman, 1994).

For example, a word becomes stored in our memory from considerable experience with or repeated exposure to it. Our brain detects and registers information of this word from sensory input, such as reading and listening. The registered information includes not only the form (such as morpheme and phoneme) and meaning(s) of this word but also contextual details (such as co-occurring word/s) and linguistic abstractions (rules about how to use this word) (e.g., Ambridge, 2020b). Instead of being unorganised, each occurrence of a word is systematically marked in a cognitive network in our memory, along with the connections between its components and with other words(s) that often co-occur (e.g., Pierrehumbert, 2001). This network is always non-linear, incremental (e.g., Ellis, 2001), recurrent (e.g., Elman, 1990, 1991; MacWhinney, 1998), hidden-layered (e.g., Ambridge, 2020a, 2020b; MacWhinney, 2020), and recursive (e.g., Christiansen & Chater, 1999). The mental representation of this word is further updated on the basis of subsequent input. The more frequently the word is encountered, the more entrenched it becomes in our memory, with the connections between its components and even between its co-occurring companion(s) becoming stronger (e.g., Bybee & Beckner, 2012; Ellis, 2002a, 2017; Goldberg, 2006; Langacker, 1987; Rumelhart & McClelland, 1986; Tomasello, 2003). The entrenchment and strong connections facilitate eventual access to this word and its components, and even, co-occurring companion(s). In this case, a frequent word is more entrenched in our memory and its processing is supposed to be faster than an infrequent word.

Importantly, proponents of the usage-based theory further argue that the above processes are not unique to items at the single word level (or even smaller units); rather, these processes should also persist in a language unit that is above the word level (e.g., Bybee & Beckner, 2012). That is to say, the usage-based principles apply to the learning, representation, and processing of multi-word sequences, whether compositional or not. To be more specific, a frequent or conventional phrase is always more entrenched in our memory and the connections between its components are stronger than a novel phrase, which is infrequent or unconventional. The entrenchment and connections greatly facilitate the processing of frequent phrases and their components, compared to novel sequences (e.g., Wray, 2002; Frost et al., 2015). Of note is that the usage-based account of the representation and processing of multi-word sequences does not completely reject syntactic analysis. Instead, this theoretical account stresses that multi-word sequences must be represented in and conform to the grammar (e.g., Ambridge, 2020b; Bybee & Beckner, 2012; Culicover et al., 2017). Therefore, it is deemed unnecessary to distinguish or establish two distinct systems for the representation and processing of analysable vs unanalysable multi-word sequences (e.g., Ambridge, 2020b; Bybee & Beckner, 2012; Lieven et al., 2020; MacWhinney, 2020). As noted by Bybee and Beckner (2012), since a multi-word sequence is represented in a rich, multifaceted, and hidden-layered network (see also Ambridge, 2020a, 2020b), there is no need to choose between storage of an unanalysable unit and an analysable unit.

3.8 Summary

This chapter allows for a number of conclusions to be drawn. The processing advantage for FL, figurative or literal, over novel language, has been widely documented in L1 adults and, to a lesser extent, in proficient L2 learners and L1 children, across different modalities (such as reading, listening, and speaking). This processing advantage for FL not only exists in alphabetic languages (such as English) but also in a logographic language – Chinese. Second, the processing advantage for FL has also been evidenced in its modified form(s), such as FSs with insertions, FSs with inflections. However, no studies have to date put the degree of modification (e.g., FS with short insertion vs FS with long insertion) to the test. Third, although researchers have started to explore modified FL processing in L2, they almost entirely focused on high proficiency learners, with those of lower proficiency largely disregarded. Finally, existing evidence of FL processing advantage almost entirely comes from adults. FL processing in children, thus, remains largely unexplored. Based on the gaps identified, three studies have been conducted with the aims of exploring the processing of Chinese FL versus novel language with different modification degrees in L1 adults, L2 learners with a range of proficiency, and L1 children. The three studies will be reported in Chapters 5, 6, and 7, respectively.

Chapter 4

Research Methods of Formulaic Language Processing: Self-paced Reading and Eye-tracking

4.1 Introduction

Compared with other modalities of language processing, such as listening and speaking, reading is relatively complex (e.g., Rayner & Pollatsek, 1989; Robeck & Wallace, 2018; Roberts & Siyanova, 2013). This is because reading requires hierarchical processing, which involves complex processes of symbol decoding and literal and even higher-level comprehension (e.g., Robeck & Wallace, 2018). Take formulaic language (FL), for example: when reading a formulaic sequence (FS), we first need to recognize its constituents by translating the signs and symbols into meanings and then integrate the constituents into larger meaningful units, such as a phrase.

A number of research methods can be adopted to investigate the process of reading. Among these methods, self-paced reading and eye-tracking have been frequently used in FL reading research (e.g., Conklin, Pellicer-Sánchez, & Carrol, 2018; Godfroid, Winke, & Conklin, 2020; Just, Carpenter, & Woolley, 1982; Marsden, Thompson, & Plonsky, 2018; Rayner, 1998). In this chapter, I will introduce these two methods and explain why they have been chosen for the current investigations.

4.2 Self-paced reading

4.2.1 What is self-paced reading and what does it tell us?

Self-paced reading is a computer-assisted research method in which participants are asked to read sentences at their own pace. Each sentence is normally broken down into words or portions (i.e., more than one word). After reading one word or portion, the reader needs to press a response button to initiate the following word or portion.

In a self-paced reading experiment, a sentence can be presented in three ways: (1) in stationary-windows, or (2) in cumulative-windows, or (3) in moving-windows. In the stationary-window presentation, each word/portion is displayed sequentially in the same location where the previous word/portion disappears and is replaced by the following word/portion (e.g., Just et al., 1982). In the cumulative-window presentation, all words/portions previously displayed remain on the screen when the following word/portion appears. However, this presentation is not recommended, as it entangles the reading time of the current word/portion with the re-reading of the previous word(s)/portion(s) (e.g., Rayner, 1998). In the moving-window presentation, only one word/portion is displayed at a time and when the reader presses the button to initiate the next word/portion, the previous word/portion will be replaced by dashes of equal length (e.g., Just et al., 1982). Compared with other forms of presentation, the moving-window is highly recommended and most frequently used, as it (directly) reflects the processing of the information in the current window and it is more analogous to the presentation of normal text, such as scanning, and the way of natural reading (e.g., Rayner, 1998).

In a self-paced reading experiment with the moving-window presentation, the reading time of the word/portion in the current window is measured between the two corresponding button presses, with the one that initiates the current window and the next one that ends the current window and/or initiates the next window. The reading time for a word/portion is largely modulated by the properties of the word/portion, such as frequency, length, familiarity, and novelty (e.g., Just et al., 1982). For example, a long and infrequent word/portion is often read relatively slower than a short and frequent word/portion.

To recap briefly, self-paced reading is a research method for reading studies that involve sentence context. This method could tell us how long it takes and how hard it is to process a specific word or phrase, by measuring the corresponding reading time.

4.2.2 Why I chose self-paced reading

Self-paced reading was chosen for the current investigation for several reasons. First, this research method offers a good way to look into the online processing of words and phrases (e.g., a FS or its final word) during sentence reading (e.g., Marsden et al., 2018).

Second, this research method has been widely used in language processing research (for a review, see Marsden et al., 2018), especially in FL processing (e.g., Babaei, Najafabadi, & Fotovatnia, 2015; Conklin & Schmitt, 2008; Gonnerman & Hayes, 2005; Kim & Kim, 2012; Libben & Titone, 2008; McGlone, Glucksberg, & Cacciari, 1994; Millar, 2011; Molinaro et al., 2013; Schmitt & Underwood, 2004; Shantz, 2017; Tremblay et al., 2009, 2011; Vespignani et al., 2009), making it possible and reasonable to compare the results of the current investigation with existing ones.

Third, this research method can be programmed on software, such as DMDX (Forster, 2002), which was used for the current investigation, E-prime (Psychology Software Tools Inc., 2020), OpenSesame (Mathôt, Schreij, & Theeuwes, 2012), PsyScope (Bonatti, 2020), PsychoPy (Open Science Tools Ltd, 2018), and Linger (Massachusetts Institute of Technology, 2020). This software is relatively affordable, available, and user-friendly to experimenters (e.g., Marsden et al., 2018).

Last but not least, this research method provides findings regarding reading time that are comparable to findings with more expensive or labour-intensive techniques for reading research, such as eye-tracking (e.g., Just et al., 1982; Keating & Jegerski, 2015; Marsden et al., 2018; Rayner, 1998), which was also adopted for the current investigation.

4.3 Eye-tracking

4.3.1 What is eye-tracking and what does it tell us?

Eye-tracking is a computer-assisted, real-time technique by which researchers can track and measure what and where the participant is looking at during reading and how long and how many times he/she lands the eye(s) in that position or region (e.g., Conklin et al., 2018). Importantly, meta-linguistic knowledge and strategy are less likely to be involved on the track of the reading process (e.g., Conklin et al., 2018; Rayner, 1998, 2009).

Eye-tracking while reading has been steeped in research history for over 100 years. It can be traced back to 1879 when Louis Emile Javal, a professor at the University of Paris, observed eye movements in reading French text. He found some fundamental characteristics of the eye movements, such as that a reader could not extract information except during pauses (i.e., fixations) (see Huey, 1908).

However, eye-tracking was not widely used in reading research until the mid-1970s when recording systems of eye movements (i.e., eye-trackers) were advanced to be more precise and more available to researchers (e.g., Rayner, 1998). Among these systems, EyeLink (SR Research Ltd., 2020), which was applied in the current investigation, Tobii (Tobii Group, 2020), and SensoMotoric Instruments (SMI, 2020) are the most frequently used eye-trackers.

These eye-trackers broadly provide two types of data: fixations and saccades (e.g., SR Research Ltd, 2017; see also Conklin et al., 2018; Rayner, 1998, 2009; Rayner & Pollatsek, 1989; Rayner, Pollatsek, Ashby, & Clifton, 2012). Fixations refer to any point “where the eye is stationary on a target”; and saccades refer to the paths and movements of the eye “from one location to another” (Conklin et al., 2018, p. 66). During reading, our eyes stop to process the information at a certain location (e.g., a word, or a phrase) and then move to another place where new information is required to be processed. The movement can be forward and also backward (i.e., regress), especially when the previously encountered text is difficult to process at a time (e.g., Conklin et al., 2018; Rayner, 1998, 2009). In most cases, however, text-based reading studies tend to focus on fixation measures only, as they are more sensitive to linguistic factors than saccades (e.g., Staub & Rayner, 2007).

In FL processing studies, four fixation measures are commonly used. They are First Fixation Duration (FFD), First Pass Reading Time (FPRT, also known as Gaze Duration), Dwell Time (DT, also known as Total Reading Time) and Fixation Count (FC). FFD refers to the duration of the first fixation on a word, regardless of whether it is the only fixation or the first of multiple fixations on the word (e.g., Rayner, 1998). FPRT refers to the sum of all fixation durations made within a region of interest (such as a word or a phrase) until the gaze exits either to the left or to the right. This measure suggests how long the target word or phrase is fixated by the reader for the first encounter (e.g., Roberts & Siyanova-Chanturia, 2013). DT refers to the sum of all fixation durations made within a region of interest (e.g., a word, or a phrase), including reading forward and backwards. Liversedge, Paterson, and Pickering (1998) suggest that this measure may contain the initial processing time and the recovering time for processing difficulties. They further note that if an effect is observed only for this measure (but not for early measures), it suggests that the manipulation might only influence late-stage processing rather than early-stage processing. FC refers to all fixations made within a region of interest. This measure does not concern processing time; rather, it indicates how many times the word or phrase is fixated on (e.g., Roberts & Siyanova-Chanturia, 2013). This measure reflects the processing difficulty and cognitive load incurred by a word or phrase (e.g., Henderson & Ferreira, 1990). As noted by Rayner (1998), a single measure of fixation of one word is a pale reflection of the reality of cognitive processing. Thus, he recommends that a reading study should observe multiple measures of different processing stages and examine different target locations (i.e., regions/areas of interest).

To reflect different stages of processing, fixation measures can be divided into two categories: early measures and late measures (e.g., Conklin et al., 2018; Inhoff & Radach, 1998; Liversedge, Paterson, & Pickering, 1998; Pellicer-Sánchez & Siyanova-Chanturia, 2018; Rayner, 1998; Rayner, Sereno, Morris, Schmauder, & Clifton, 1989; Roberts & Siyanova-Chanturia, 2013; Siyanova-Chanturia, 2013; Winke, Godfroid, & Gass, 2013). The early measures reflect initial lexical access and indicate automatic processing, while the late measures reflect the reanalysis of text information and indicate more controlled and strategic processing (e.g., Patterson et al., 1999; Pellicer-Sánchez & Siyanova-Chanturia, 2018; Rayner et al., 1989; Roberts & Siyanova-Chanturia, 2013). The early measures include measures such as FFD and FPRT (note that this measure is sometimes considered a middle-stage measure, see e.g., McDonald & Shillcock, 2003a, 2003b; Vonk, Radach, & van Rijn, 2000). The late measure include measures such as DT and FC.

In FL processing studies, two regions of interest (also known as 'areas of interest') are recommended: the whole phrase and its final word (e.g., Carrol & Conklin, 2014b; Conklin et al., 2018; Conklin & Pellicer-Sánchez, 2016). The rationale for examining the final word is that, although the final word is always identical or matched in a FS and its control phrase, the processing of this word in the two phrase conditions could not be identical, but different (e.g., Carrol & Conklin, 2014b; Conklin et al., 2018; Conklin & Pellicer-Sánchez, 2016; Jiang et al., 2020). This is because in most cases, the final word in FL is more predictable than in novel language. The final word is thus considered the predicted locus of the processing differences (e.g., Vilkaitė, 2016a). Although examining the final word is informative, some detailed yet important cases could still be missed, such as the processing difference/advantage taking effect not on the final word only, but also on the component word(s) before it. In this case, the whole phrase should also be examined. In fact, as suggested by Carrol and Conklin (2014b), the whole phrase, which allows for direct comparison between FSs and novel phrases, is even more informative than its final word. It is noteworthy that the two interest areas are not always sensitive to the same eye-tracking measure(s). While the final word is sensitive to early and/or late measures (e.g., Conklin et al., 2018; Conklin & Pellicer-Sánchez, 2016), the whole phrase is primarily sensitive to late measures (e.g., Conklin & Pellicer-Sánchez, 2016; Jiang et al., 2020; Siyanova-Chanturia, 2013; Siyanova-Chanturia et al., 2011a). The different sensitivity to distinct eye-tracking measures suggests that it is necessary to examine both the whole phrase and its final word, rather than only one of them.

No matter what and how many fixation measures are chosen for the target region to observe, they are interpreted on the same assumption. That is, the number and/or duration of fixations can immediately and directly reflect the cognitive effort required to process the information within the target region during reading (e.g., Conklin & Pellicer-Sánchez, 2016; Just & Carpenter, 1980; Pellicer-Sánchez & Siyanova-Chanturia, 2018). This means longer durations and/or more fixations suggest greater processing effort, while shorter durations and/or fewer fixations suggest less processing effort. The amount of processing effort is associated with lexical properties, such as frequency, length, contextual constraint (e.g., Rayner, 1998, 2009). For example, highly frequent phrases are easier to process with shorter and/or fewer fixations than infrequent phrases (e.g., Rayner & Duffy, 1986); longer phrases are more difficult to process with longer and/or more fixations than shorter phrases (e.g., Rayner, Sereno, & Raney, 1996).

Last but not least, the duration of fixations can be different among distinct languages. In particular, the average fixation duration in Chinese is longer than in alphabetic languages (e.g., Li & Pollasek, 2020; Liversedge, Drieghe, Li, Yan, Bai, & Hyönä, 2016; Rayner, Li, Juhasz, & Yan, 2005). For example, Liversedge et al. (2016) found that on average, L1 Chinese users spent "245 milliseconds" on each fixation when reading Chinese text, while L1 English users spent "207 milliseconds" when reading comparable English text (p. 7).

To recap briefly, eye-tracking is a research method that tracks the fixations and saccades of a reader during reading. This method can tell us how long and how hard it is to process a specific phrase and its final word at early versus late processing-stages, by observing different measures and distinct target regions.

4.3.2 Why I chose eye-tracking

Eye-tracking was chosen for the current investigation for several reasons. First, eye-tracking is regarded as the "gold standard" for reading research (Rayner, 2009, p. 1474). Unlike other real-time techniques, eye-tracking can "tell us more about what is going on in the mind than [eye-tracking] realistically can" (Conklin & Pellicer-Sánchez, 2016, p. 454).

Second, compared with other real-time techniques (such as priming, judgment task, self-paced reading; e.g., Roberts, 2012), eye-tracking can provide extremely rich data from a variety of measures, reflect the process of reading from both macroscopic and microscopic views by observing different sizes of interest areas, and tell apart the early and late processing stages of reading.

Third, compared with other research methods, eye-tracking enables researchers to investigate more nuanced differences between experimental conditions and depict the reading pattern in a more fine-grained way (e.g., Godfroid, Winke, & Conklin, 2020).

Fourth, unlike other real-time techniques (e.g., priming, judgment task, self-paced reading), eye-tracking can be used to study not only what is fixated but also what is re-fixated (i.e., regression, which is good for assessing comprehension, e.g., Metzner, von der Malsburg, Vasishth, & Rösler, 2017) and what is skipped.

Fifth, unlike other real-time techniques (e.g., priming, judgment task, self-paced reading), eye-tracking allows the participant to read as naturally as possible in a laboratory setting (e.g., Conklin et al., 2018; Duyck, Van Assche, Drieghe, & Hartsuiker, 2007; Metzner et al., 2017).

Finally, eye-tracking can be used to triangulate with other behavioural (e.g, self-paced reading) and neurolinguistic techniques (e.g., EEG) for more valid and complete data interpretation (e.g., Siyanova-Chanturia, 2013; Godfroid et al., 2020).

It is of note that due to the unavailability of an eye-tracker for the L2 learner participants (Study 2) and child participants (Study 3), I chose the self-paced reading methodology instead.

4.4 Summary

To sum up, self-paced reading and eye-tracking are research methods that have been widely used and highly recommended in (FL) reading research. Both methods are good options to study FL processing, in that they can provide a window into the online processing of certain words and phrases in sentence context, such that the observation of the reading process can be more specific and fine-grained than other real-time techniques. Apart from this, eye-tracking can help researchers tease apart the early and late stages of processing during reading.

Based on the advantages listed above, both of the research methods have been chosen for the current thesis, with eye-tracking adopted in Study 1 (see Chapter 5), and self-paced reading adopted in Study 2 (see Chapter 6) and Study 3 (see Chapter 7).

Chapter 5

Study 1: The Processing of Modified Chinese Formulaic Sequences in L1 Users: Evidence from Eye-tracking

5.1 Introduction

The literature review in Chapter 3 suggests several prominent gaps in the research into formulaic language (FL) processing. First, although modified FL is common in natural language (e.g., Renouf & Sinclair, 1991), only a handful of studies have looked at their processing (e.g., Molinaro et al., 2013; Vilkaitė, 2016a). Critically, according to Santelmann and Jusczyk (1998), modification degree, such as insertion length, is likely to affect language processing. However, no study to date has explored the influence of insertion length in FL processing. Second, although research into English FL processing is plentiful, it is still scarce with respect to other languages, for example, Chinese. Among the few studies involving Chinese, only the processing of idioms (e.g., Liu, Li, Shu, Zhang, & Chen, 2010; Yu et al., 2016; Zhang, Yang, Gu, & Ji, 2013; Zhou, Zhou, & Chen, 2004), and n-grams (e.g., Kong, Zhang, & Zhang, 2016; Yi, Lu, & Ma, 2017) have received some attention, with other types of FL, such as collocations, being largely disregarded.

To address the gaps mentioned above, Study 1 was carried out to explore the processing of modified Chinese FSs in sentence context, by employing the eye-tracking paradigm on L1 adults. Five questions were examined:

(1) Do L1 Chinese FSs have a processing advantage over novel phrases?
(2) If so, does the processing advantage extend to modified L1 Chinese FSs? If so, how does the degree of modification, operationalised as insertion length (i.e., number of inserted word/s) affect the processing advantage?
(3) Does the processing of L1 Chinese FSs vary at early versus late processing stages?
(4) Is the processing of L1 Chinese FS in the whole phrase similar to the processing of its final word?
(5) In addition, does the processing of L1 Chinese FSs differ between distinct kinds, such as Adjective+Noun collocations versus Verb+Noun collocations, relative to novel phrases?

5.2 Methodology

5.2.1 Materials

For the purpose of the current study and in line with existing research (e.g., Jiang et al., 2020; Siyanova et al., 2011b; Vilkaitė, 2016a), I adopt the frequency-based approach to define FSs. The Adjective+Noun (A+N) (e.g., *白衬衫* '*white shirt*') and Verb+Noun (V+N) Chinese collocations (e.g., *坐飞机* '*take the plane*') were chosen for the present investigation. These FSs were selected because (1) they are more frequent than other types of FL, such as idioms (e.g., Vilkaitė, 2016b). (2) Unlike idioms, collocations are transparent, compositional (i.e., the meaning of individual components contributes fully to the meaning of the phrasal configuration), and can be easily modified (e.g., Li & Guo, 2016; Vilkaitė, 2016a). (3) Collocations have been employed as stimuli in studies investigating alphabetic languages (e.g., modified Italian collocations in Molinaro et al. (2013); modified English Verb+Noun collocations in Vilkaitė (2016a)), which will allow for useful comparisons with Chinese. (4) It has been proposed that V+N collocations and A+N collocations may differ in how easily they are acquired and used, which may also affect their respective processing (e.g., Peters, 2016; Wolter & Yamashita, 2015). For example, A+N collocations are thought to be easier to learn than V+N collocations (e.g., Peters, 2016). Thus we may expect that A+N collocations could be processed faster and the magnitude of FL advantage or phrase frequency effect on A+N collocations could be larger than V+N collocations. (5) It has been found that the degree of dependency between components is different in A+N vs V+N collocations, which results in processing difference between the two structures (e.g., Wolter & Yamashita, 2015). The reasons mentioned above warrant the comparison of the processing of A+N vs V+N collocations.

The original (i.e., unmodified) Adjective+Noun (e.g., *白衬衫* '*white shirt*') and Verb+Noun collocations (e.g., *坐飞机* '*take the plane*'), as well as their controls (e.g., *黄衬衫* '*yellow shirt*'; *修飞机* '*fix the plane*'), were retrieved in pairs from the BCC corpus of Chinese[2] (BLCU Corpus Center, 2016) as per the following criteria. First, each pair included

[2] The BCC corpus is "an online data system with a size of about 15 billion Chinese characters, ideal as a data source for studies in linguistics as well as applied linguistics" (Xun, Rao, Xiao, & Zang, 2016, p.118). This corpus is regarded as the largest Chinese online corpus to date, and one of the "three most widely used Chinese general-purpose corpora" (Xu, 2015, p. 219). Entries are extracted from texts in newspapers, Weibo (spoken text), technology, literature, comprehensive discourses, and other text genres (Xun et al., 2016). This mega-

a high-frequency phrase as collocation (phrase frequency > 500) and a low-frequency phrase as its matched control (phrase frequency < 50). Second, all phrases were grammatical, transparent, and decomposable configurations, consisting of two high-frequency component words (frequency > 10,000). Third, Word 1 in each pair (e.g., in Adjective+Noun phrases: 白 '*white*' in 白衬衫 '*white shirt*', 黄 '*yellow*' in 黄衬衫 '*yellow shirt*'; in Verb+Noun phrases: 坐 '*take*' in 坐飞机 '*take the plane*', 修 '*fix*' in 修飞机 '*fix the plane*') was matched in frequency, length (1–3 characters) and part of speech, and matched for character structure, stroke number, and concreteness. Fourth, Word 2 (1–2 characters, e.g., in Adjective+Noun phrases: 衬衫 '*shirt*' in 白衬衫 '*white shirt*' and 黄衬衫 '*yellow shirt*'; in Verb+Noun phrases: 飞机 '*plane*' in 坐飞机 '*take the plane*' and 修飞机 '*fix the plane*') was always identical in the collocation and control conditions (e.g., Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). Finally and crucially, all phrases were modifiable and allowed for insertions. Following the above criteria, a pool of 104 pairs of phrases was selected.

To ensure any processing differences were not due to the (forward) semantic association between the two component-words in the collocation and control conditions (e.g., Conklin et al., 2018; Siyanova-Chanturia et al., 2011b; Vilkaitė, 2016a), a norming study in the form of a questionnaire was carried out. In this study, component Word 1 of the original phrase pairs were presented to 10 L1 users of Chinese[3] who did not participate in the reading experiment. Pilot participants provided the word that first came to mind when seeing a given word (e.g., Collocation condition: 白 '*white*' ______, 坐 '*sit*' _______; Control condition: 黄 '*yellow*' ______, 修 '*fix*' _______). The proportion of responses (out of 10) that were identical to Word 2 in the target phrases was calculated for each phrase and was taken as its association strength (e.g., none of the pilot participants provided the noun 衬衫 '*shirt*' upon seeing the adjective 白 '*white*' _____; thus, the association strength was taken as 0.00). Any phrase pair with a significant difference in the association strength (difference > 0.40) was excluded. This ended up with 48 pairs of phrases remaining. *T*-test showed the 48 pairs had no significant associative differences between the two constituent-words in the collocation

corpus not only allows researchers to extract frequencies for collocates, but also to search unique units such as insertable expressions (Li & Guo, 2016).

[3] They were recruited via emails from the population of Chinese students and academic staff at universities, either in New Zealand or Mainland China, in line with the ethical procedures of the Human Ethics Committee at Victoria University of Wellington (VUW, Ethics Approval: 0000026508).

and control conditions ($mean_{collocation}$ = 0.08, $mean_{control}$ = 0.06, t = 1.05, $p > .10$, effect size = 0.21).

The 48 pairs included 23 pairs of Adjective+Noun (A+N) phrases and 25 pairs of Verb+Noun (V+N) phrases. The characteristics of these two types of collocations and their respective controls are presented in Table 9. *T*-tests showed no significant differences between Word 1 frequencies in each phrase pair ($mean_{collocation}$ = 112392.17, $mean_{control}$ = 146187.29, t = -0.62, $p > .10$, effect size = 0.13).[4] Character structure and stroke number of Word 1 in each phrase pair were also matched ($\chi^2_{character\ structure}$ = 29.60, $p > .10$; $t_{character\ stroke\ number}$ = 1.10, $p > .10$, effect size = 0.12).[5] However, the collocations and controls differed greatly in phrase frequency as shown in BCC corpus ($median_{collocation}$ = 1564.50, $median_{control}$ = 17.50; $mean_{collocation}$ = 4000.46, $mean_{control}$ = 19.75, t = 2.96, $p < .01$, effect size = 0.60).

Table 9:

Summary of the Frequency Information and the Association Strength of the Original Phrases

Part of speech	**Phrase type**	**Example**	**Mean phrase frequency**	**Mean word 1 frequency**	**Mean word 2 frequency**	**Mean phrase length**	**Mean association strength**
Adjective+Noun	Collocation	白衬衫 (white shirt)	1317.65	122310.96	216234.83	4.48	0.10
Adjective+Noun	Control	黄衬衫 (yellow shirt)	21.00	164204.00	216234.83	4.48	0.07
Verb+Noun	Collocation	坐飞机 (take the plane)	6468.64	103266.88	178700.24	3.16	0.07
Verb+Noun	Control	修飞机 (fix the plane)	18.60	129611.92	178700.24	3.16	0.05

Notes. English translations are word by word. Phrase length is calculated by the number of Chinese characters. The maximum association strength is 1.00.

On top of the original version mentioned above, two modified versions were created for each pair of phrases. One was the short-insertion version, with two Chinese characters

[4] All *t*-tests in the materials section were done using the *stats* package (version 3.6.2) in R software (version 3.6.1, 2019). Effect sizes of the *t*-tests in this section were calculated using the *Cohen's d* estimate in the *effsize* package (version 0.7.6, 2019).

[5] This was calculated using the *Pearson's Chi-Squared* test (for count data) in the *stats* package in R.

inserted before the noun[6] (e.g., *白**色长**衬衫* '*white **long** shirt*', *黄**色长**衬衫* '*yellow **long** shirt*'; *坐**一次**飞机* '*take the plane **for once***', *修**一次**飞机* '*fix the plane **for once***'). Two characters were specified as the short insertion length, in that this is the most common length of one Chinese word (it is estimated that over 70% of Chinese words are made up of two characters, e.g., The Commercial Press, 2008; Zhou & Marslen-Wilson, 2000). The other modified condition was the long-insertion version, with four characters inserted in the same position (e.g., *白**格子长袖**衬衫* '*white-**plaid long-sleeve** shirt*', *黄**格子长袖**衬衫* '*yellow-**plaid long-sleeve** shirt*'; *坐**那架真的**飞机* '*take **that real** plane*', *修**那架真的**飞机* '*fix **that real** plane*'). This version was created for the following reasons: (1) to be comparable with the short-insertion condition; and (2) to ensure that the lengthened phrase (collocation or control) can still sit within the perceptual span[7] of readers. According to Yan, Zhou, Shu, and Kliegl (2015), on average, Chinese readers can only extract information from one character to the left of the current fixation, to three characters to the right of the fixation (see also Li & Pollatsek, 2020). The inserted words/characters always carried meaning but were not strong collocates for the first word (i.e., Word 1 in the original version) or the final word (i.e., Word 2 in the original version) of the phrase. To avoid a repetition effect, the inserted words/characters in the short-insertion version never appeared in the long-insertion version.

The six versions of each phrase (original, short-insertion, and long-insertion collocation conditions and their respective control conditions) were embedded in identical sentence contexts. Care was taken to keep the phrase positioned approximately in the middle of the sentence, at least two words from the beginning and one word from the end (to avoid the wrap-up effect, e.g., Rayner & Pollatsek, 1989). All words in the sentence contexts were highly frequent (word frequency > 10,000).

To ensure any processing differences were not due to the context predictability or sentence naturalness (e.g., Siyanova-Chanturia et al., 2017; Conklin et al., 2018) and in line with existing research (e.g., Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019), another two norming studies (in the form of a questionnaire) were carried out. Sixty L1 Chinese adults,[8] who did

[6] The inserted words in the current study are always in the middle of the phrase. The word order of the inserted words might be different in the English translation.

[7] Perceptual span reflects how much information can be extracted in single fixation during natural reading.

[8] These participants were also recruited via emails from the population of Chinese students and academic staff at universities, either in New Zealand or in Mainland China, in line with the ethical procedures of the Human Ethics Committee at VUW (Ethics Approval: 0000026508).

not take the association test mentioned earlier and did not participate in the eye-tracking experiment, were invited. Thirty of them took part in the context predictability study (10 people for each list), and the other 30 people participated in the sentence naturalness study (10 people for each list).

The context predictability study was analogous to a pseudo-cloze test. This test consisted of three lists, with 96 target sentences (and 48 filler sentences) in each list. Each list consisted of 48 pairs of phrases in their two phrase type conditions but with different insertion lengths (e.g., a collocation in its original condition and its relative control in the short-insertion (or long-insertion) condition appearing in one list) within partial sentence context. Each sentence was not displayed completely, with the context ending before Word 2 (in the original phrases) or the final word (in the modified phrases) of the phrase (e.g., Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). For example:

- List 1:
    - Original collocation: 这件白_______________;
      (This white __________________)
    - Control with short-insertion: 这件黄色长_____________;
      (This yellow long ______________)
- List 2:
    - Collocation with long-insertion: 这件白格子长袖_____________;
      (This white-plaid long-sleeve_______________)
    - Original control: 这件黄_______________;
      (This yellow________________)
- List 3:
    - Collocation with short-insertion: 这件白色长_____________;
      (This white long ___________________________)
    - Control with long-insertion: 这件黄格子长袖_____________.
      (This yellow-plaid long-sleeve________________)

Thirty participants were asked to complete the sentences in each list. The proportion of responses (out of 10) that were identical to the Word 2/final word of target items (e.g., *衬衫* '*shirt*' in *这件白<u>衬衫</u>* '*This white <u>shirt</u>*') was calculated for each phrase as its context predictability score (e.g., three people provided the noun *衬衫* '*shirt*' upon seeing the context

*这件白*________ ‘*This white*________’; thus the context predictability was taken as 0.30). Note that the calculation did not consider the context after the Word 2/final word (e.g., ‘*这件白衬衫**是他哥哥的***’ (*This white shirt **is his elder brother’s***) equals to ‘*这件白衬衫**不是我的***’ (*This white shirt **is not mine***)). *T*-test showed no significant differences in context predictability between collocations and controls, in their unmodified/original form ($mean_{collocation}$ = 0.15, $mean_{control}$ = 0.10, t = 1.11, p > .10, effect size = 0.23) and modified forms (short-insertion condition: $mean_{collocation}$ = 0.38, $mean_{control}$ = 0.27, t = 1.66, p = .10, effect size = 0.34; long-insertion condition: $mean_{collocation}$ = 0.51, $mean_{control}$ = 0.45, t = 0.86, p > .10, effect size = 0.18).

The sentence naturalness study also included three lists, with 96 target sentences (and 48 filler sentences) in each list. Each list consisted of the 48 pairs of phrases (in both phrase type conditions, but with different insertion lengths) within (full) sentence context. For example:

- List 1:
  - Original collocation: 这件白衬衫是他哥哥的;
    (This white shirt is his elder brother’s)
  - Control with short-insertion: 这件黄色长衬衫是他哥哥的;
    (This yellow long shirt is his elder brother’s)
- List 2:
  - Collocation with long-insertion: 这件白格子长袖衬衫是他哥哥的;
    (This white-plaid long-sleeve shirt is his elder brother’s)
  - Original control: 这件黄衬衫是他哥哥的;
    (This yellow shirt is his elder brother’s)
- List 3:
  - Collocation with short-insertion: 这件白色长衬衫是他哥哥的;
    (This white long shirt is his elder brother’s)
  - Control with long-insertion: 这件黄格子长袖衬衫是他哥哥的.
    (This yellow-plaid long-sleeve shirt is his elder brother’s)

Thirty participants were asked to rate the naturalness of each sentence on a 10-point Likert scale in each list (e.g., 请用数字 1-10 对下列句子的自然度进行评分 ‘Please rate the naturalness of the following sentences on a scale of 1–10, where 1 is unnatural and 10 is natural'. 例如 'For example', *他们还在图书馆里学习* (10, '*They are still studying in the library*'); *我们在这里说了一下午篮球* (7, '*We are talking about basketball here for the whole afternoon*'); *我跟姐姐抓了一些汉语* (1, '*My elder sister and I caught some Chinese language*')). *T*-tests showed no significant differences in naturalness between sentences embedded with collocations versus controls, in their unmodified/original ($mean_{collocation}$ = 8.85, $mean_{control}$ = 8.59, t = 1.60, p > .10, effect size = 0.33) and modified forms (short-insertion condition: $mean_{collocation}$ = 8.55, $mean_{control}$ = 8.31, t = 1.56, p > .10, effect size = 0.32; long-insertion condition: $mean_{collocation}$ = 8.58, $mean_{control}$ = 8.28, t = 1.58, p > .10, effect size = 0.32).

The characteristics of the target sentences are summarized in Table 10.

Table 10:

Summary of the Context Predictability and Naturalness of the Sentences Embedded with Phrases in Their Original and Modified Forms

Part of speech	Phrase type	Insertion length	Example	Mean phrasal length	Mean sentence length	Mean context predictability	Mean naturalness
Adjective+Noun	Collocation	Original	这件白衬衫是他哥哥的。 (This white shirt is his elder brother's.)	4.48	13.48	0.19	8.87
Adjective+Noun	Control	Original	这件黄衬衫是他哥哥的。 (This yellow shirt is his elder brother's.)	4.48	13.48	0.17	8.64
Adjective+Noun	Collocation	Short insertion	这件白色长衬衫是他哥哥的。 (This white long shirt is his elder brother's.)	6.48	15.48	0.38	8.43
Adjective+Noun	Control	Short insertion	这件黄色长衬衫是他哥哥的。 (This yellow long shirt is his elder brother's.)	6.48	15.48	0.35	8.23
Adjective+Noun	Collocation	Long insertion	这件白格子长袖衬衫是他哥哥的。 (This white-plaid long-sleeve shirt is his elder brother's.)	8.48	17.48	0.46	8.39
Adjective+Noun	Control	Long insertion	这件黄格子长袖衬衫是他哥哥的。 (This yellow-plaid long-sleeve shirt is his elder brother's.)	8.48	17.48	0.43	8.04
Verb+Noun	Collocation	Original	我要是能坐飞机就好了。 (I wish that I could take the plane.)	3.16	10.80	0.11	8.83
Verb+Noun	Control	Original	我要是能修飞机就好了。 (I wish that I could fix the plane.)	3.16	10.80	0.04	8.53

Part of speech	Phrase type	Insertion length	Example	Mean phrasal length	Mean sentence length	Mean context predictability	Mean naturalness
Verb+Noun	Collocation	Short insertion	我要是能坐一次飞机就好了。 (I wish that I could take the plane for once.)	5.16	12.80	0.38	8.66
Verb+Noun	Control	Short insertion	我要是能修一次飞机就好了。 (I wish that I could fix the plane for once.)	5.16	12.80	0.20	8.38
Verb+Noun	Collocation	Long insertion	我要是能坐那架真的飞机就好了。 (I wish that I could take that real plane.)	7.16	14.80	0.56	8.75
Verb+Noun	Control	Long insertion	我要是能修那架真的飞机就好了。 (I wish that I could fix that real plane.)	7.16	14.80	0.46	8.51

Notes. English translations are word by word. Sentence length is calculated by the number of Chinese characters. The maximum context predictability and naturalness rating are 1.00 and 10.00, respectively.

5.2.2 Participants

Forty-five[9] L1 users of Mandarin (36 female adults, mean age = 26.73) with normal or corrected to normal vision, were recruited from Victoria University of Wellington (VUW). None of these participants participated in any of the norming studies. These participants were full-time students at VUW taking a great variety of subjects. All were born and raised in Mainland China and came to New Zealand to study after age 18. All of the participants had been living in New Zealand for less than half a year.

Each participant received a NZ$15.00 supermarket voucher for their participation (funded by the Faculty of Humanities and Social Sciences Research Grant, No. 220797). All participants were informed of their rights and then gave their written consent for participation. This study was carried out in line with the ethical procedures of the Human Ethics Committee at VUW (Ethics Approval: 0000026508).

5.2.3 Apparatus and procedures

The material presentation and eye-movement recording were done using Eyelink 1000 plus (sampling rate: 1000 Hz, SR Research Ltd., Ontario, Canada). The materials were displayed on the screen of a desktop Dell computer (also known as the presentation computer, which is a 21-inch CRT monitor, resolution: 1024 × 768 pixels, refresh rate: 150 Hz), which was

[9] According to Conklin et al. (2018), the sample size of a quantitative study is normally determined by the number of lists and experimental conditions; 10 participants for each list/condition is the minimum. As three lists were included in the current study (which required at least 30 participants), 45 participants were considered a reasonable sample size.

connected to a desk-mount eye tracker. Another desktop Dell computer (i.e., the recording computer) was used to record eye movements registered by the desk-mount eye tracker.

The target sentences consisted of 48 pairs of phrases (23 pairs of Adjective+Noun phrases and 25 pairs of Verb+Noun phrases) embedded in sentence context. Each phrase pair included two phrase types (collocation and control) in three modified versions – the original version, the short-insertion version, and the long-insertion version. The 288 target sentences (48 phrase pairs × 2 phrase types × 3 modified versions) were presented in a pseudorandom order and allocated to three lists. Each list contained 96 target sentences and 48 filler sentences. The target sentences in each list included 48 phrase pairs (23 Adjective+Noun phrase pairs and 25 of Verb+Noun phrase pairs). Each pair of phrases in each list included both phrase types (collocation condition and control condition) but in different insertion length conditions (i.e., modified versions). The filler sentences (e.g., *她和姐姐去商场买书包* '*She went to the shopping mall with her elder sister to buy a schoolbag*') were made up of Chinese characters and words that are highly frequent (frequency > 10,000 as shown in BCC corpus). These filler sentences were equal in length to the target sentences.

Each sentence was presented across a single line in the middle of the screen. The Chinese characters in each sentence were shown in black (RGB: 0, 0, 0) bold *Song* font (one of "the most popular Chinese fonts", Zhang, 2016, p. 425), size 25 (to situate the longest sentence, which included 26 characters in one line in the middle of the screen) on a white (RGB: 255, 255, 255) background.

The reading experiment was carried out individually in the Eye-tracking lab at the School of Psychology, VUW. Each participant was first given written instructions and a verbal explanation of the procedure. His/her background information (i.e., age, gender, major of study) was collected and the consent form was signed before the experiment. After that, the participant was invited to the lab and seated approximately 60 cm from the eye-tracker (e.g., Liu, Li, & Han, 2015; Ma & Li, 2015; Shen, Li, & Pollatsek, 2018; Yi, Lu, & Ma, 2017). Their head was supported by a chin and forehead rest to minimize head movements during the experiment. The height of the chin rest and/or the chair were adjusted when needed. Before the experiment, a nine-point grid calibration procedure was conducted, in which the participant fixated on each of nine locations to establish an exact correspondence between pupil position in the eye-tracker (camera) image and gaze position on the display screen. A validation was run immediately after the calibration procedure, in which the participant again looked at these nine locations to ensure the accuracy of the system in predicting gaze position from pupil position (SR Research Ltd., 2017). The maximal error of the validation was 0.5 degrees in visual angle (e.g., Gu & Li, 2015; Liu et al., 2015; Ma, Li, & Rayner, 2015; Ma, Pollatsek, Li, & Li, 2016; Shen et al., 2018; Yi et al., 2017; Zhou, Ma, Li, & Taft, 2017). Although participants read sentences binocularly, only the right eye was recorded (e.g., Conklin et al., 2018; Gu & Li, 2015; Ma, Li, Xu, & Li, 2018; Rayner, 2009; Shen et al., 2018; Wei, Li, & Pollatsek, 2013; Yi et al., 2017; Zhou et al., 2017).

Each participant first completed a practice session with six trials and then the recording session with 144 trials (96 target sentences + 48 filler sentences). Each trial started with a drift check (a character-sized fixation point needed to be successfully gazed at the location of the first character of the upcoming sentence, e.g., Conklin & Pellicer-Sánchez, 2016; Conklin et al., 2018; Li & Logan, 2008) and followed by a sentence. The participant was required to read the sentences silently for comprehension at his/her own pace. Once the participant had finished reading the sentence, he/she pressed the spacebar to proceed. To check the participant's comprehension, 48 sentences (32 target sentences + 16 fillers, i.e., one-third of the trial sentences) were followed by a comprehension question (e.g., Carrol & Conklin, 2020). The participant responded to the questions (e.g., *这件衬衫是他哥哥的吗?* (*Is this shirt his brother's?*)) by pressing 'Left Alt' for *NO* and 'Right Alt' for *YES* on the keyboard. The 144 sentences in each list were divided into three blocks (with 48 sentences in each block), and the participant could choose to take a short break after finishing each block. Another set of calibration and validation procedure was done when the experiment resumed at the beginning of the block (e.g., Conklin et al., 2018). The experiment took about 15 minutes from start to finish. All the 45 participants successfully completed the experiment.

5.3 Analysis and results

5.3.1 Data analysis

Prior to the analyses, a Four-Stage Fixation Cleaning procedure using the Eyelink Data Viewer software (version 3.1.246) was performed to exclude fixations shorter than 80ms or longer than 800ms (e.g., Carrol & Conklin, 2020; Chen, Li, & Yang, 2012; Conklin et al., 2018; Ma, Pollatsek, Li, & Li, 2016). This accounted for 6.84% of the sentences (original version: 0.50% from the Adjective+Noun collocation condition, 0.55% from the Adjective+Noun control condition, 0.46% from the Verb+Noun collocation condition, 0.55% from the Verb+Noun control condition; short-insertion version: 0.57% from the Adjective+Noun collocation condition, 0.65% from the Adjective+Noun control condition, 0.55% from the Verb+Noun collocation condition, 0.53% from the Verb+Noun control condition; and long-insertion version: 0.68% from the Adjective+Noun collocation condition, 0.63% from the Adjective+Noun control condition, 0.64% from the Verb+Noun collocation condition, 0.54% from the Verb+Noun control condition). Participants showed no difficulty in comprehending these sentences (the mean percentage of correct answers was 95.69%).

As suggested by researchers (e.g., Carrol & Conklin, 2014b; Conklin & Pellicer-Sánchez, 2016) and previous FL processing studies (e.g., Jiang et al., 2020; Underwood et al., 2004; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019), two interest areas – the whole phrase and its final word – were observed.

In line with existing literature, three eye-tracking measures were selected for data analysis (e.g., Conklin & Pellicer-Sánchez, 2016; Rayner, 1998; Roberts & Siyanova-Chanturia, 2013; Siyanova-Chanturia, 2013; Siyanova-Chanturia et al., 2011a, 2011b; Underwood et al., 2004; Vilkaitė, 2016a; Yi et al., 2017). These measures are First Pass Reading Time (FPRT, early measure), Dwell Time (DT, late measure), and Fixation Count (FC, late measure).

An example of a hypothetical eye-movement record (examined in the three measures) for the six conditions of a sentence is illustrated in Figure 1.

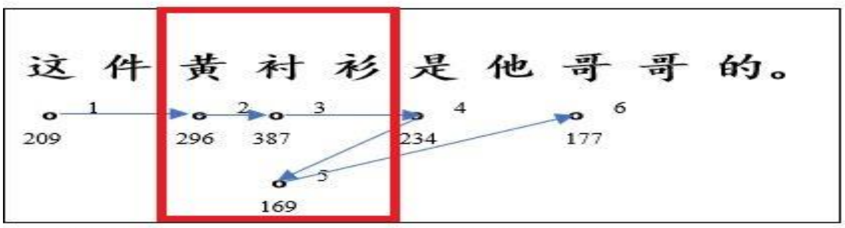
Figure 1:

An Example of a Hypothetical Eye-movement Record for the Six Conditions of a Sentence, with the Red Box Representing the Area of Interest in the Whole Phrase

Non-insertion collocation:

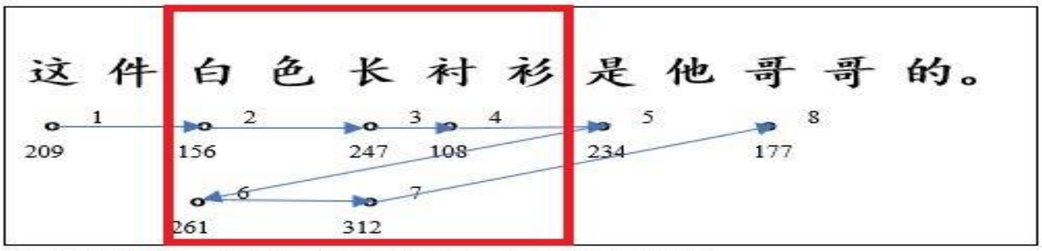
(1a) FPRT = 2 + 3 (duration of fixation)
(1b) DT = 2 + 3 (duration of fixation)
(1c) FC = 2 + 3 (number of fixation)

Non-insertion control:



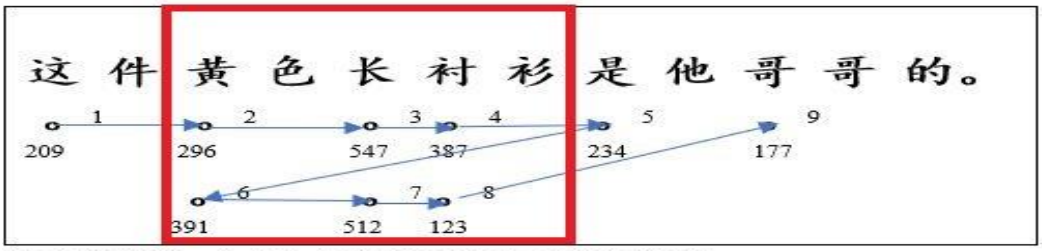
(2a) FPRT = 2 + 3 (duration of fixation)
(2b) DT = 2 + 3 + 5 (duration of fixation)
(2c) FC = 2 + 3 + 5 (number of fixation)

Short-insertion collocation:



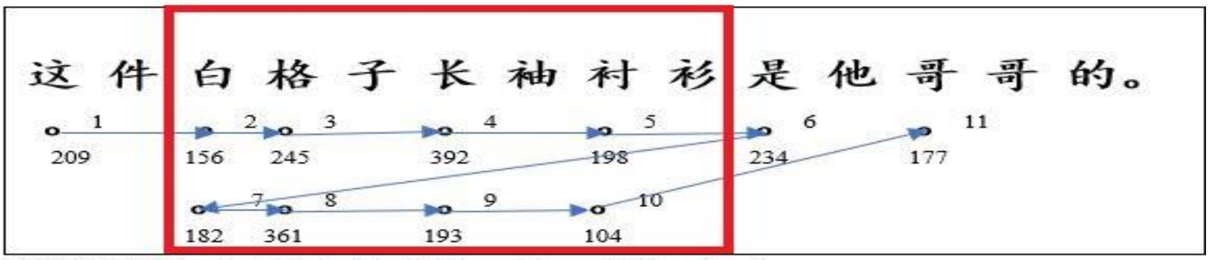
(3a) FPRT = 2 + 3 + 4 (duration of fixation)
(3b) DT = 2 + 3 + 4 + 6 + 7 (duration of fixation)
(3c) FC = 2 + 3 + 4 + 6 + 7 (number of fixation)

Short-insertion control:



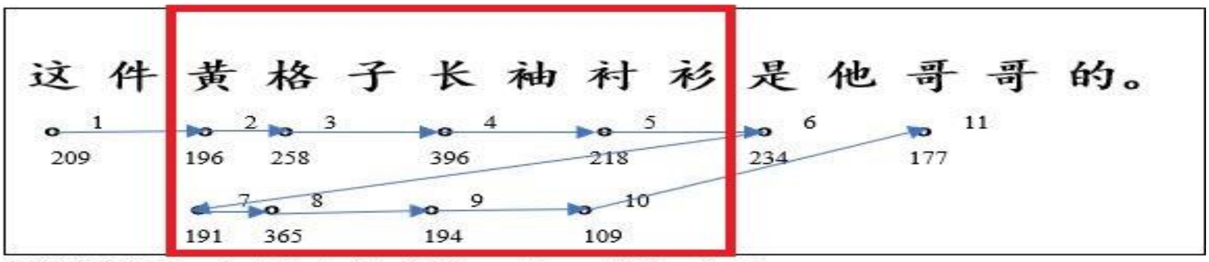
(4a) FPRT = 2 + 3 + 4 (duration of fixation)
(4b) DT = 2 + 3 + 4 + 6 + 7 + 8 (duration of fixation)
(4c) FC = 2 + 3 + 4 + 6 + 7 + 8 (number of fixation)

Long-insertion collocation:



(5a) FPRT = 2 + 3 + 4 + 5 (duration of fixation)
(5b) DT = 2 + 3 + 4 + 5 + 7 + 8 + 9 + 10 (duration of fixation)
(5c) FC = 2 + 3 + 4 + 5 + 7 + 8 + 9 + 10 (number of fixation)

Long-insertion control:



(6a) FPRT = 2 + 3 + 4 + 5 (duration of fixation)
(6b) DT = 2 + 3 + 4 + 5 + 7 + 8 + 9 + 10 (duration of fixation)
(6c) FC = 2 + 3 + 4 + 5 + 7 + 8 + 9 + 10 (number of fixation)

Note. The blue arrows represent the hypothetical eye movements of the participant when reading the sentence.

For the phrase-final word analysis, trials in which the word was skipped during first-pass reading (i.e., first fixation = 0ms) were excluded from the analyses of the FPRT and DT measures. This is due to the consideration that in single word analysis, including any skipped items in the calculation of means for duration measures or in any further analysis would affect the accuracy of the results (e.g., Conklin et al., 2018; see also Vilkaitė, 2016a). This exclusion resulted in 8.37% of data loss (original version: 0.59% from the Adjective+Noun collocation condition, 0.51% from the Adjective+Noun control condition, 0.93% from the Verb+Noun collocation condition, 0.49% from the Verb+Noun control condition; short-insertion version: 0.69% from the Adjective+Noun collocation condition, 0.42% from the Adjective+Noun control condition, 0.90% from the Verb+Noun collocation condition, 0.66% from the Verb+Noun control condition; and long-insertion version: 0.87% from the Adjective+Noun collocation condition, 0.57% from the Adjective+Noun control condition, 0.96% from the Verb+Noun collocation condition, 0.78% from the Verb+Noun control condition).

Descriptive statistics for the 12 experimental conditions (two phrase types × three insertion lengths × two parts of speech) across the three eye-tracking measures in the two interest areas are presented in Table 11.

Table 11:

Mean Fixation Durations (in Milliseconds) and Mean Fixation Counts in the Whole Phrase and its Final Word Analyses of the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition with Different Insertion Lengths, with Standard Deviation (SD) in Parentheses

	Original						Short Insertion						Long Insertion					
	A+N Phrase			*V+N Phrase*			*A+N Phrase*			*V+N Phrase*			*A+N Phrase*			*V+N Phrase*		
	COL	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***
Phrase-final word																		
FPRT	235.72 (110.44)	242.20 (118.04)	.25	254.07 (129.92)	272.64 (131.45)	.05	249.15 (132.70)	253.59 (143.13)	.35	240.80 (126.23)	257.13 (128.97)	.07	225.71 (101.07)	240.53 (121.24)	.07	232.88 (109.02)	233.99 (110.17)	.45
DT	331.15 (206.70)	389.93 (295.17)	< .01	300.80 (163.79)	377.36 (227.05)	< .001	350.92 (249.38)	385.32 (280.27)	.07	301.73 (206.96)	359.81 (223.42)	< .001	295.39 (187.51)	326.20 (247.81)	.06	284.14 (162.22)	302.58 (185.69)	.11
FC	1.14 (0.92)	1.40 (1.19)	< .001	0.89 (0.76)	1.34 (0.96)	< .001	1.14 (1.02)	1.43 (1.21)	< .001	0.95 (0.91)	1.24 (1.06)	< .001	0.91 (0.86)	1.17 (1.03)	< .001	0.86 (0.77)	1.04 (0.87)	< .01
Whole phrase																		
FPRT	359.64 (225.78)	396.55 (247.36)	.02	341.57 (192.11)	399.87 (229.42)	< .001	502.43 (335.43)	572.57 (371.65)	< .01	515.53 (323.79)	561.88 (342.11)	.03	636.55 (410.70)	704.70 (433.59)	.02	676.94 (433.72)	720.48 (429.32)	.08
DT	626.45 (449.32)	736.10 (469.95)	< .001	459.10 (256.02)	613.50 (395.24)	< .001	907.32 (593.12)	997.26 (652.00)	.03	763.11 (472.29)	903.47 (537.58)	< .001	1117.74 (751.94)	1228.62 (862.95)	.04	1045.07 (661.75)	1114.97 (660.56)	.08
FC	2.80 (1.82)	3.27 (1.92)	< .001	1.99 (1.03)	2.57 (1.47)	< .001	4.12 (2.40)	4.55 (2.87)	.02	3.36 (1.90)	3.98 (2.23)	< .001	5.22 (3.20)	5.61 (3.73)	.07	4.65 (2.70)	4.96 (2.65)	.05

Notes. FPRT = First Pass Reading Time, DT = Dwell Time, FC = Fixation Count. The mean fixation durations and fixation counts are the original values (not log-transformed). The estimated value of each measure was rounded and kept to two decimal places.

The eye movement data were analyzed using Linear Mixed-Effects Models in R (version 3.6.1, 2019), package *lme4* (version 1.1-21, 2019). The *p*-value for each predictor (variable) was estimated using the *lmerTest* package (version 3.1-1, 2019). Mixed-effects models for the two duration measures (i.e., FPRT and DT, which provide continuous data) were constructed using the *lmer* function in the *lme4* package. For the count measure (i.e., FC, which provides discrete data), a generalized linear model was established using the *glmer* function in the *lme4* package (e.g., Vilkaitė, 2016a), specifying the *Poisson* distribution (Conklin et al., 2018; see also Vilkaitė, 2016a).

All continuous dependent variables (i.e., FPRT and DT) were log-transformed before the model fitting. All categorical independent variables were coded. To this aim, *phrase type* was coded as '1' for the collocation condition and '-1' for the control condition. A similar coding procedure was conducted for *insertion length* ('0' for the original condition, '2' for the short-insertion condition, and '4' for the long-insertion condition) and *part of speech* ('1' for the Verb+Noun phrases, including collocations and controls; and '-1' for the Adjective+Noun phrases, including collocations and controls).

Six separate models were constructed (one model × three eye-tracking measures × two interest areas). Each of these original models included a maximal structure of fixed effects and random effects. To answer Research Question (1), phrase type was first added to the model as a main predictor. To answer Research Questions (2) (3) (4), insertion length and its interaction with phrase type were added to the model. To answer Research Question (5), the two-way interaction between part of speech and phrase type and the three-way interaction involving part of speech, phrase type, and insertion length were added to the model. All of these fixed effects, including main effects and interactions, were kept in the original model as predictors for the eye-tracking measure, regardless of their significance values. A maximal structure of random effects was also retained in the model (e.g., Barr, Levy, Scheepers, & Tily, 2013), with random intercepts for the participants and items respectively, as well as random slopes for the main predictors (variables) involved.

A backward model-fitting procedure was carried out as per Bates, Kliegl, Vasishth, and Baayen (2015). To obtain the final converged and fitted model, predictors in the random effects structure were incrementally removed from the maximal model, in which the removal of a given predictor was no longer justified (e.g., Siyanova-Chanturia & Janssen, 2018). The justification for the removal was determined by model comparisons using the *chi-square* test in the *anova* function (in the *lmerTest* package, see Baayen, Davidson, & Bates, 2008; see also Siyanova-Chanturia & Janssen, 2018). The *ranova* function (also in the *lmerTest* package) was also checked for the random effect removal and model reduction (note this function is available for models of duration measures, but unavailable for the model of count measure). *Kappa* value was checked for the multicollinearity issue in each model (*Kappa* < 3.51) and *vif* value was referred for the collinearity issue for each fixed effect (*vif* < 2.50).

Further comparisons/contrasts (including *post-hoc* analysis) between the conditions within categorical variables and the interactions involved (including calculating effect sizes for each comparison/contrast) were performed using the *emmeans* package (version 1.4.3.01, 2019). Interactions were plotted by applying the *ggplot2* package (version 3.2.1, 2019).

5.3.2 Results

5.3.2.1 The phrase-final word analysis

Fixed effects and random effects of the selected models for the three eye-tracking measures – First Pass Reading Time (FPRT), Dwell Time (DT), and Fixation Count (FC) – are presented in Table 12. *Phrase type* was not a significant predictor in the early measure ($t_{FPRT} = -1.66$, $p = .10$), but was significant in the late measures ($t_{DT} = -4.19$, $p < .001$; $z_{FC} = -4.92$, $p < .001$). Further comparisons revealed that the final word in collocations was processed faster and required fewer fixations than in controls ($t_{DT} = 4.27$, $p < .001$, effect size = 0.23; $z_{FC} = 5.84$, $p < .001$, effect size = 0.26).

Table 12:

Summary of Selected Models in the Phrase-Final Word Analysis

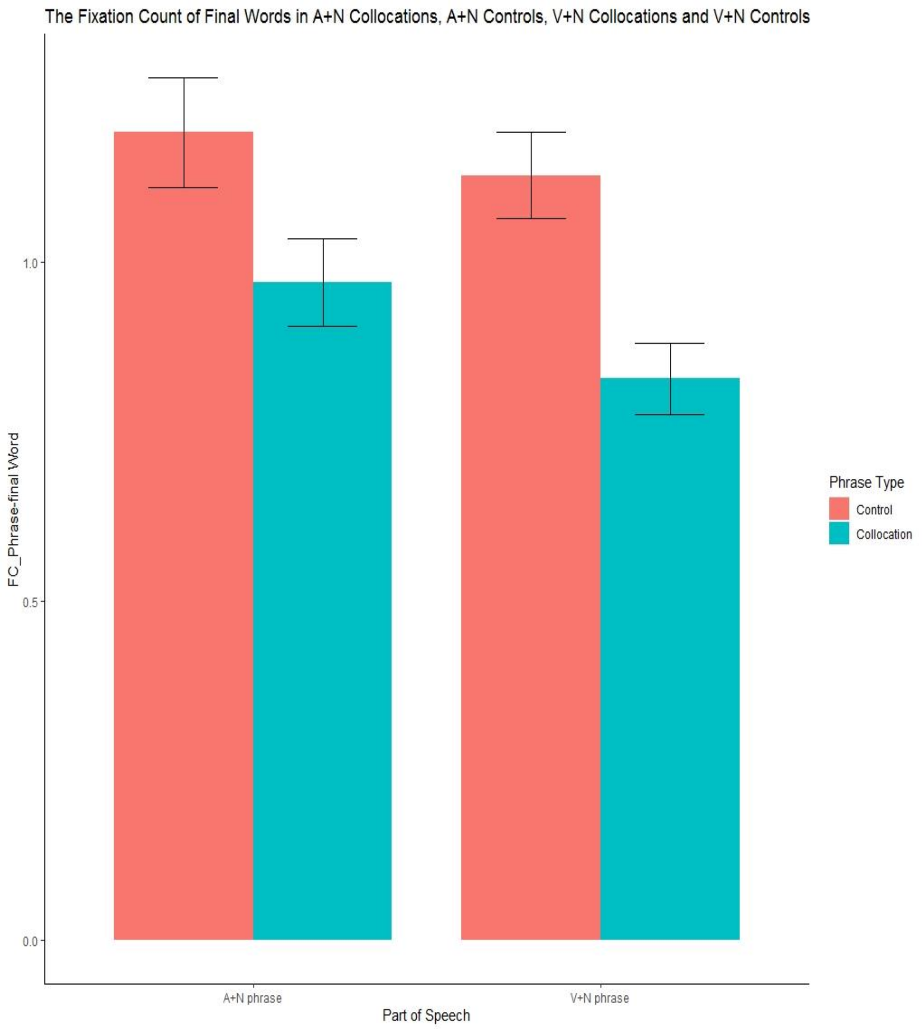
	First Pass Reading Time					**Dwell Time**					**Fixation Count**			
Fixed effects	*Estimate*	*SE*	*df*	*t*	*pr*	*Estimate*	*SE*	*df*	*t*	*pr*	*Estimate*	*SE*	*z*	*pr*
Intercept	5.43	0.02	66.16	228.29	< .001	5.68	0.04	69.38	148.87	< .001	0.11	0.05	2.01	.04
Phrase type	-0.02	0.01	732.90	-1.66	.10	-0.08	0.02	279.90	-4.19	< .001	-0.15	0.03	-4.92	< .001
Insertion length	-0.02	0.00	1136.00	-3.28	< .01	-0.03	0.01	972.10	-4.88	< .001	-0.04	0.01	-3.16	< .01
Phrase type × Insertion length	0.00	0.00	1059.00	0.16	.87	0.01	0.01	1020.00	1.58	.12	0.01	0.01	0.77	.44
Phrase type × Part of speech	-0.02	0.01	807.00	-1.29	.20	-0.02	0.02	743.60	-1.17	.24	-0.06	0.03	-1.99	.05
Phrase type × Insertion length × Part of speech	0.01	0.00	725.10	1.20	.23	0.01	0.01	806.40	1.04	.30	0.02	0.01	1.36	.17
Random effects	*Variance*	*SD*				*Variance*	*SD*				*Variance*	*SD*		
Item intercept	0.00	0.05				0.01	0.12				0.01	0.11		
Phrase type | Item	-	-				0.01	0.09				0.01	0.10		
Part of speech | Item	-	-				-	-				0.03	0.17		
Insertion length | Item	-	-				-	-				0.01	0.08		
Participant intercept	0.02	0.13				0.05	0.22				0.08	0.28		
Part of speech | Participant	-	-				-	-				0.00	0.04		
Insertion length | Participant	-	-				-	-				0.00	0.01		
Residual	0.17	0.41				0.25	0.50				-	-		

Notes. The estimated value of each predictor was rounded and kept to two decimal places. Some random effects were discarded due to model fitting, and model comparisons showed no significant differences between the fitted model and the previous one(s).

It is noteworthy that *phrase type* significantly interacted with *part of speech* in the (late) count measure (z_{FC} = -1.99, *p* = .05), but not in the duration measures (t_{FPRT} = -1.29, *p* >.10; t_{DT} = -1.17, *p* >.10). This interaction is plotted in Figure 2. Further comparisons suggested that the processing of final words in collocations required fewer fixations than in control phrases, and the magnitude of the processing facilitation for collocations was larger in Verb+Noun collocations (z_{FC_V+N} = 5.24, *p* < .001, effect size = 0.30) than in Adjective+Noun collocations (z_{FC_A+N} = 3.33, *p* < .001, effect size = 0.22).

Figure 2:

The Interaction between Phrase Type and Part of Speech for the Fixation Count in the Phrase-Final Word Analysis, with 95% Confidence Intervals for Each Condition



No significant interactions were found between *phrase type* and *insertion length* (t_{FPRT} = 0.16, *p* >.10; t_{DT} = 1.58, *p* >.10; z_{FC} = 0.77, *p* >.10), or between *phrase type, insertion length*, and *part of speech* across the three eye-tracking measures (t_{FPRT} = 1.20, *p* >.10; t_{DT} = 1.04, *p* >.10; z_{FC} = 1.36, *p* >.10).

5.3.2.2 The whole-phrase analysis

The reading of the whole phrase was further analysed. Fixed effects and random effects of the selected models for the three eye-tracking measures (FPRT, DT, and FC) are presented in Table 13.

Phrase type was a significant predictor across all eye-tracking measures (t_{FPRT} = -2.69, $p < .01$; t_{DT} = -5.31, $p < .001$; z_{FC} = -4.86, $p < .001$). Further comparisons revealed that collocations were processed faster and required fewer fixations than controls (t_{FPRT} = 3.47, $p < .001$, effect size = 0.19; t_{DT} = 5.76, $p < .001$, effect size = 0.34; z_{FC} = 5.16, $p < .001$, effect size = 0.47).

Table 13:

Summary of Selected Models in the Whole Phrase Analysis

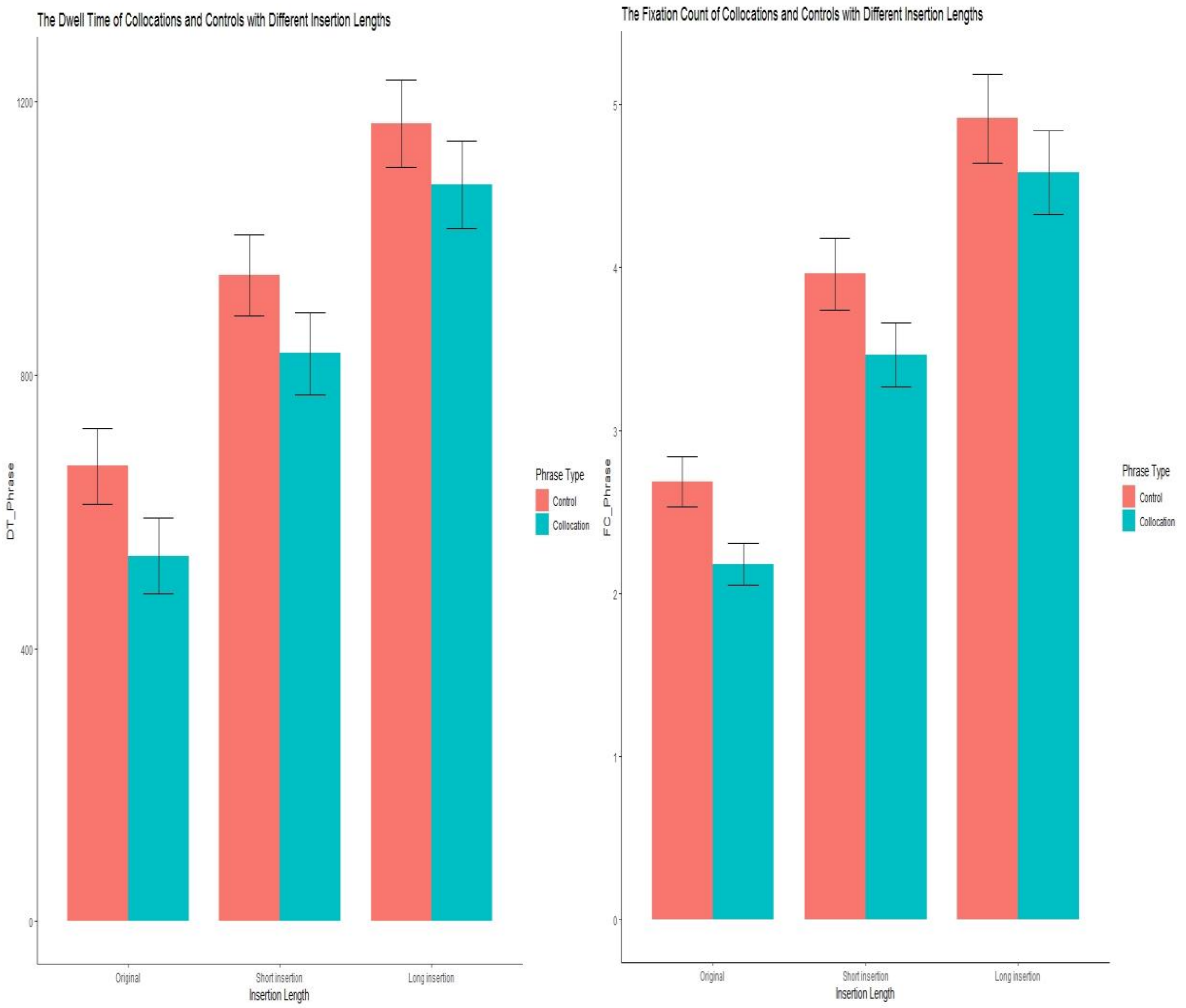
	First Pass Reading Time					Dwell Time					Fixation Count			
Fixed effects	*Estimate*	*SE*	*df*	*t*	*pr*	*Estimate*	*SE*	*df*	*t*	*pr*	*Estimate*	*SE*	*z*	*pr*
Intercept	5.77	0.04	60.13	143.82	< .001	6.23	0.06	54.96	103.46	< .001	0.92	0.05	17.52	< .001
Phrase type	-0.06	0.02	125.49	-2.69	< .01	-0.11	0.02	134.27	-5.31	< .001	-0.10	0.02	-4.86	< .001
Insertion length	0.14	0.01	148.84	14.40	< .001	0.16	0.01	278.18	19.77	< .001	0.17	0.01	21.62	< .001
Phrase type × Insertion length	0.00	0.01	279.64	0.20	.84	0.02	0.01	278.17	2.09	.04	0.02	0.01	2.17	.03
Phrase type × Part of speech	-0.01	0.02	125.43	-0.48	.63	-0.02	0.02	134.17	-1.08	.28	-0.03	0.02	-1.37	.17
Phrase type × Insertion length × Part of speech	0.01	0.01	280.00	0.55	.58	0.01	0.01	278.19	0.70	.48	0.01	0.01	0.89	.37
Random effects	*Variance*	*SD*				*Variance*	*SD*				*Variance*	*SD*		
Item intercept	0.03	0.17				0.04	0.19				0.00	0.00		
Phrase type | Item	-	-				-	-				-	-		
Part of speech | Item	-	-				-	-				0.03	0.16		
Insertion length | Item	0.00	0.02				0.00	0.00				-	-		
Participant intercept	0.05	0.23				0.14	0.38				0.11	0.32		
Part of speech | Participant	-	-				0.00	0.04				0.00	0.04		
Insertion length | Participant	0.00	0.03				-	-				-	-		
Residual	0.31	0.56				0.20	0.45				-	-		

Notes. The estimated value of each predictor was rounded and kept to two decimal places. Some random effects were discarded due to model fitting, and model comparisons showed no significant differences between the fitted model and the original model(s).

Crucially, a significant interaction between *phrase type* and *insertion length* was observed in the late, but not early, measures (t_{FPRT} = 0.20, $p > .10$; t_{DT} = 2.09, p = .04; z_{FC} = 2.17, p = .03). This interaction is plotted in Figure 3. Further comparisons showed that the collocations were processed significantly faster and required fewer fixations than controls across all insertion length conditions; and, more crucially, the processing advantage for collocations diminished with longer insertions (original condition: t_{DT} = 5.11, $p < .001$, effect size = 0.50, z_{FC} = 4.16, $p < .001$, effect size = 0.49; short-insertion condition: t_{DT} = 3.05, $p < .01$, effect size = 0.32, z_{FC} = 3.07, $p < .01$, effect size = 0.48; long-insertion condition: t_{DT} = 2.04, p = .04, effect size = 0.20, z_{FC} = 1.75, p = .08, effect size = 0.33).

Figure 3:

The Interaction between Phrase Type and Insertion Length for the Dwell Time and Fixation Count in the Whole Phrase Analysis, with 95% Confidence Intervals for Each Condition



No significant interactions were found between *phrase type* and *part of speech* (t_{FPRT} = -0.48, p > .10; t_{DT} = -1.08, p > .10; z_{FC} = -1.37, p > .10), or between *phrase type*, *insertion length*, and *part of speech* across all eye-tracking measures (t_{FPRT} = 0.55, p > .10; t_{DT} = 0.70, p > .10; z_{FC} = 0.89, p > .10).

5.4 Discussion

The processing advantage for L1 FSs, figurative or literal, over infrequent novel phrases has long been affirmed in the literature (e.g., Arnon & Snider, 2010; Carrol & Conklin, 2020; Ellis et al., 2009; Janssen & Barber, 2012; Jolsvai et al., 2013; Millar, 2011; Molinaro et al., 2008; Siyanova-Chanturia et al., 2017; Siyanova-Chanturia & Lin, 2018; Swinney & Cutler, 1979; Tremblay et al., 2011; Vespignani et al., 2009). However, the relevant evidence pertains almost exclusively to L1 FSs in their canonical form in English and other alphabetic languages. Whether or not modified FSs in non-alphabetic languages, such as Chinese, enjoy a processing advantage over novel phrases has to date been largely disregarded. In the present investigation, I sought to address this gap and the research questions brought up at the beginning of the current chapter in particular, by asking L1 users of Chinese (Mandarin) to read phrases varying in frequency, embedded in sentence context. I recorded the eye movements of these readers and, in particular, I examined how *phrase type* (collocation vs control), *degree of modification* (i.e., insertion length, with original vs short-insertion vs long-insertion), and *part of speech* (Adjective+Noun phrase vs Verb+Noun phrase) might have affected the early- and late-stage processing of the whole phrase and its final word.

Regarding Research Question (1): Do L1 Chinese FSs have a processing advantage over novel phrases? The answer is yes. That is, L1 Chinese FSs had a processing advantage over novel phrases. I found a significant main effect of *phrase type*, with Chinese collocations read faster than control phrases. This effect was observed in the early and, late measures in particular, in both analyses of the whole phrase and its final word. This finding is reminiscent of several earlier studies (e.g., Jiang et al., 2020; Siyanova-Chanturia et al., 2011b; Sonbul, 2014; Underwood et al., 2004; Yi et al., 2017). For example, in an eye-tracking study, Yi et al. (2017) asked L1 Chinese (Mandarin) users to read sentences embedded with frequent adverbial sequences (e.g., *仍不* ‘*still did not*’) and infrequent control phrases (e.g., *仍没* ‘*still had not*’). Yi et al. (2017) found adverbial sequences were read significantly faster with fewer fixations than control phrases across the early and late eye-tracking measures. The finding in the present investigation is in line with earlier research, such as Yi et al. (2017).

Regarding Research Question (2): Does the processing advantage of Chinese FSs extend to their modified form? If so, how does the degree of modification, operationalised as insertion length (i.e., number of inserted word/s) affect the processing advantage? The answer is also yes. That is, the processing advantage for L1 Chinese FSs in their canonical form extended to their modified form; but, with the increase of modification degree, the magnitude of the processing advantage diminished. *Modification degree* was found to significantly interact with *phrase type* in the analysis of the whole phrase for the late measures (Dwell Time and Fixation Count), such that collocations were read faster with fewer fixations than relative controls in their original ($p < .001$), short-insertion ($p < .01$), and long-insertion forms ($p = .04$). This finding is evocative of several existing studies (e.g., Haeuser et al., 2020; Molinaro et al., 2013; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). For example, in an eye-tracking study, Vilkaitė (2016a) asked L1 English speakers to read sentences. These sentences were embedded with Verb+Noun collocations and control phrases, in their original (e.g., *provide information* vs *compare information*) and modified forms (with three words inserted in the middle of the phrase; e.g., *provide some of the information* vs *compare some of the information*). However, Vilkaitė (2016a) did not find a significant interaction between modification and phrase type in the whole phrase in any of the eye-tracking measures. This finding was taken to show that the magnitude of processing advantage for the collocations over relative controls kept consistent across the original and modified forms. The finding of the present study, however, points to a different picture. Unlike the finding in Vilkaitė (2016a), the interaction between modification degree and phrase type was observed in the whole phrase for the late measures. More interestingly, the magnitude of the processing advantage for collocations over respective controls was not consistent across the original and modified forms; instead, with the increase of modification degree, the processing advantage for collocations diminished.

Regarding Research Question (3): Does the processing of L1 Chinese FSs vary at early versus late processing stages? The answer is also yes. That is, L1 Chinese FSs were processed differently at the early and late processing stages and this fact remains not only in FSs in their original form but also in their modified forms. I found that the processing advantage for the original and modified collocations over their respective controls in the late, but not early, eye-tracking measures. This finding reminds us of some existing studies (e.g., Jiang et al., 2020; Siyanova-Chanturia et al., 2011a). For example, in a recent eye-tracking study, Jiang et al. (2020) asked L1 Chinese users to read sentences embedded with Verb+Noun collocations and control phrases, and examined the early (Fixation Duration), middle (First Pass Reading Time, which can be also regarded as an early measure, see e.g., Conklin et al., 2018; Pellicer-Sánchez & Siyanova-Chanturia, 2018; Rayner, 1998), and late measures (Dwell Time and Fixation Count) during reading. Jiang et al. (2020) found the processing advantage for collocations and higher frequency phrases in the late, but never in the early or middle, measures. The present finding is in line with earlier research, such as Jiang et al. (2020).

Regarding Research Question (4): Is the processing of L1 Chinese FS in the whole phrase similar to the processing of its final word? The answer is no. That is, the processing of L1 Chinese FSs was not similar but varied in the whole phrase and its final word. For example, I found a significant interaction between phrase type and modification degree in the whole phrase, but not in its final word in any of the eye-tracking measures. This finding suggests that the magnitude of FL advantage is kept consistent in the final word across the original and modified forms. With regard to the whole phrase, however, the magnitude of FL advantage was not consistent but diminished with the increase of modification degree. In another eye-tracking study (Vilkaitė, 2016a), the interaction between phrase type and modification was not found in the whole phrase but in the phrase-final word. This finding suggests that the magnitude of FL advantage is kept consistent in the whole phrase across the original and modified forms. With regard to its final word, however, the magnitude of FL advantage was not consistent but diminished in the modified form. Although the present study found the magnitude of FL advantage consistent across the original and modified forms in the phrase-final word, this finding is not surprising. It suggests that, although a FS was stretched and knocked apart, which might have affected its formulaicity and the strength of the observed phrase frequency effect, the predictability and facilitation of its final word persisted nearly as well as in the original form. This finding is in line with Bolinger (1979), who argues that the connection between FL's components has a "magical property of persisting, even when we knock some of them apart and put them together in unpredictable ways" (p. 96).

Regarding Research Question (5): Does the processing of L1 Chinese FSs differ between distinct kinds, such as Adjective+Noun collocations versus Verb+Noun collocations, relative to novel phrases? The answer is yes. That is, the processing of Chinese Verb+Noun collocations versus novel phrases was different from the processing of Adjective+Noun collocations versus novel phrases, to some degree. I found that *phrase type* significantly interacted with *part of speech* for the count measure in the phrase-final word analysis, with the final words in both Adjective+Noun collocations and Verb+Noun collocations processed significantly faster than in relative control phrases; and more interestingly, the magnitude of the processing advantage was slightly larger for final words in Verb+Noun collocations (effect size = 0.30) than in Adjective+Noun collocations (effect size = 0.22). This finding suggests that, although Chinese FL enjoys a processing advantage over novel language, the magnitude of the processing advantage might be different between distinct kinds of FSs. However, this tenet is speculative and requires further exploration using various types of FSs. Despite all this, the finding is still in line with existing research (e.g., Wolter & Yamashita, 2015). In a lexical decision study, Wolter and Yamashita (2015) found a significant processing advantage for English (Japanese-translated) Verb+Noun collocations over novel phrases in L1 users; however, the processing advantage never manifested in Adjective+Noun collocations. Wolter and Yamashita (2015) took this finding as "verbs subcategorize noun phrases, but adjectives subcategorize nothing" (p. 22). According to Wolter and Yamashita (2015), the different magnitude of processing advantage between Verb+Noun and Adjective+Noun collocations might be due to that the constituents of Verb+Noun collocations are more directly and closely connected with each other than the constituents of Adjective+Noun collocations. As suggested by the current finding, this tenet may also apply to FSs in their modified forms.

5.5 Summary

Chapter 5 reports on an eye-tracking study on the processing of modified Chinese FL in L1 users. The analysis of participants' eye movements suggests that Chinese FSs – Adjective+Noun collocations and Verb+Noun collocations – enjoy a processing advantage over their respective controls. The processing advantage was found significant both in the entire phrase and its final word, especially in the late processing stage. Crucially, the processing advantage for FSs persisted in their modified forms, with two and even four Chinese characters inserted in the middle of the FS; and, with the increase of modification degree, the magnitude of the processing advantage became smaller.

The present study has confirmed the processing advantage for modified Chinese FSs over novel phrases in L1 users. Whether or not this finding extends to L2 learners will be explored in Chapter 6.

Chapter 6

Study 2: The Processing of Modified Chinese Formulaic Sequences in L1 Users and L2 Learners: Evidence from Self-Paced Reading

6.1 Introduction

The previous chapter (Study 1) focused on the processing of modified Chinese formulaic sequences (FSs) in L1 users. Using the eye-tracking methodology, Study 1 confirmed that the processing advantage for formulaic language (FL) over novel language existed in Chinese FL and the FL advantage extended to its two modified forms, with two and four Chinese characters inserted in the middle of the phrase. However, this issue remains largely unexplored in L2. Due to limited experience (exposure) with the target language and the incongruency (in some but not all cases) between learners' L1 and the target language, the processing of L2 FL could be different from FL processing in L1, to some extent (for more information, please refer to *3.3 The processing of formulaic language in L2 learners*). Further, there could be processing differences among L2 learners of different proficiencies and from distinct L1 backgrounds (e.g., Carrol et al., 2016; Shantz, 2017; Siyanova-Chanturia et al., 2011b; Wolter & Gyllstad, 2011).

Existing studies on L2 FL processing have mainly focused on just one group of proficient learners (e.g., Jiang & Nekrasova, 2007; Siyanova-Chanturia & Janssen, 2018; Yi et al., 2017). Other studies observed learners of different language proficiencies and required them to self-rate their L2 proficiency (e.g., Siyanova-Chanturia et al., 2011b; Wolter & Yamashita, 2015, 2018) or do a vocabulary size test for the target language (e.g., Shantz, 2017; Sonbul & El-Dakhs, 2020; Wolter & Yamashita, 2015). Such scores, from self-rating or a vocabulary test, have been used to differentiate the language proficiency of L2 learners and then treated as a covariate or factor for language processing. Proficiency, derived from such measures, has been found to be a significant factor in phrasal processing: the higher the proficiency, the more accelerated and accurate the processing (e.g., Shantz, 2017; Siyanova-Chanturia et al., 2011b; Sonbul & El-Dakhs, 2020; Wolter & Yamashita, 2015, 2018). Despite how language proficiency is measured and established, most of the existing studies have focused on higher-proficiency learners, with L2 FL processing in other proficiency groups, especially those at lower proficiency levels being largely disregarded. According to

Wray (2002), there is a developmental trajectory of L2 FL processing, and the processing pattern of higher proficiency learners can be different from those at lower proficiency levels. In fact, existing studies have found that apart from faster processing speed, higher proficiency learners are more sensitive to FL than lower proficiency learners (e.g., e.g., Shantz, 2017; Siyanova-Chanturia et al., 2011b; Sonbul & El-Dakhs, 2020; Wolter & Yamashita, 2015, 2018). Thus, it is warranted to study L2 FL processing in learners of different proficiencies.

In addition, as suggested by existing research (e.g., Beck & Weber, 2016a, 2016b; Carrol & Conklin, 2014a, 2017; Carrol et al., 2016; Wolter & Gyllstad, 2011; Yamashita & Jiang, 2010), L1 background plays a role in L2 FL processing. However, existing studies have focused on either one background of L2 learners (e.g., L1 Chinese students who are learning English as their L2; e.g., Carrol & Conklin, 2014a, 2017; Carrol et al., 2016), or various L1 backgrounds but without comparisons (e.g., L2 English learners who are from China, France, Germany, Greece, Italy, Saudi Arabia, Spain, or The Netherlands, and so forth; e.g., Sonbul, 2014; Vilkaitė & Schmitt, 2019). To the best of my knowledge, no study has to date investigated L2 FL processing in learners from distinct L1 backgrounds, for example, Japanese and Thais, represented by a sizable population with the aim of making direct comparisons between them.

To address the gaps mentioned above, a self-paced reading study was conducted to answer the following questions:

(1) Do L2 Chinese FSs show a processing advantage over novel phrases?
(2) If so, does this processing advantage extend to modified FSs? If so, does the degree of modification, operationalised as insertion length, affect the processing advantage?
(3) Is there any difference in the processing of Chinese FSs relative to novel phrases in L2 learners of different language proficiencies?
(4) Is there any difference in the processing of Chinese FSs relative to novel phrases in L2 learners from distinct L1 backgrounds, such as Japanese versus Thai?
(5) In addition, is there any difference in the processing of L2 Chinese FSs between distinct kinds, such as Adjective+Noun collocations versus Verb+Noun collocations, relative to novel phrases?

6.2 Method

6.2.1 Materials

The materials were identical to those used in Study 1. Although the target items in the materials were extracted from a L1 corpus (BCC corpus), they were deemed appropriate for L2 learners. As noted by Wray (2002), L2 FL research tends to identify FSs based on the norms of what is formulaic for L1 users, rather than the norms of the learner population.

All individual words within the target FS were checked and confirmed to be highly frequent (frequency > 10,000 as registered in BCC corpus) and marked as Level 4 and under in the lexical syllabus of the Chinese Proficiency Test (HSK, Chinese Testing International, 2012). This lexical syllabus is the official syllabus for the HSK, an international standardized test of Chinese language proficiency that assesses the abilities of non-native Chinese speakers in using the Chinese language in their daily, academic, and professional lives. The HSK consists of six levels, with Level 1 as the lowest and Level 6 as the highest. According to the lexical syllabus, Chinese language learners at Level 1 should acquire 150 words, at Level 2 should acquire 300 words, at Level 3 should acquire 600 words, at Level 4 should acquire 1,200 words, at Level 5 should acquire 2,500 words, and at Level 6 should acquire 5,000 words.

6.2.2 Participants

Sixty-seven L2 learners of Chinese participated in this study (62 female adults, mean age = 22.16). Among them, 37 participants were from Japan and 30 were from Thailand. On average, they first learned Chinese (Mandarin) at the age of 17.52 (ranging from three years old to 38 years old). They had been living in China for 15.04 months on average (ranging from two months to 120 months) and had been learning Chinese for 3.69 years on average (ranging from one year to 17 years). Their Chinese language proficiency was self-rated in a background information form, on a 10-point Likert scale (ranging from 1 '*很差* (*very poor*)' to 10 '*很好* (*excellent*)'), for reading, writing, listening, and speaking skills. Their average rating scores for these four language skills were 5.61 (ranging from 2 to 10), 5.47 (ranging from 2 to 9), 5.99 (ranging from 1 to 10), and 5.51 (ranging from 1 to 10), respectively.

These two L1 backgrounds were chosen to investigate whether the L1-L2 congruency in word order may affect L2 phrasal processing. In the Japanese language, the word order in Adjective+Noun collocations is congruent with Chinese (e.g., *白衬衫 (white shirt) vs 白いシャツ (white shirt)*; e.g., Bunt, 2003), while it is largely incongruent with Chinese in terms of

Verb+Noun collocations (e.g., *坐飞机 (take a plane) vs 飛行機に乗る (a plane take)*; e.g., Bunt, 2003). In the Thai language, the word order for Verb+Noun collocates is congruent with Chinese (e.g., *包饺子 (make dumplings) vs ทำเกี๊ยว (make dumplings)*; e.g., Iwasaki & Ingkaphirom, 2005), while it is largely incongruent with Chinese in the case of Adjective+Noun collocations (e.g., *白衬衫 (white shirt) vs เสื้อเชิ้ตสีขาว (shirt white)*; e.g., Iwasaki & Ingkaphirom, 2005). Since learners draw on their L1 knowledge to process L2 phrases, we would expect that Japanese speakers may show greater FL advantages for and sensitivity to Adjective+Noun collocations, and Thai speakers may show greater FL advantages for and sensitivity to Verb+Noun collocations.

As a control group, 39 full-time university students in Beijing who were L1 users of Chinese (Mandarin), were also recruited. They consisted of 33 female and six male adults, between the ages 18 and 28, with a mean age of 21.46. All of them were born and raised in Mainland China. Although the Chinese participants were not required to rate their language proficiency, the highest score (i.e., ‘10’) was nevertheless assigned to them for statistical modelling (see Siyanova-Chanturia et al., 2011b, for a similar procedure).

All participants received a supermarket voucher for their participation (funded by the Faculty of Humanities and Social Sciences Research Grant, No. 221046). All participants were informed of their rights and gave their written consent before the experiment. This study was carried out in line with the ethical procedures of the Human Ethics Committee at VUW (Ethics Approval: 0000026508).

6.2.3 Apparatus and procedures

The self-paced reading experiment was carried out individually in a quiet laboratory at a university in Beijing. Each participant was first given Chinese written instructions and a verbal explanation of the procedure before the experiment. After that, a self-reported questionnaire was given to the L2 learners to record their demographic information in general, and their Chinese language proficiency in particular. The questionnaire asked about the participant’s gender, age, L1 background, number of years studying Chinese, age of first contact with Chinese, and number of months living in Mainland China, followed by the 10-point scale to rate their proficiency in Chinese reading, writing, listening, and speaking. Their HSK scores were also noted.

The presentation of materials and the collection of reaction times were accomplished using the DMDX software (version 5.3.1.3, Forster, 2002) on a Dell laptop. Each sentence was presented in a portion-by-portion fashion across one line in the centre of the screen using a self-paced reading procedure with non-cumulative moving windows (e.g., Molinaro et al., 2013). Each sentence was divided into three portions (e.g., Tremblay et al., 2011): (1) portion 1: everything prior to the target phrase; (2) portion 2: the target phrase, which is identical to the whole phrase region of interest in Study 1 (see Chapter 5); and (3) portion 3: everything after the target phrase. For example:

(1) portion 1: 这件################## (This##################);

(2) portion 2: ####白衬衫########## (####white shirt##########);

(3) portion 3: ##########是他哥哥的 (##########is his elder brother's).

The Chinese characters in each sentence were displayed in *Song* font (identical to Study 1), size 14 (to allow the longest sentence to appear across one line in the DMDX program), without word spaces.

Each trial began with a button press ('RB' button) on a gamepad (Logitech Gamepad F310) connected to the display laptop. After pressing this button, portion 1 appeared on the screen with the other two portions masked by hashtags. When the participant pressed the 'RB' button for the second time, the target phrase portion (portion 2) appeared on the screen with portions 1 and 3 replaced by hashtags. When the 'RB'button was pressed for the third time, portions 1 and 2 were replaced by hashtags, with only portion 3 displayed. Reaction times (RTs) were measured on the critical portion (i.e., portion 2) from the onset of portion 2 to the onset of portion 3.

Akin to the material presentation procedure in Study 1, the experimental session was preceded by a practice session, which included six trial sentences. After that, the participant (either a L1 user or a L2 learner) was assigned to one of the three lists. Each list contained 144 sentence trials, including 96 target sentences and 48 filler sentences (see *5.2.3 Apparatus and procedures* section in Study 1). Participants were required to read the sentences for comprehension. Every three sentences on average (e.g., Carrol & Conklin, 2020), participants were asked to answer a True/False comprehension question, by pressing 'LB' (False) or 'RB' (True) on the gamepad. No feedback was given following the button press (e.g., Conklin & Schmitt, 2008; Jiang & Nekrasova, 2007; Kim & Kim, 2012; Tremblay et al., 2011). The 144 sentences were separated into three blocks, and the participant could take a short break after each block. This experiment lasted approximately 25 minutes from start to finish. All 106 participants successfully completed the experiment.

6.3 Data analysis and results

6.3.1 Data analysis

Prior to the analysis, reaction times with incorrect responses to the comprehension question were excluded from the entire data. This resulted in 3.26% of data loss (see Table 14 for more details).

Table 14:

The Exclusions of L1 and L2 Chinese Speakers' Incorrect Responses for the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition, with Different Insertion Lengths

	Original				**Short insertion**				**Long insertion**			
	A+N phrase		*V+N phrase*		*A+N phrase*		*V+N phrase*		*A+N phrase*		*V+N phrase*	
	COL	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**
L1 speakers (Chinese)	0.06%	0.06%	0.02%	0.04%	0.04%	0.21%	0.05%	0.12%	0.01%	0.12%	0.09%	0.04%
L2 speakers (Japanese)	0.14%	0.09%	0.05%	0.13%	0.13%	0.14%	0.10%	0.12%	0.14%	0.27%	0.15%	0.15%
L2 speakers (Thai)	0.02%	0.11%	0.04%	0.11%	0.05%	0.17%	0.04%	0.04%	0.03%	0.15%	0.06%	0.03%

Note. The estimated value of each exclusion rate was rounded and kept to four decimal places.

As suggested by existing studies (e.g., Gu, Li, & Liversedge, 2015; Kim & Kim, 2012), reaction times that are above or below three standard deviations (3 SDs) from the mean of each participant should be discarded from the data analysis. This procedure was adopted after the exclusion of incorrect responses, which led to 1.82% of data loss (see Table 15 for more details).

Table 15:

The Exclusions of L1 and L2 Chinese Speakers' Irregular Responses for the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition, with Different Insertion Lengths

	Original				Short insertion				Long insertion			
	A+N phrase		*V+N phrase*		*A+N phrase*		*V+N phrase*		*A+N phrase*		*V+N phrase*	
	COL	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**
L1 speakers (Chinese)	0.02%	0.05%	0.02%	0.12%	0.07%	0.09%	0.05%	0.04%	0.06%	0.06%	0.04%	0.06%
L2 speakers (Japanese)	0.10%	0.06%	0.05%	0.07%	0.06%	0.05%	0.01%	0.06%	0.04%	0.07%	0.01%	0.05%
L2 speakers (Thai)	0.03%	0.08%	0.01%	0.03%	0.06%	0.07%	0.00%	0.05%	0.06%	0.07%	0.01%	0.01%

Note. The estimated value of each exclusion rate was rounded and kept to four decimal places.

All participants answered comprehension questions with ease (mean accuracy rate for Chinese readers: 95.41%, for Japanese readers: 90.93%, for Thai readers: 94.10%). Descriptive statistics for the 12 experimental conditions (two phrase types × three insertion lengths × two parts of speech) across the three language background groups are presented in Table 16.

Table 16:

L1 and L2 Chinese Speakers' Mean Reaction Times (in Milliseconds) for the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition in Different Insertion Lengths, with Standard Deviation (SD) in Parentheses

	Original						**Short Insertion**						**Long Insertion**					
	A+N Phrase			*V+N Phrase*			*A+N Phrase*			*V+N Phrase*			*A+N Phrase*			*V+N Phrase*		
	COL	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***
L1 speakers (Chinese)	512.51 (202.12)	659.27 (310.32)	< .001	483.84 (199.77)	652.03 (234.48)	< .001	606.60 (307.66)	722.85 (315.16)	< .001	579.53 (302.59)	739.82 (371.36)	< .001	774.70 (535.86)	881.38 (572.56)	.01	738.88 (406.31)	841.13 (489.73)	< .01
L2 speakers (Japanese)	1410.50 (959.37)	1775.63 (1204.92)	< .001	1170.48 (749.55)	1591.79 (992.33)	< .001	2068.77 (1454.70)	2347.85 (1240.60)	< .01	1934.35 (1253.75)	2367.28 (1173.90)	< .001	2784.07 (1732.97)	3106.05 (2023.81)	.03	2761.57 (1689.51)	2788.97 (1575.88)	.42
L2 speakers (Thai)	1775.22 (1043.01)	2156.56 (1266.87)	< .001	1384.33 (886.76)	1913.36 (1050.02)	< .001	2792.39 (1882.22)	3081.70 (1921.59)	.06	2367.73 (1378.52)	2835.71 (1619.67)	< .001	3655.11 (2851.76)	3805.21 (2165.08)	.27	3314.46 (1745.09)	3450.18 (2061.33)	.22

Notes. The mean reaction times are the original values (not log-transformed). The estimated value of each mean reaction time was rounded and kept to two decimal places.

A linear mixed-effects model was established to analyse the reaction time measure (dependent variable) in R (version 3.6.1, 2019), package *lme4* (version 1.1-21, 2019). The *p*-value for each predictor (variable) was estimated using the *lmerTest* package (version 3.1-1, 2019). All reaction times were log-transformed before model fitting. The coding method for each categorical variable (i.e., *phrase type*, *insertion length*, and *part of speech*) was identical to Study 1. Besides, *language background* was regarded as a nominal variable, assigning 'c' to L1 Chinese speakers (reference group), 'j' to Japanese students (L2 Chinese learners), and 't' to Thai students (L2 Chinese learners). The self-rating score, which was averaged among ratings for the four language skills (i.e., Chinese reading, writing, listening, and speaking) was used as the *language proficiency* for each participant (e.g., Siyanova-Chanturia et al., 2011b). This numeric (continuous) variable was centred before model fitting.

The model construction was akin to Study 1. To address Research Question (1), phrase type was first added to the maximal (original) model as a main predictor. To address Research Question (2), insertion length and its interaction with phrase type were added to the model. To address Research Question (3), language proficiency and its two-way interaction with phrase type and the three-way interaction involving language proficiency, phrase type, and insertion length were added to the model. To address Research Question (4), language background and its two-way interaction with phrase type and the three-way interaction involving language background, phrase type, and insertion length were added to the model. To address Research Question (5), the two-way interaction between part of speech and phrase type and the three-way interaction involving part of speech, phrase type, and insertion length were added to the model. However, due to the convergence issue and based on the suggestion of the *step* function (a function in the *stats* package (version 3.6.2), developed for formula-based model selection by the Akaike Information Criterion in a stepwise algorithm), some non-significant interactions were removed from the maximal model. The removal included interactions between language background, phrase type, and insertion length, and between part of speech, phrase type, and insertion length. Model comparisons showed that the removal did not significantly affect the maximal model ($p > .10$).

The random-effects structure included the participant random intercept, item random intercept, and random slopes of the main predictors by participants and by items. The procedures of model fitting and further comparisons/contrasts of categorical variables were identical to the practice adopted in Study 1. Further analysis of the continuous variable (i.e., language proficiency in the current study) was done by employing the *aov* function in the *stats* package (version 3.6.2), with its effect size calculated using the *etaSquared* function in the *lsr* package (version 0.5). The multicollinearity issue was also checked for the model (*Kappa* = 31.51) and the predictors (*vif* < 6).

6.3.2 Results

Fixed effects and random effects of the selected model for the reaction times of L1 users and L2 learners are presented in Table 17. *Phrase type* was revealed to be a significant predictor ($t = -8.60$, $p < .001$). Further comparisons showed that collocations were processed faster than controls ($t = 9.34$, $p < .001$, effect size = 0.56).

Language proficiency was another significant predictor ($t = -2.66$, $p < .01$). Further analysis revealed that higher proficiency readers read faster than lower proficiency readers ($F = 6359.00$, $p < .001$, effect size = 0.40).

Language background was also shown to be a significant predictor ($t_{Japanese\ vs\ Chinese} = 3.49$, $p < .001$; $t_{Thai\ vs\ Chinese} = 6.76$, $p < .001$). Post-hoc analysis revealed that Japanese participants read faster than Thais ($t = -3.68$, $p < .01$, effect size = 0.89).

Table 17:

The Selected Mixed Effects Model for Study 2

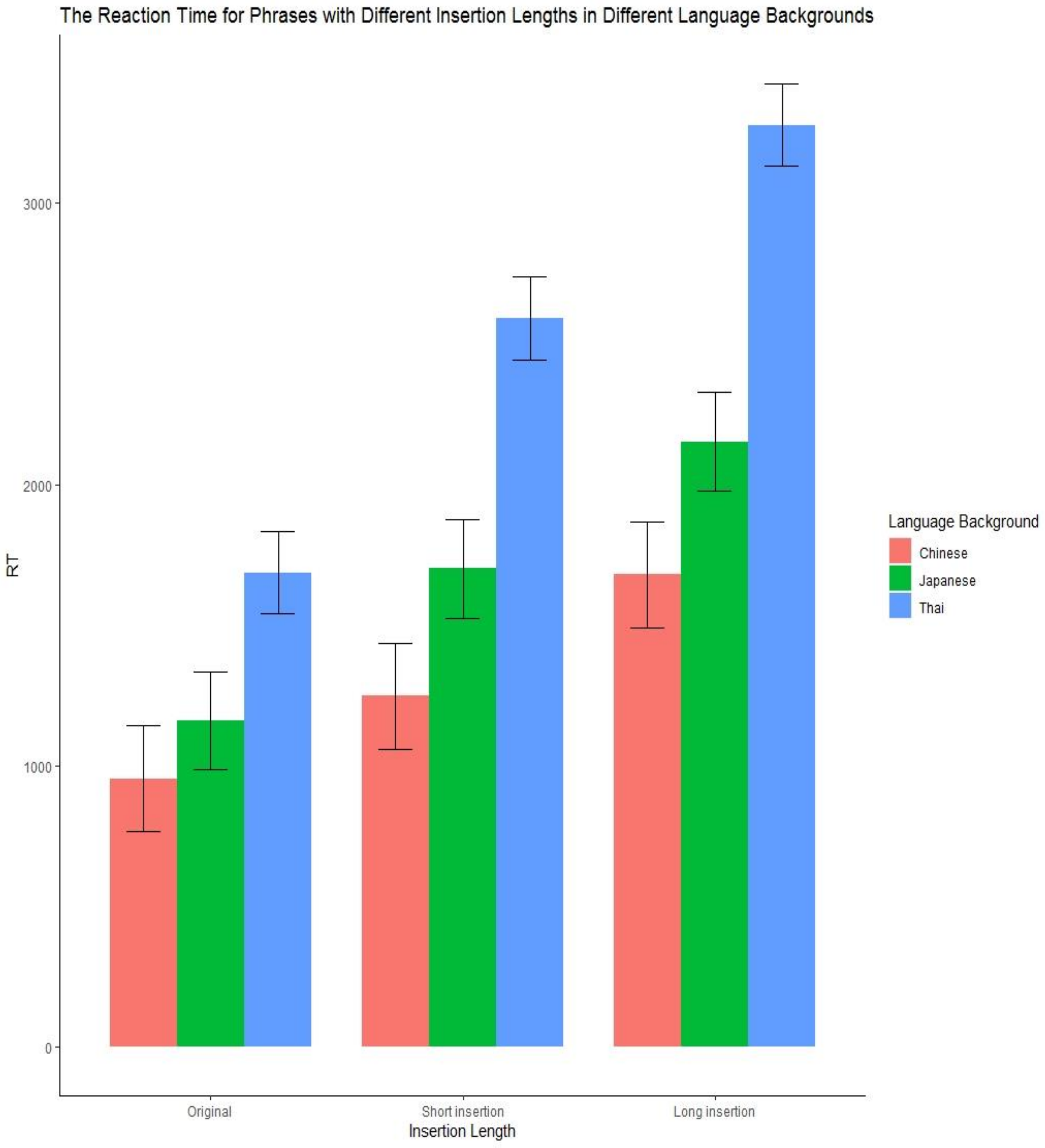
Fixed effects	*Estimate*	*SE*	*df*	*t*	*pr*
Intercept	6.48	0.09	113.10	70.81	< .001
Phrase type	-0.15	0.02	279.80	-8.60	< .001
Insertion length	0.07	0.01	921.40	7.81	< .001
Language proficiency	-0.07	0.02	106.50	-2.66	< .01
Language background (Japanese vs Chinese)	0.53	0.15	106.40	3.49	< .001
Language background (Thai vs Chinese)	0.84	0.12	106.30	6.76	< .001
Phrase type × Language proficiency	0.00	0.00	9277.00	0.98	0.33
Insertion length × Language proficiency	-0.00	0.00	9303.00	-0.73	0.47
Phrase type × Insertion length	0.02	0.01	280.20	3.48	< .001
Insertion length × Language background (Japanese vs Chinese)	0.09	0.01	9300.00	8.46	< .001
Insertion length × Language background (Thai vs Chinese)	0.10	0.01	9292.00	10.84	< .001
Phrase type × Insertion length × Language proficiency	-0.00	0.00	9277.00	-2.61	< .01
Random effects	*Variance*	*SD*			
Item intercept	0.03	0.17			
Participant intercept	0.10	0.32			
Residual	0.13	0.36			

Notes. The estimated value of each predictor was rounded and kept to two decimal places.

Language background was found to interact with *insertion length* ($t_{Japanese\ vs\ Chinese \times Insertion\ length}$ = 8.46, p < .001; $t_{Thai\ vs\ Chinese \times Insertion\ length}$ = 10.84, p < .001). This interaction is plotted in Figure 4. Further analysis showed that Japanese students read faster than Thai students across all insertion length conditions ($t_{original}$ = -3.47, p < .01, effect size = 0.85; $t_{short\text{-}insertion}$ = -3.76, p < .001, effect size = 0.92; $t_{long\text{-}insertion}$ = -3.67, p < .01, effect size = 0.90).

Figure 4:

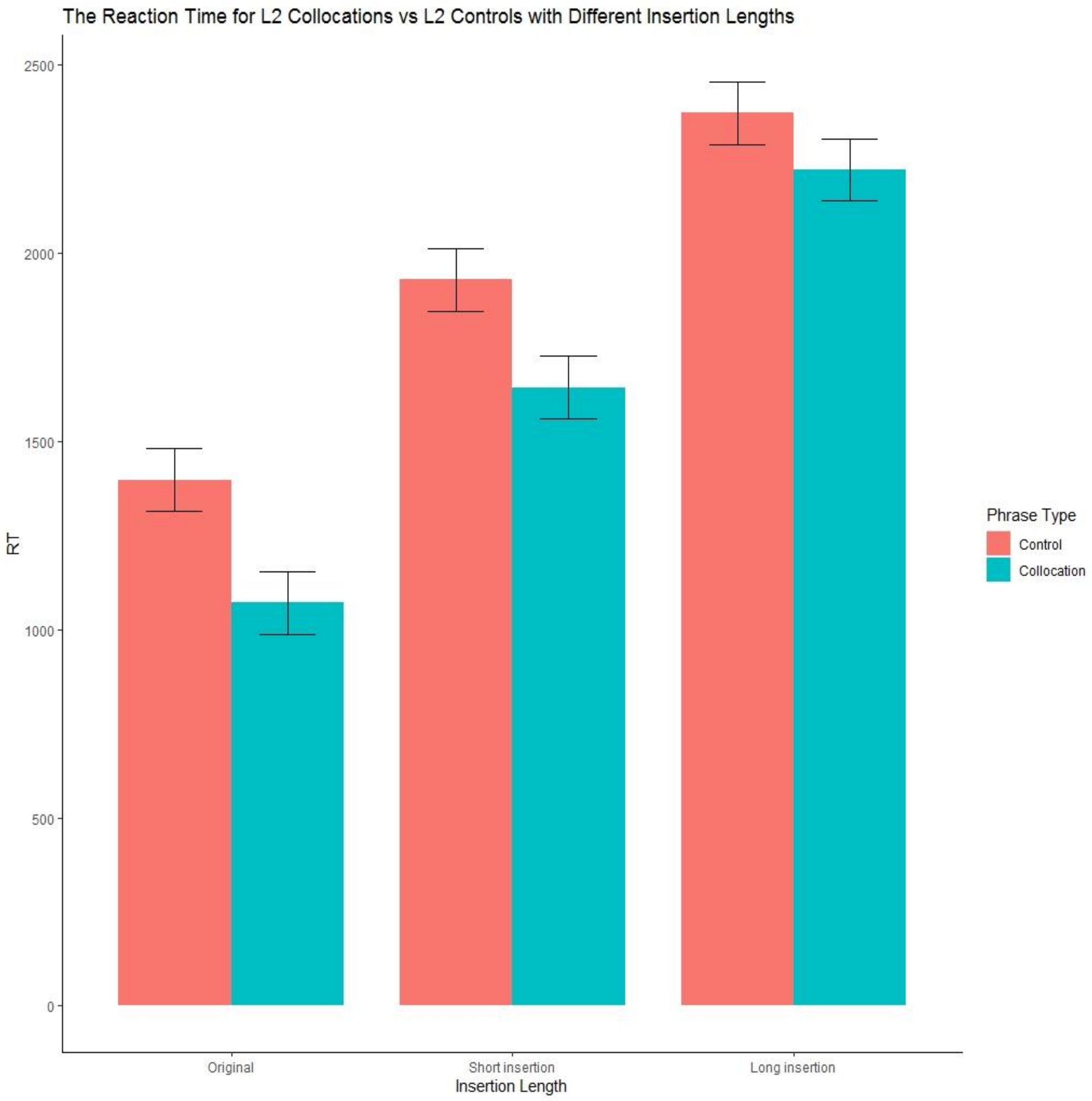
The Interaction between Language Background and Insertion Length, with 95% Confidence Intervals for Each Condition



Crucially, *phrase type* was found to significantly interact with *insertion length* (t = 3.48, p < .001). This interaction is plotted in Figure 5. Further analysis revealed that collocations were processed faster than controls across all insertion length conditions, and, more crucially, the advantage for collocations diminished with longer insertions ($t_{original}$ = 7.78, p < .001, effect size = 0.81; $t_{short\text{-}insertion}$ = 5.55, p < .001, effect size = 0.58; $t_{long\text{-}insertion}$ = 2.85, p < .01, effect size = 0.30).

Figure 5:

The Interaction between Phrase Type and Insertion Length in L1 and L2 Speakers, with 95% Confidence Intervals for Each Condition



It is noteworthy that *language proficiency* interacted with *phrase type* and *insertion length* (t = -2.61, p < .01). This interaction is plotted in Figure 6. Further analysis showed that for L1 users, collocations were processed significantly faster than relative controls across all insertion length conditions ($t_{original}$ = 6.95, p < .001, effect size = 0.79; $t_{short\text{-}insertion}$ = 5.59, p < .001, effect size = 0.64; $t_{long\text{-}insertion}$ = 3.84, p < .001, effect size = 0.44). When it came to L2 learners, the processing advantage for collocations was found significant in the original and short-insertion conditions ($t_{original}$ = 7.64, p < .001, effect size = 0.82; $t_{short\text{-}insertion}$ = 5.05, p < .001, effect size = 0.54), and somewhat smaller in the long-insertion condition ($t_{long\text{-}insertion}$ = 1.98, p = .05, effect size = 0.21). A more fine-grained figure for this three-way interaction in L2 learners is plotted in Figure 7.

Figure 6:

The Interaction between Phrase Type and Insertion Length in L1 vs L2 Speakers, with 95% Confidence Intervals for Each Condition

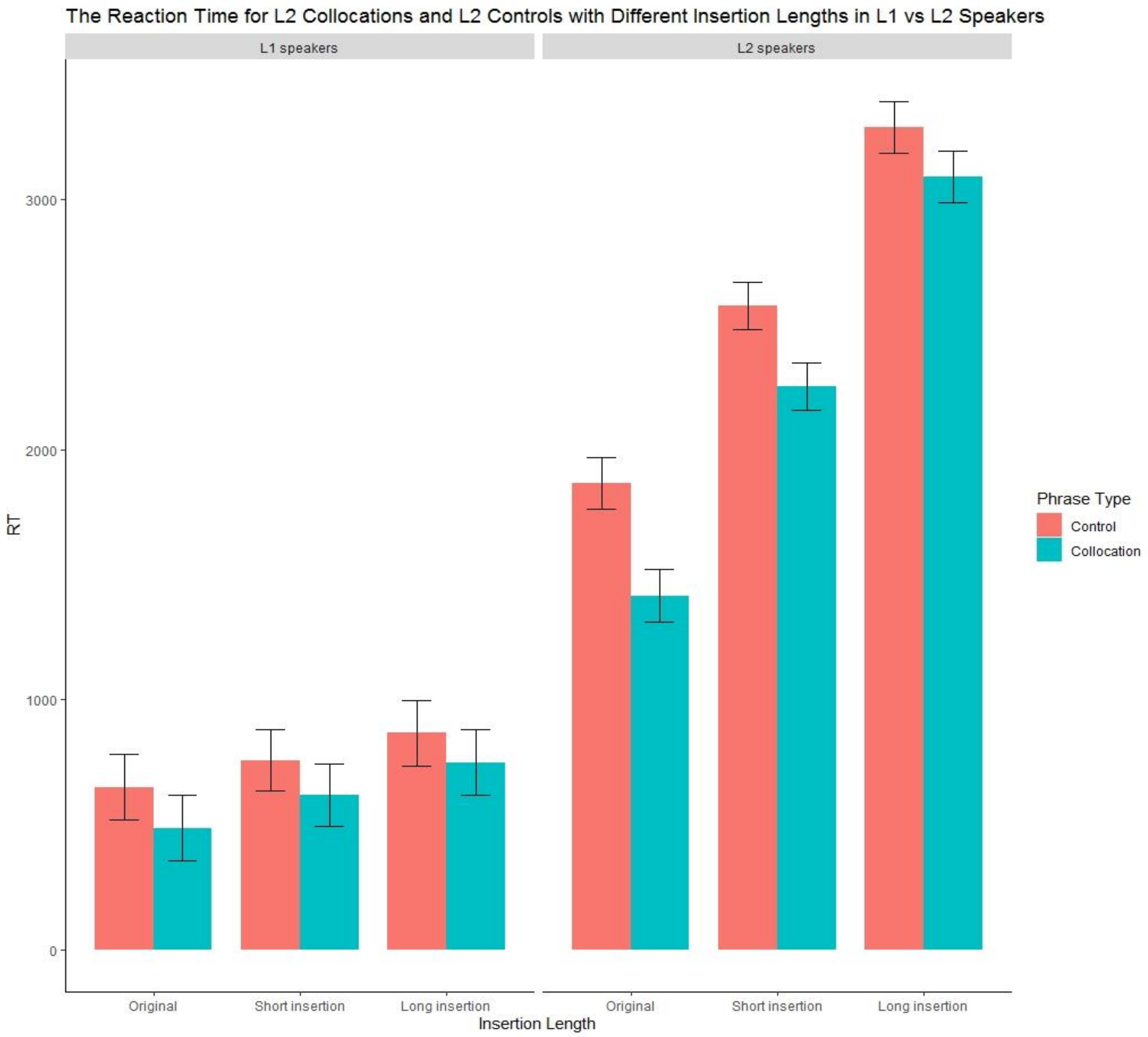
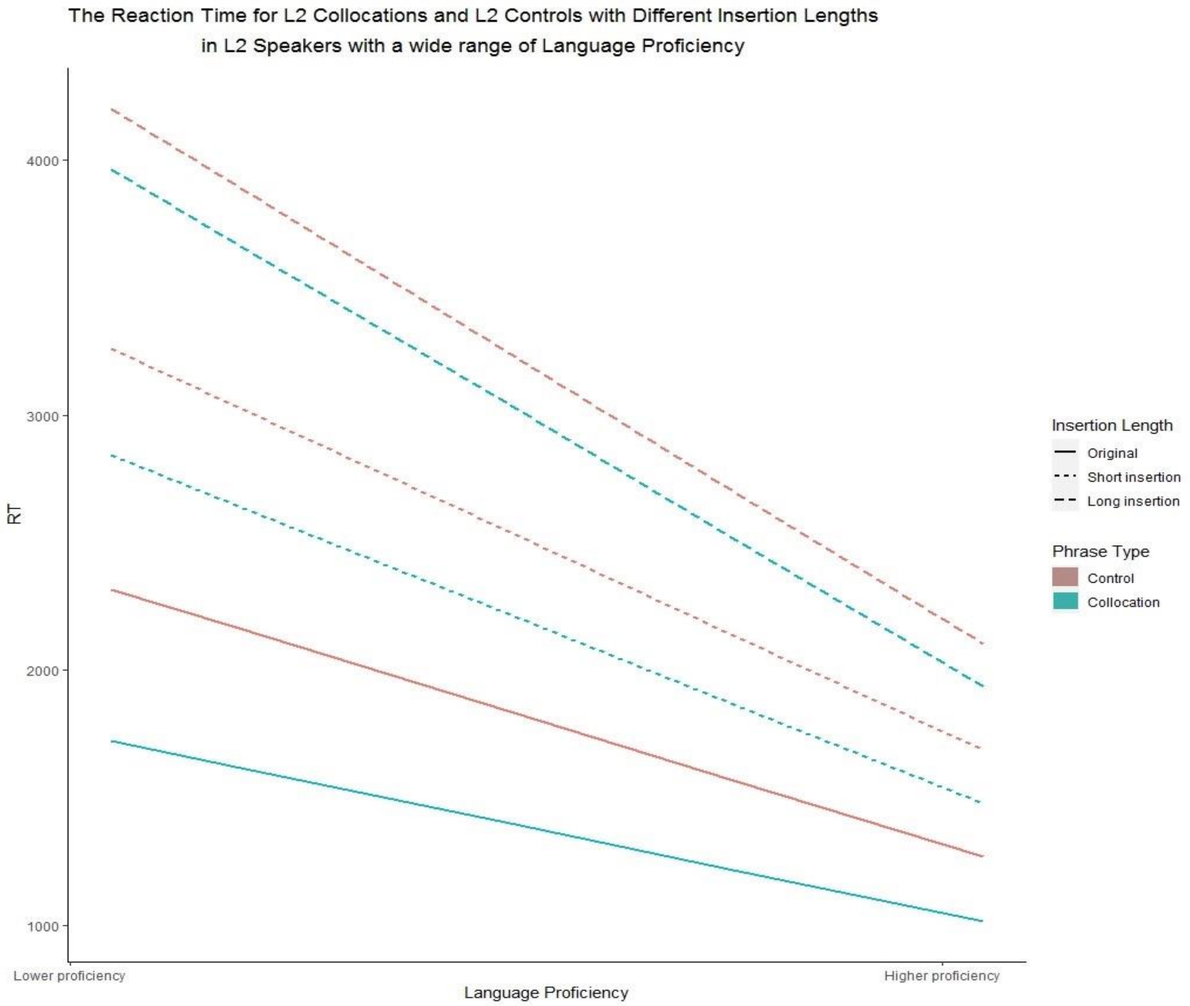


Figure 7:

The Interaction between Language Proficiency, Phrase Type, and Insertion Length in L2 Speakers



No significant interactions were found between *language background* and *phrase type* ($p > .10$), between *language background, phrase type*, and *insertion length* ($p > .10$), between *part of speech* and *phrase type* ($p > .10$), or between *part of speech, phrase type*, and *insertion length* ($p > .10$). These insignificant interactions were excluded from the final model due to model fitting.

6.4 Discussion

Chinese FL and its processing advantage over novel language have been affirmed in the existing literature (e.g., Jiang et al., 2020; Kong et al., 2016; Liu et al., 2010; Shen et al., 2017; Yu et al., 2016; Zhang et al., 2013; Zhou et al., 2004). However, the relevant evidence pertains almost exclusively to FL processing in L1 users. To the best of my knowledge, only one published study investigated the processing of Chinese FL in L2 learners (Yi et al., 2017). In this existing study, Yi et al. (2017) examined the processing of frequent FSs in their canonical form with a group of proficient L2 learners from a wide range of L1 backgrounds. Whether or not L2 Chinese learners from distinct L1 backgrounds and of different language proficiencies exhibit a processing advantage for FL over novel language, and whether this FL advantage remains in its modified form, has to date been largely disregarded. In the present investigation, I sought to address this gap and the research questions brought up at the beginning of the current chapter in particular, by asking L1 users (as a reference group) and L2 learners from Japan and Thai with a relatively wide range of language proficiency to read Chinese (Mandarin) phrases varying in frequency, embedded in sentence context. I recorded the reaction times for phrases of these readers and, in particular, I examined how *phrase type* (collocation vs control), *degree of modification* (i.e., insertion length, with original vs short-insertion vs long-insertion), *language proficiency* (continuum), *language background* (Japanese vs Thai), and *part of speech* (Adjective+Noun phrase vs Verb+Noun phrase) might have affected the reading of L2 Chinese phrases.

Regarding Research Question (1): Do L2 Chinese FSs show a processing advantage over novel phrases? The answer is yes. That is, L2 Chinese FSs demonstrated a significant processing advantage over novel phrases. I found a significant main effect of *phrase type* with collocations read faster than control phrases. This finding reminds us of several existing studies (e.g., Conklin & Schmitt, 2008; Gyllstad & Wolter, 2016; Jiang & Nekrasova, 2007; Kim & Kim, 2012; Northbrook & Conklin, 2019; Shantz, 2017; Siyanova-Chanturia et al., 2011b; Sonbul, 2014; Underwood et al., 2004; Vilkaitė & Schmitt, 2019; Yi, 2018). For example, in a self-paced reading study, Kim and Kim (2012) asked L1 and L2 English speakers to read sentences embedded with Verb+out combinations at four phrase-frequency levels: high frequency (1200–2000 occurrences; e.g., *find out*), mid–high frequency (500–1000 occurrences; e.g., *sort out*), mid–low frequency (200–400 occurrences; e.g., *miss out*), and low frequency (20–100 occurrences, e.g., *test out*). Each sentence was presented in a portion-by-portion pattern. Kim and Kim (2012) found both L1 and L2 speakers read higher frequency phrases significantly faster than lower frequency phrases, suggesting that L2 frequent collocations had a processing advantage over infrequent phrases. The finding in the present investigation is in line with existing research, such as Kim and Kim (2012).

Regarding Research Question (2): Does the processing advantage for L2 FSs extend to their modified form? If so, does the degree of modification, operationalised as insertion length, affect the processing advantage? The answer is also yes. That is, the processing advantage for L2 Chinese FSs persisted in their modified forms; but, with the increase of modification degree, the magnitude of the processing advantage diminished. I found *modification degree* significantly interacted with *phrase type* in the analysis of reaction times, such that L2 Chinese collocations were read faster than respective controls not only in their original form ($p < .001$, effect size = 0.81), but also in their short-insertion form ($p < .001$, effect size = 0.58) and long-insertion form ($p < .01$, effect size = 0.30), although the processing facilitation greatly diminished with longer insertions. This finding reminds us of another modified L2 FS reading study (Vilkaitė & Schmitt, 2019). In this study, Vilkaitė and Schmitt (2019) asked L2 English learners to read sentences embedded with Verb+Noun collocations and relative controls, in their original (e.g., *provide information* vs *compare information*) and modified forms (with three words inserted in the middle of the phrase: e.g., *provide* <u>some of the</u> *information* vs *compare* <u>some of the</u> *information*). Vilkaitė and Schmitt (2019) found a significant interaction between phrase type and modification: while L2 collocations were processed significantly faster than control phrases in their original form ($p < .001$), the processing advantage for collocations disappeared in their modified form ($p > .10$). The finding of the present study, however, points to a different picture. Unlike the finding in Vilkaitė and Schmitt (2019), the processing advantage for L2 Chinese collocations existed in their original form and did not entirely disappear, instead, remained to their modified forms, although the magnitude of processing advantage diminished with the increase of modification degree. This finding is novel and noteworthy, as none of the previous studies found the processing advantage for modified L2 FL.

Regarding Research Question (3): Is there any difference in the processing of Chinese FSs relative to novel phrases in L2 learners of different language proficiencies? The answer is also yes. That is, the processing of Chinese FSs versus novel phrases was different between learners of different L2 proficiencies, with higher proficiency learners processing FSs and respective controls faster than lower proficiency learners, and this fact remained true even if phrases were modified. First, I found a significant main effect of *language proficiency* for the reaction times, with higher proficiency learners processing L2 phrases faster than lower proficiency learners. This is because, compared with lower proficiency learners, higher proficiency learners had gained more knowledge of, and experience (exposure) with the target language, which facilitated their processing of L2 phrases. This finding is consistent with existing research (e.g., Matlock & Heredia, 2002; Shantz, 2017; Siyanova-Chanturia et al., 2011b; Sonbul, 2014; Sonbul & El-Dakhs, 2020; Vilkaitė & Schmitt, 2019), suggesting that language proficiency plays a role in L2 phrasal processing.

More importantly, *language proficiency* was found to significantly interact with *phrase type* and *modification degree*, such that Chinese collocations were read faster than relative controls across their original and modified forms by L2 learners across the continuum of language proficiency, and, with the development of language proficiency, the reading time of L2 phrases decreased. This finding indicates that language proficiency plays a role in the processing of L2 phrases, not only in their original form but also in their modified forms. This finding reminds us of a recent study on the reading of modified L2 English FSs (Vilkaitė & Schmitt, 2019). In this existing study, Vilkaitė and Schmitt (2019) also found a significant interaction between language proficiency, phrase type, and modification. However, further analysis showed that although English learners across the continuum of language proficiency exhibited a significant processing advantage for L2 collocations over relative controls, this processing advantage was almost exclusively confined to collocations in their original form and disappeared in their modified form. One possible explanation for the different results of the processing advantage for modified L2 FSs in Vilkaitė and Schmitt (2019) and the current study is that, the L2 FSs used in the current study could be more familiar to learners than those in Vilkaitė and Schmitt (2019). For example, '*包（虾仁）饺子*' (*make dumplings (stuffed with prawns)*) could be more familiar to L2 learners than '*provide (some of the) information*', such that the former FS could be easier for L2 learners to identify and process than the latter FS.

Regarding Research Question (4): Is there any difference in the processing of Chinese FSs relative to novel phrases in L2 learners from distinct L1 backgrounds, such as Japanese versus Thai? The answer is also yes. That is, the processing of Chinese FSs versus novel phrases was different between L2 learners from Japan and Thailand, with Japanese students processing collocations and controls significantly faster than Thai students, and this fact remained even when phrases are modified. I found a significant main effect of *language background*, which also interacted with *modification degree* for the reaction times during L2 Chinese phrasal processing. Despite the fact that the word order of Verb+Noun combinations in Chinese is congruent with Thai but incongruent with Japanese, Japanese students were always faster than Thai students in the processing of L2 Chinese phrases, both in their original and modified forms. One possible explanation for this finding is that, compared with the L1-L2 similarities between Thai and Chinese texts, there are more similarities between Japanese and Chinese texts. This is because the writing system of Japanese uses characters with identical meaning(s) derived from Chinese (e.g., *Kanji* characters, *Kana* characters, etc.), to a great extent (e.g., Rayner et al., 2012; Rayner & Pollatsek, 1989; Taylor, 1981). In this case, Japanese students could be more familiar with Chinese text than Thai students and the familiarity further facilitated their L2 Chinese phrasal processing.

Regarding Research Question (5): Is there any difference in the processing of L2 Chinese FSs between distinct kinds, such as Adjetive+Noun collocations versus Verb+Noun collocations, relative to novel phrases? The answer might be no, as I did not find any significant interactions between *part of speech* and *phrase type* (and *modification degree*). This finding is consistent with existing studies (e.g., Sonbul & El-Dakhs, 2020; Wolter & Yamashita, 2015). For example, in a lexical decision study, Wolter and Yamashita (2015) found that unlike L1 users who processed Verb+Noun collocations (e.g., *answer phone*) faster than Adjective+Noun collocations (e.g., *heavy traffic*), L2 learners processed these two collocation types at a similar pace. This finding suggests that the processing of L2 FSs might not be different between distinct kinds. However, this tenet is speculative and requires further exploration using various types of L2 FSs and more sensitive research techniques, such as eye-tracking and EEG.

6.5 Summary

Chapter 6 reports on a self-paced reading study on the processing of modified L2 Chinese FSs in L1 users and, in particular, L2 learners from distinct L1 backgrounds and of different language proficiencies. The analysis of the reading times suggests that akin to L1 users, L2 learners exhibited a processing advantage for Chinese FSs – Adjective+Noun collocations and Verb+Noun collocations – over respective novel phrases. Crucially, the processing advantage for FSs persisted in their modified forms, with two and even four Chinese characters inserted in the middle of the FS, but, with the increase of modification degree, the magnitude of the processing advantage diminished. In addition, learners of different L2 proficiencies were found to show different speed for Chinese FS processing, with higher proficiency learners reading faster than lower proficiency learners. Furthermore, L2 learners from distinct L1 backgrounds also showed different speed for Chinese FS processing, with Japanese students reading faster than Thai students.

The studies in the present and previous chapters have confirmed the processing advantage for modified Chinese FSs over novel phrases in L1 and L2 adults. In the following chapter, I explore whether or not the processing advantage for modified FL over novel language exists in children.

Chapter 7

Study 3: The Processing of Modified Chinese Formulaic Sequences in L1 Children and Adults: Evidence from Self-Paced Reading

7.1 Introduction

Study 1 and Study 2 presented in earlier chapters focused on adult processing of modified Chinese formulaic sequences (FSs). However, the processing of modified FSs in children has remained unexplored in the current literature, and it could be problematic to directly extend the results of adults to children without verification (e.g., Jiang et al., 2020; Wray, 2002). Given that formulaic language (FL) plays an important role in language development (e.g., Clark, 1974; Foster-Cohen, 1999; Havron & Arnon, 2020; Lieven et al., 1992; Nelson, 1973; Peters, 1977; Wray, 2002), a better understanding of how it is processed by children at different age groups is needed. In this chapter, a self-paced reading study was conducted to examine these four questions:

(1) Do children show a processing advantage for Chinese FSs over novel phrases?

(2) If so, does the processing advantage extend to modified Chinese FSs in children? If so, does the degree of modification, operationalised as insertion length, affect Chinese FS processing in children?

(3) Does the age of children play a role in the processing of Chinese FSs?

(4) In addition, do children show any difference in the processing of distinct kinds of Chinese FSs, such as Adjective+Noun collocations versus Verb+Noun collocations, relative to novel phrases?

7.2 Methodology

7.2.1 Materials

The materials were identical[10] to those in Study 1. To ensure that all words appearing in the materials were familiar to the child participants, a checklist was created and sent to the Chinese teachers of the child participants before the self-paced reading experiment. After the experiment, I also checked whether there were any constituent words not familiar to the child

[10] Child corpus was not used for item selection in the current study, as no such corpora are available.

participants. No words were reported (by both the children and their teachers) as unfamiliar to the child participants.

7.2.2 Participants

Three age groups of Mandarin monolinguals took part in the experiment, including two child groups and one adult group. The two child-groups consisted of 36 healthy third graders (23 girls, mean age = 9.12, age ranging from eight years and seven months old to nine years and 10 months old) and 36 healthy sixth graders (15 girls, mean age = 12.01, age ranging from 10 years and 11 months old to 13 years old) from the same primary school in Beijing, China. They had received formal schooling and had been learning to read and write Chinese for approximately 2.5 years and 5.5 years, respectively. The adult group comprised the 39 full-time university students from Study 2 (33 female adults) with an average age of 21.46 (ranging from 18 years old to 28 years old). These adults were considered a control (reference) group for the two child-groups.

Third and sixth graders were examined for the following reasons. The age around nine years is when “Chinese children have mastered enough characters to begin to read well” (Wu et al., 2009, p. 47). Thus, third graders were chosen as the youngest age group capable of participating in a reading experiment.[11] As the highest grade of (Chinese) primary schools, Grade Six was chosen because 12 year olds are considered experienced readers whose reading is “highly similar to that of adult readers” (Li et al., 2013, p. 1581).

All participants were residents of Beijing (China). They received a gift (for the child participants, or an equivalent supermarket voucher for the adult participants) for their participation (funded by the Faculty of Humanities and Social Sciences Research Grant, No. 221046). All participants and guardians of the child participants were informed of their rights. Written consent was obtained from all participants (in the case of the child participants, written consent was also obtained from their parents/guardians) before the experiment. This study was carried out in line with the ethical procedures of the Human Ethics Committee at VUW (Ethics Approval: 0000026508).

[11] According to Shu, Chen, Anderson, Wu, and Xuan (2003), “1686” Chinese characters are introduced to third graders and “2570” characters are taught to sixth graders (p. 31).

7.2.3 Apparatus and procedures

The self-paced reading experiment was carried out individually in a quiet laboratory at a primary school in Beijing. Each participant was first given written instructions and a verbal explanation of the procedure. The demographic information (e.g., age, gender) of participants was then collected. All the procedures were identical to Study 2. This experiment lasted approximately 15 minutes from start to finish. All 111 participants successfully completed the experiment.

7.3 Data analysis and results

7.3.1 Data analysis

The data cleaning procedure was identical to Study 2. All incorrect responses were first excluded from the data. This accounted for 3.20% of the entire data (see Table 18 for more details).

Table 18:

The Exclusions of Third Graders, Sixth Graders, and Adults' Incorrect Responses for the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition, with Different Insertion Lengths

	Original				**Short insertion**				**Long insertion**			
	A+N phrase		*V+N phrase*		*A+N phrase*		*V+N phrase*		*A+N phrase*		*V+N phrase*	
	COL	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**
Grade Three	0.09%	0.08%	0.05%	0.08%	0.11%	0.25%	0.03%	0.11%	0.08%	0.21%	0.08%	0.03%
Grade Six	0.12%	0.10%	0.08%	0.06%	0.11%	0.17%	0.13%	0.06%	0.09%	0.15%	0.06%	0.07%
Adults	0.06%	0.06%	0.02%	0.04%	0.04%	0.20%	0.05%	0.11%	0.01%	0.11%	0.08%	0.04%

Note. The estimated value of each exclusion rate was rounded and kept to four decimal places.

The subsequent data exclusion involved those reaction times that were above or below 3 SDs from the mean of each participant (e.g., Gu, Li, & Liversedge, 2015; Kim & Kim, 2012). This led to 1.67% of data loss (see Table 19 for more details).

Table 19:

The Exclusions of Third Graders, Sixth Graders, and Adults' Irregular Responses for the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition, with Different Insertion Lengths

	Original				**Short insertion**				**Long insertion**			
	A+N phrase		*V+N phrase*		*A+N phrase*		*V+N phrase*		*A+N phrase*		*V+N phrase*	
	COL	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**	**COL**	**CON**
Grade Three	0.05%	0.03%	0.02%	0.06%	0.08%	0.02%	0.02%	0.06%	0.04%	0.02%	0.05%	0.03%
Grade Six	0.03%	0.10%	0.02%	0.07%	0.05%	0.03%	0.02%	0.02%	0.05%	0.04%	0.03%	0.08%
Adults	0.02%	0.05%	0.02%	0.12%	0.08%	0.09%	0.05%	0.04%	0.07%	0.06%	0.04%	0.06%

Note. The estimated value of each exclusion rate was rounded and kept to four decimal places.

Participants had no difficulty answering comprehension questions (mean accuracy rate for Grade Three: 92.59%, Grade Six: 92.65%, adults: 95.41%). Descriptive statistics for the 12 experimental conditions (two phrase types × three insertion lengths × two parts of speech) across the three age groups are presented in Table 20.

A linear mixed-effects model was established to analyse the reaction time measure (dependent variable) in R (version 3.6.1, 2019), package *lme4* (version 1.1-21, 2019). The *p*-value for each predictor (variable) was estimated using the *lmerTest* package (version 3.1-1, 2019). The reaction time was log-transformed before model fitting. The coding method for each categorical variable (i.e., *phrase type*, *insertion length*, and *part of speech*) was identical to Study 1. Besides, *age group* was treated as a discrete and ordinal variable to observe a developmental trajectory of phrasal processing, rather than the comparisons between two child groups and the adult group. To this aim, '3' was assigned to Grade Three, '6' to Grade Six, and '9' to adults.

Table 20:

Third Graders, Sixth Graders, and Adults' Mean Reaction Times (in Milliseconds) for the Adjective+Noun (A+N) Phrases and Verb+Noun (V+N) Phrases in Their Collocation (COL) Condition and Control (CON) Condition with Different Insertion Lengths, with Standard Deviation (SD) in Parentheses

	Original						**Short Insertion**						**Long Insertion**					
	A+N Phrase			*V+N Phrase*			*A+N Phrase*			*V+N Phrase*			*A+N Phrase*			*V+N Phrase*		
	COL	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***	**COL**	**CON**	***p***
Grade Three	897.46 (405.14)	978.50 (419.44)	.01	815.94 (350.04)	918.15 (383.84)	< .001	1126.61 (631.34)	1202.50 (618.81)	.09	1011.52 (542.59)	1144.93 (579.02)	< .01	1272.95 (796.92)	1344.98 (759.85)	.15	1268.18 (686.58)	1331.20 (793.24)	.15
Grade Six	703.25 (330.39)	811.39 (404.45)	< .001	632.71 (307.01)	766.12 (336.70)	< .001	928.38 (581.56)	1032.45 (660.17)	.03	826.66 (505.89)	978.97 (580.66)	< .001	1072.95 (732.78)	1166.22 (750.18)	.08	1123.77 (707.27)	1213.10 (704.84)	.06
Adults	512.51 (202.12)	659.27 (310.32)	< .001	483.84 (199.77)	652.03 (234.48)	< .001	606.60 (307.66)	722.85 (315.16)	< .001	579.53 (302.59)	739.82 (371.36)	< .001	774.70 (535.86)	881.38 (572.56)	.01	738.88 (406.31)	841.13 (489.73)	< .01

Notes. The mean reaction times are the original values (not log-transformed). The estimated value of each mean reaction time was rounded and kept to two decimal places.

The procedure of model building was akin to Study 2. To answer Research Question (1), phrase type was first added to the maximal model as a main predictor. To answer Research Question (2), insertion length and its interaction with phrase type were added to the model. To answer Research Question (3), age group and its two-way interaction with phrase type and the three-way involving age group, phrase type, and insertion length were added to the model. To answer Research Question (4), the two-way interaction between part of speech and phrase type as well as the three-way interaction involving part of speech, phrase type, and insertion length were added to the model. However, due to the convergence issue and based on the suggestion of the *step* function, some non-significant predictors were removed from the maximal model. The removal included all interactions involving part of speech. Model comparisons showed that the removal did not significantly affect the maximal model ($p > .10$).

The random-effects structure included the participant random intercept, item random intercept, and random slopes of the main predictors by participants and by items. The procedures of model fitting and further comparisons/contrasts were identical to the practice adopted in Study 2. The multicollinearity issue was also checked for the model (*Kappa* = 19.17) and predictors (*vif* < 4).

7.3.2 Results

Fixed effects and random effects of the selected model for the reaction times of the children and adults are presented in Table 11. *Phrase type* was not a significant predictor for the reaction times ($t = -0.57$, $p > .10$).

On the contrary, *age group* was a significant predictor ($t = -7.34$, $p < .001$). Post-hoc analysis revealed that sixth graders read faster than third graders ($t = 2.59$, $p = .03$, effect size = 0.59), and adults read faster than sixth graders ($t = 3.88$, $p < .001$, effect size = 0.87).

Table 21:

The Selected Mixed Effects Model for Study 3

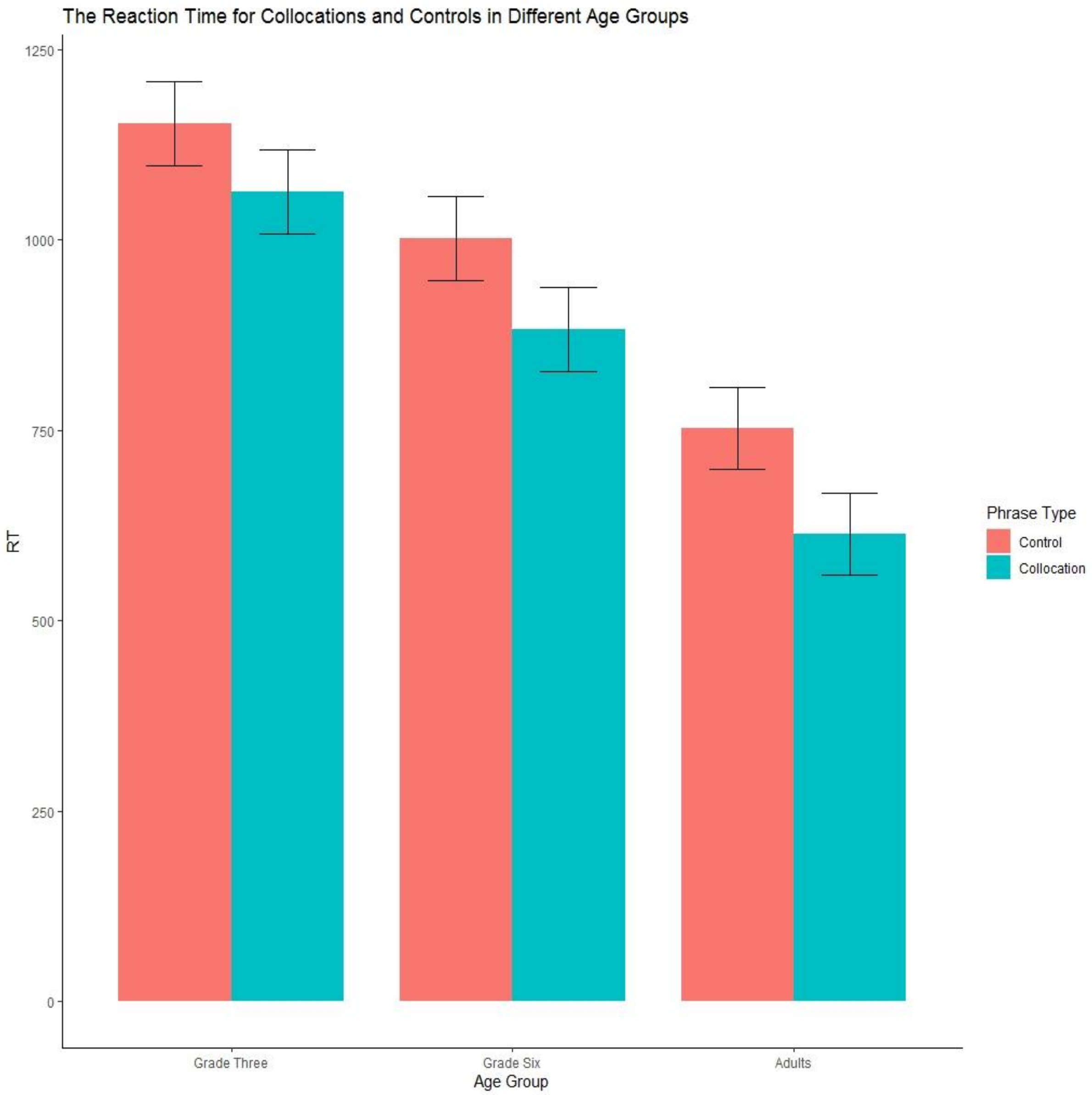
Fixed effects	*Estimate*	*SE*	*df*	*t*	*pr*
Intercept	6.94	0.07	122.10	101.91	< .001
Phrase type	-0.01	0.02	654.30	-0.57	0.57
Insertion length	0.09	0.01	164.20	6.58	< .001
Age group	-0.07	0.01	108.60	-7.34	< .001
Phrase type × Insertion length	0.00	0.01	655.00	0.12	0.91
Phrase type × Age group	-0.01	0.00	9655.00	-6.81	< .001
Insertion length × Age group	-0.00	0.00	107.30	-1.12	0.27
Phrase type × Insertion length × Age group	0.00	0.00	9655.00	1.84	.07
Random effects	*Variance*	*SD*			
Item intercept	0.03	0.17			
Participant intercept	0.07	0.26			
Insertion length | Participant	0.00	0.05			
Residual	0.10	0.32			

Note. The estimated value of each predictor was rounded and kept to two decimal places.

Age group was found to interact with *phrase type* (t = -6.81, p < .001). This interaction is plotted in Figure 8. Further analysis revealed that collocations were processed significantly faster than relative controls across all age groups, and, crucially, the magnitude of the processing advantage for collocations increased with older ages (Grade Three: t = 3.85, p < .001, effect size = 0.28; Grade Six: t = 6.14, p < .001, effect size = 0.45; adults: t = 9.55, p < .001, effect size = 0.69).

Figure 8:

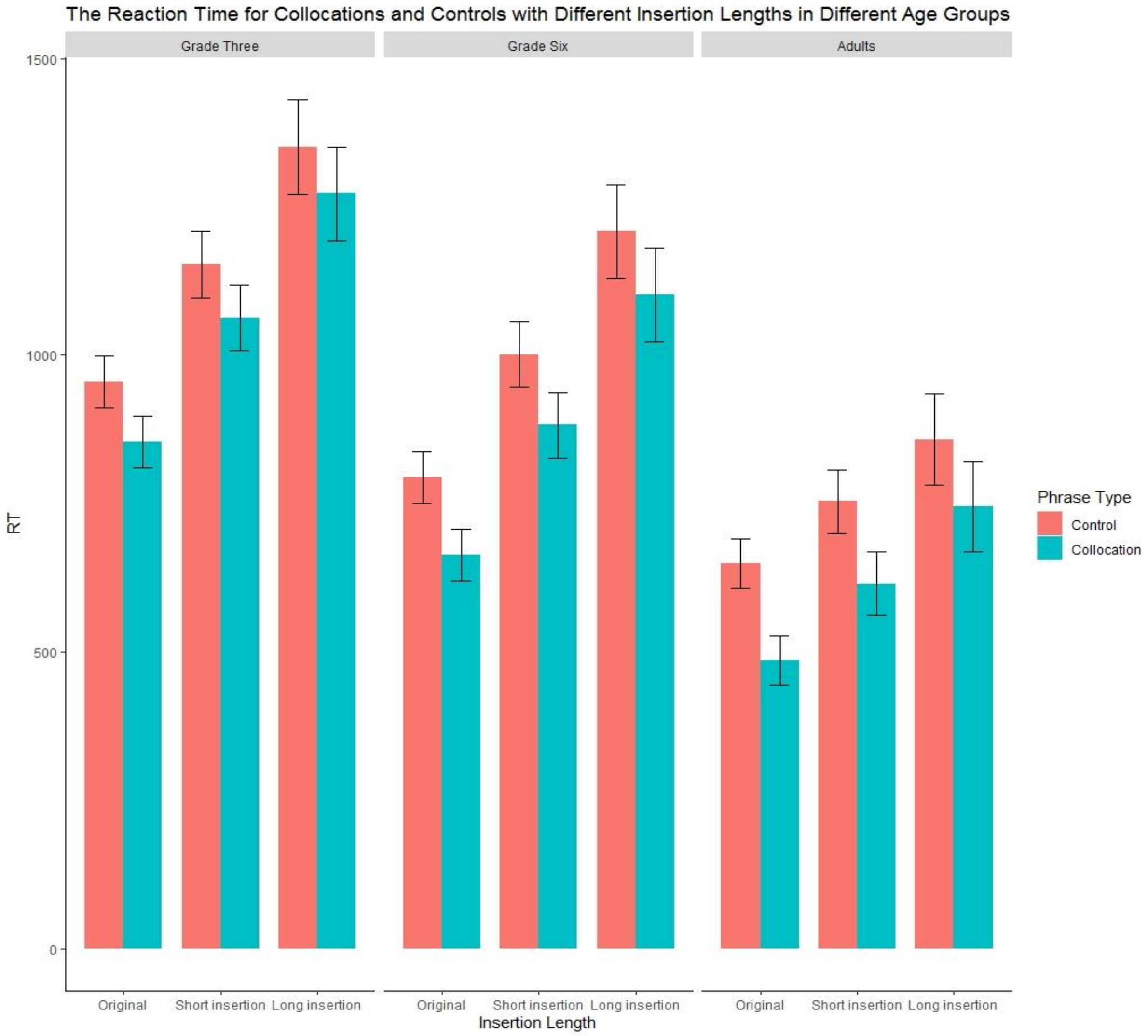
The Interaction between Age Group and Phrase Type, with 95% Confidence Intervals for Each Condition



Crucially, *age group* was also found to interact with *phrase type* and *insertion length* (t = 1.84, p = .07). This interaction is plotted in Figure 9. Further analysis showed that collocations were processed significantly faster than relative controls in the original (Grade Three: t = 3.57, p < .001, effect size = 0.34; Grade Six: t = 5.52, p < .001, effect size = 0.53; adults: t = 9.34, p < .001, effect size = 0.88) and short-insertion phrases across all age groups (Grade Three: t = 2.28, p = .02, effect size = 0.32; Grade Six: t = 3.16, p < .01, effect size = 0.44; adults: t = 5.03, p < .001, effect size = 0.70). The processing advantage for collocations remained significant in the long-insertion phrases among sixth graders and adults (Grade Six: t = 2.56, p = .01, effect size = 0.37; adults: t = 3.42, p < .001, effect size = 0.49), but not in the third graders (Grade Three: t = 1.23, p > .10, effect size = 0.18).

Figure 9:

The Interaction between Age Group, Phrase Type, and Insertion Length, with 95% Confidence Intervals for Each Condition



However, no significant interactions were found between *part of speech* and *phrase type* (*p* > .10), or between *part of speech*, *phrase type*, and *insertion length* (*p* > .10). These insignificant interactions were excluded from the final model due to model fitting.

7.4 Discussion

The processing advantage for FL over novel language has long been evidenced in the literature. However, the relevant evidence pertains almost exclusively to young adults (educated university students in most cases; e.g., Arnon & Snider, 2010; Carrol & Conklin, 2020; Ellis et al., 2009; Janssen & Barber, 2012; Jolsvai et al., 2013; Millar, 2011; Molinaro et al., 2008; Siyanova-Chanturia et al., 2017; Siyanova-Chanturia & Lin, 2018; Swinney & Cutler, 1979; Tremblay et al., 2011; Vespignani et al., 2009), with children being largely disregarded. To the best of my knowledge, only three published studies explored FL processing in children (i.e., Bannard & Matthews, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020; Skarabela et al., 2021 (unpublished)). One of these studies pertains to Chinese FL processing. To be more specific, Jiang et al. (2020) examined how third and fourth graders processed Chinese FSs in the canonical form. Whether or not Chinese FSs in their modified form enjoy a processing advantage over novel phrases has so far been unexplored. In the present investigation, I sought to address this gap and the research questions brought up at the beginning of the current chapter in particular, by asking third and sixth graders, as well as adults (as a reference group), to read Chinese phrases varying in frequency, embedded in sentence context. I recorded the reaction times for these phrases of the readers and, in particular, I examined how *phrase type* (collocation vs control), *degree of modification* (i.e., insertion length, with original vs short-insertion vs long-insertion), *age* (third grader vs sixth grader vs adult), and *part of speech* (Adjective+Noun phrase vs Verb+Noun phrase) might have affected the reading of the Chinese phrases.

Regarding Research Questions (1) and (3): Do children show a processing advantage for Chinese FSs over novel phrases? Does the age of children play a role in the processing of Chinese FSs? The answers are yes. That is, children exhibited a processing advantage for Chinese FSs over novel phrases, and the age of children affected the processing of Chinese FSs. First, I found a significant main effect of *age* for the reading times, with Grade Three readers being the slowest, adult readers being the fastest, and Grade Six readers being in the middle. The degressive reaction times across different age groups suggest a developmental trajectory of reading and underline the development of reading fluency and experience in readers of various ages (e.g., Jiang et al., 2020; Landerl & Wimmer, 2008).

More crucially, I found that *age* significantly interacted with *phrase type* in the analysis of reaction times, such that collocations were processed faster than relative controls by adults ($p < .001$, effect size = 0.69), sixth graders ($p < .001$, effect size = 0.45), as well as third graders ($p < .001$, effect size = 0.28); and, with the increase of age, the magnitude of the processing advantage for collocations increased. This finding reminds us of several existing studies (e.g., Bannard & Matthews, 2008; Jiang et al., 2020). For example, in a phrase elicitation study, Bannard and Matthews (2008) asked two year olds and three year olds with English as their L1, to repeat four-word sequences varying in frequency (e.g., *sit in your chair*, log frequency = 4.26 vs *sit in your truck*, log frequency = 0.00). These sequences were retrieved from the speech exposed to a male child between the ages of two and five registered in the Max Planck Child Language Corpus. Bannard and Matthews (2008) found a significant main effect of sequence frequency, with higher frequency sequences produced faster and more accurately than lower frequency sequences by children, especially the three year olds. However, Bannard and Matthews (2008) did not find a significant interaction between sequence frequency and age. This finding was taken to indicate that the magnitude of the processing advantage for higher frequency sequences over lower frequency sequences kept consistent across ages. The finding in the current study, however, points to a different picture. Although the processing advantage for Chinese collocations emerged both in third and sixth graders, the magnitude of the processing advantage was smaller in the third graders than in the sixth graders. One possible explanation for the different results of the varying magnitude of the processing advantage across ages in Bannard and Matthews (2008) and the current study is that different tasks were used in these two studies (listening+articulating versus reading). In this case, the findings obtained from the two studies may not be directly comparable. Another possible explanation is that the age interval in these two studies is different, with on average, one year (two-year-olds versus three-year-olds) in Bannard and Matthews (2008) and three years in the present study (third graders versus sixth graders). The processing difference between FL and novel language might be larger and more detectable in children with a wide age interval than in those with a narrow age interval. However, these reasons are speculative and need further studies to verify.

Regarding Research Questions (2) and (3): Does the processing advantage for Chinese FSs extend to their modified form in children? If so, does the degree of modification, operationalised as insertion length, affect Chinese FS processing? Does the age of children play a role in the processing of modified Chinese FSs? The answers are also yes. That is, the processing advantage for Chinese FSs persisted in their two modified forms in children, with two and four characters inserted in the FS; but, with the increase of modification degree, the processing advantage diminished; besides, the processing of modified Chinese FSs was affected by children's age. I found that *modification degree* interacted with *phrase type* and *age* in the analysis of reaction times. While the processing advantage for Chinese collocations over relative controls existed in their original form (adults: $p < .001$, effect size = 0.88; sixth graders: $p < .001$, effect size = 0.53), and extended to their short-insertion form (adults: $p < .001$, effect size = 0.70; sixth graders: $p < .01$, effect size = 0.44) and long-insertion form in adults ($p < .001$, effect size = 0.49) and sixth graders ($p = .01$, effect size = 0.37), the processing advantage for the collocations was confined to their original ($p < .001$, effect size = 0.34) and short-insertion ($p = .02$, effect size = 0.32), but not long-insertion forms ($p > .10$, effect size = 0.18) in third graders. Interestingly, with the increase of modification degree, the magnitude of the processing advantage decreased across all age groups. On the contrary, with the increase of age, the magnitude of the processing advantage for collocations increased across all modification versions. These findings suggest that FS processing was greatly affected by age and modification degree.

Regarding Research Question (4): Do children show any difference in the processing of distinct kinds of Chinese FSs, such as Adjective+Noun collocations versus Verb+Noun collocations, relative to novel phrases? The answer might be no, as I did not find any significant interaction between *part of speech* and *phrase type* (and *modification degree*). This finding suggests that the processing of Chinese FSs might not be different between distinct kinds in children. However, this tenet is speculative and requires further exploration using various types of FSs and more sensitive research techniques, such as eye-tracking and EEG.

7.5 Summary

Chapter 7 reports on a self-paced reading study on the processing of modified Chinese FSs in L1 adults and, in particular, children at Grades Three and Six. The analysis of the reading times suggests that akin to adults, children exhibited a processing advantage for Chinese FSs – Adjective+Noun collocations and Verb+Noun collocations – over respective novel phrases. Crucially, the processing advantage persisted in FSs in their modified forms, with two and even four Chinese characters inserted in the middle of the FS; but, with the increase of modification degree, the magnitude of the processing advantage diminished. The age of children was found to play an important role in FL processing: sixth graders always read faster than third graders; more importantly, while sixth graders exhibited a processing advantage for FSs across original and modification forms, the processing advantage was confined to the original and the short-insertion forms in third graders; and, with the increase of age, the magnitude of the processing advantage increased accordingly.

The studies in the present and previous chapters have confirmed the processing advantage for modified FL over novel language in L1 adults, L2 learners, as well as L1 children. In the following chapter, I will discuss what these findings mean to the current literature on FL processing and the knowledge of relevant theories, such as the emergentist approaches.

Chapter 8: General Discussion and Conclusions

8.1 Introduction

Chapters 5 to 7 have provided the answers to the corresponding research questions and interpreted the findings that emerged from each study. In this chapter, I will draw on existing theories, models, and approaches to account for the main findings and will talk about what these findings indicate to language users in the real world. The current chapter will by and large cover three topics: (1) the processing of modified formulaic language (FL) versus novel language in L1 adults; (2) the processing of modified FL versus novel language in L2 learners; and (3) the processing of modified FL versus novel language in children. Then, I will talk about the general contributions of the current thesis. At the end of this chapter, I will point out a number of limitations and, importantly, will discuss possible directions for future research, followed by the conclusion of the current thesis.

8.2 The processing of modified formulaic language vs novel language in L1 adults, L2 learners, and L1 children

8.2.1 The processing of modified formulaic language vs novel language in L1 adults

In the studies reported in the present thesis, L1 Chinese users exhibited a processing advantage for FL (e.g., 坐飞机 '*take the plane*') over novel language (e.g., 修飞机 '*fix the plane*'). This finding is in line with existing research involving adults (e.g., Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Carrol & Conklin, 2020; Janssen & Barber, 2012; Jiang et al., 2020; Millar, 2011; Siyanova-Chanturia et al., 2011a, 2011b; Sonbul, 2014; Tremblay et al., 2007, 2011; Underwood et al., 2004; Yi et al., 2017). For example, Carrol and Conklin (2020), Millar (2011), Siyanova-Chanturia et al. (2011b), Sonbul (2014), Tremblay et al. (2011), Underwood et al. (2004), and Yi et al. (2017) compared the processing of frequent FSs (e.g., *salt and pepper*; *ideal partner*; *in any case*) with infrequent novel phrases in L1 readers (e.g., *pepper and salt*; *best partner*; *in your case*). All of these studies demonstrated that FL enjoys a processing advantage over novel language in L1 adults, with FSs processed faster and more accurately and effortlessly than novel phrases. The processing advantage for FL indicates that it exists in the mental lexicon of L1 users. The psychological reality of FL further points to an important role of frequency, exposure, or experience in L1 processing, learning, and use.

Crucially, the processing advantage for FL over novel language has also been observed in their modified forms, with two (i.e., short-insertion; e.g., 坐***一次***飞机 ‘*take the plane **for once***’ vs 修***一次***飞机 ‘*fix the plane **for once***’) and even four Chinese characters inserted in the middle of the phrase (i.e., long-insertion; e.g., 坐***那架真的***飞机 ‘*take **that real** plane*’ vs 修***那架真的***飞机 ‘*fix **that real** plane*’). This finding is in line with existing research (e.g., Geeraert et al., 2017; Kyriacou et al., 2020; Molinaro et al., 2013; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). For example, in Vilkaitė (2016a) and Vilkaitė and Schmitt (2019), L1 English users were asked to read sentences embedded with Verb+Noun collocations and relative controls in their original and modified forms, with three words inserted in the middle of the phrase (e.g., *provide (some of the) information* vs *compare (some of the) information*). The eye movements of readers in Vilkaitė (2016a), Vilkaitė and Schmitt (2019), as well as the current study suggest that FL enjoys a processing advantage over novel language even when substantially modified. This finding implies that at least for literal FL, modification can not destroy the psychological reality of FL in L1 users' mental lexicon.

While the present finding is consistent with most earlier studies pertaining to modified FL processing (e.g., Geeraert et al., 2017; Kyriacou et al., 2020; Molinaro et al., 2013; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019), it is somewhat inconsistent with McGlone et al. (1994) and Haeuser et al. (2020). In McGlone et al. (1994) and Haeuser et al. (2020), a group of L1 English adults were asked to read sentences embedded with ambiguous idioms and relative controls in their original and modified forms. These researchers found that, while idioms had a processing advantage over novel phrases in their original form (e.g., *Larry **bit the bullet** and bought diamond earrings for his wife's birthday* vs *Larry **hid the bullet** so the police would not find the crucial evidence*), the processing advantage disappeared in their modified form (e.g., *Larry **bit the iron bullet** and bought diamond earrings for his wife's birthday* vs *Larry **hid the iron bullet** so the police would not find the crucial evidence*). One possible explanation for the different results concerning the processing advantage for modified FL in the earlier (McGlone et al., 1994; Haeuser et al., 2020) and current research is that different types of FSs were examined, with figurative FSs used in McGlone et al. (1994) and Haeuser et al. (2020), whereas literal FSs used in the current investigation. As Schmitt (2005) noted, FL is more fixed than novel language, and the degree of fixedness differs greatly among various FS types. Compared with literal FSs, such as collocations (e.g., *provide (some of the) information*), n-grams (e.g., *in (clear) contrast to*), figurative FSs, such as idioms (e.g., *hit the nail on the head*), are more fixed in meaning (e.g., Sinclair, 1991;

Wray, 2002). Therefore, modification of meaning, such as insertion, may pose more threat to the formulaicity of figurative FSs relative to literal FSs, which may further result in the sensitivity to figurative FSs decreasing or even disappearing in their modified form(s). On the contrary, literal FSs are less fixed and allow for more modifications, in comparison with figurative FSs. Thus, modification of meaning, such as insertion, could not entirely jeopardise the formulaicity of literal FSs, instead, enrich their original meaning (e.g., Molinaro et al., 2013). In this case, the processing advantage for literal FSs does not disappear but persists in their modified form. The different magnitudes of processing advantage for modified literal FL versus figurative FL indicate that at least in the case of insertion, different FL types are affected by modification to various degrees: while insertion jeopardises our sensitivity to figurative FL to a larger degree, it poses minor threat to literal FL.

It is noteworthy that the degree of modification affected FL processing, and the magnitude of the processing advantage for FL was not consistent across different modification degrees. This finding is novel and not consistent with earlier research (e.g., Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). In the earlier studies, Vilkaitė (2016a) and Vilkaitė and Schmitt (2019) asked L1 English readers to read sentences embedded with Verb+Noun collocations and relative controls both in their original and modified forms. For the modified form, Vilkaitė (2016a) and Vilkaitė and Schmitt (2019) inserted three words in the middle of each collocation and its relative control, which was the only level of modification degree. These researchers found the processing advantage for FL both in its original and modified forms and the magnitude of the processing advantage was consistent across the two forms.

However, the findings in the current investigation point to a different picture. The current investigation not only compared the processing of original and modified FSs with relative novel phrases as what was done in Vilkaitė (2016a) and Vilkaitė and Schmitt (2019) but also took different modification degrees into consideration. Two modification degrees were examined: the short-insertion condition (with two Chinese characters inserted in the middle of the phrase) and the long-insertion condition (with four Chinese characters inserted in the middle of the phrase). Crucially, the current finding shows that, although the processing advantage for FL extended to its two modified forms, the magnitude of the processing advantage was not consistent with it in the original form; and, with the increase of modification degree, the magnitude of the processing advantage diminished. This finding suggests that, apart from frequency, modification degree also plays a role in L1 phrasal processing. With the increase of modification degree, L1 users' sensitivity to phrase frequencies diminishes, but importantly, it is still observable.

8.2.2 The processing of modified formulaic language vs novel language in L2 learners

In the current investigation, in addition to L1 users, L2 learners exhibited a processing advantage for FL over novel language. This finding is in line with existing research (e.g., Beck & Webber, 2016a; Conklin & Schmitt, 2008; Edmonds, 2014; Ellis et al., 2008; Hernández et al., 2016; Jiang & Nekrasova, 2007; Kim & Kim, 2012; Northbrook & Conklin, 2019; Siyanova-Chanturia et al., 2011b; Sonbul, 2014; Sonbul & El-Dakhs, 2020; Underwood et al., 2004; Vilkaitė & Schmitt, 2019; Yi et al., 2017; Yi, 2018). For example, Siyanova-Chanturia et al. (2011b), Sonbul and El-Dakhs (2020), and Yi et al. (2017) found that frequent L2 FSs (e.g., *bride and groom*; *I don't know why*; *point out*) were read faster relative to infrequent novel phrases (e.g., *groom and bride*; *I don't know who*; *test out*), either in isolation or sentence context. This finding suggests that FL enjoys a processing advantage over novel language not only in L1 but also in L2. It is noteworthy that the FL advantage was not exclusive to higher proficiency learners (e.g., Edmonds, 2014; Jiang & Nekrasova, 2007; Siyanova-Chanturia et al., 2011b; Siyanova-Chanturia & Janssen, 2018; Yi et al., 2017), but also observed in lower proficiency learners (e.g., Hernández et al., 2016; Isobe, 2011; Northbrook & Conklin, 2019; Sonbul & El-Dakhs, 2020; Wolter & Yamashita, 2018), as shown in the present investigation. The processing advantage for L2 FL demonstrates that akin to FL in L1, it exists in the mental lexicon of L2 learners. The psychological reality of L2 FL further points to an important role of frequency, exposure, or experience in L2 processing, learning, and use.

Importantly, the processing advantage for L2 FL also manifested itself in the two modified forms. This is a novel finding as the facilitation was not observed in earlier research (Vilkaitė & Schmitt, 2019). In Vilkaitė and Schmitt (2019), the processing advantage for L2 collocations over relative controls existed only in their original form, and disappeared in their modified form, with three words inserted in the middle of the phrase. The finding of the current investigation points to a different picture. The processing advantage for L2 FL emerged in its original form extended to its modified forms, with two (i.e., short insertion) and even four Chinese characters inserted in the middle of the phrase (i.e., long insertion), respectively. This finding suggests that akin to FL in L1, after L2 FL is acquired, it becomes entrenched in the mental lexicon of learners, and this fact remains even if FL is modified.

It is noteworthy that the magnitude of the processing advantage for modified L2 FL over novel language was not equal across different modification degrees (i.e., short insertion vs long insertion). With the increase of modification degree, the magnitude of the processing advantage decreased. This finding is different from what was observed in Vilkaitė (2016a), in which the magnitude of the processing advantage was found consistent across the original and modified forms of English FSs in L1. On the contrary, the variant magnitudes of FL advantage across different modification degrees in L2 learners is consistent with what was found in the current study on modified L1 FL processing. This finding suggests that, akin to modification degree in L1 FL processing, this factor also affects the processing of FL in L2. With the increase of modification degree, L2 learners' sensitivity to FL and phrase frequencies is not consistent but weakened. In this case, although modified L2 FL enjoys a processing advantage over novel language, this facilitation decreases with major modification.

Although the downward trend of the processing advantage for L1 and L2 FL over novel language was analogous across the original and modified forms, the magnitude of the FL facilitation was found to be smaller in L2 learners than in L1 users. This finding is not unexpected and in line with existing research (e.g., Edmonds, 2014; Gyllstad & Wolter, 2016; Kim & Kim, 2012; Underwood et al., 2004; Vilkaitė & Schmitt, 2019). Compared with L1 users, L2 learners had less exposure to and experience with the target language, which weakened their sensitivity to FL and frequencies during phrasal processing. This finding suggests that learners' sensitivity to FL and frequencies is affected by language proficiency, which develops with the increase of experience. This finding further offers support to the emergentist approaches that experience plays a crucial role in L2 processing, learning, and use (e.g., Conklin, 2020; Van Lancker Sidtis, 2015; Wiechmann et al., 2013).

8.2.3 The processing of modified formulaic language vs novel language in children

In addition to L1 and L2 adults, children at Grades Three and Six exhibited a processing advantage for FL over novel language. This finding is in line with Skarabela et al. (2021), Bannard and Matthews (2008), Hallin and Van Lancker (2017), and Jiang et al. (2020). For example, in an eye-tracking study, Jiang et al. (2020) asked third and fourth graders, who were monolingual speakers of Chinese, to read sentences embedded with Verb+Noun collocations (e.g., 参加会议 '*attend a meeting*') and relatively infrequent controls (e.g., 参加游戏 '*attend a game*'). Jiang et al. (2020) found collocations were read faster than control phrases by not only adults but also fourth graders. The present study confirmed this finding on Chinese adults and sixth graders, and extended it to third graders. The processing advantage for FL over novel language illustrates that FL exists in the mental lexicon of not only adults but also children, which helps them reliably distinguish between higher- and lower-frequency phrases. The psychological reality of FL further points to the important role of frequency, exposure, or experience in language processing, learning, and use in children. As shown in the current study, children as young as third graders have already established the sensitivity to distributional properties of input information greatly exposed to them in daily life (e.g., Bates et al., 1988; Clark, 1974; Foster-Cohen, 1999; Goldberg, 2006; Hakuta, 1974; Lieven et al., 1992, 2003, 2009; Nelson, 1973; Peters, 1977).

Although FL advantage emerged in third graders in the present study, it did not manifest in the third graders of Jiang et al. (2020). One explanation for the different findings is that different materials were used in the two studies. Compared with the FSs in Jiang et al. (2020; e.g., *参加会议* '*attend a meeting*'), the FSs in the current study (e.g., *坐飞机* '*take the plane*') might be more familiar to third graders, thus third graders were more sensitive to the FSs in the current study than the FSs in Jiang et al. (2020). Another explanation is that different operationalised definitions of 'third graders' were used in the two studies: while the third graders recruited by the current study were at late Grade Three (mean age = 9 years and 1 month), the third graders in Jiang et al. (2020) were at early Grade Three (mean age = 8 years and 7 months). The third graders in the current study were, in fact, older than the third graders in Jiang et al. (2020), thus the former group of third graders could be more experienced than the latter group of third graders.

Crucially, akin to adults, children exhibited a processing advantage for modified FL. This is a novel finding not previously reported in the literature (Bannard & Matthews, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020; Skarabela et al., 2021). This finding suggests that after FL is acquired, it becomes entrenched in children's mental lexicon; and, whether it is in original or modified form, children are able to recognize it with ease.

Furthermore, akin to the FL facilitation in adults, the magnitude of FL advantage in children was not invariant across modification degrees (i.e., short-insertion vs long-insertion). Instead, children showed a downward trend of the processing advantage for FL, as modification degree increases. This finding suggests that, although modification did not entirely disrupt the entrenchment of FL in children's mental lexicon, it nevertheless weakened their sensitivity to FL, with the increase of modification degree.

Although children showed a similar downward trend of FL advantage across original and modified forms as adults did, the magnitude of the processing advantage was found to be larger in adults than in children. This finding is in line with existing research (e.g., Hallin & Van Lancker, 2017; Jiang et al., 2020). Compared with L1 adults, children had relatively limited exposure to target FSs, which decreased their sensitivity to FL in its original and modified forms. This finding suggests that at least in reading modality, children' sensitivity to the frequency of input increases with age (e.g., Arnon, 2019; Raviv & Arnon, 2017; Shufaniya & Arnon, 2018).

Interestingly, as learners of the Chinese language, L1 children and L2 adults showed some similarities and differences in the processing of FL in its original and modified forms. First, although children read phrases faster than L2 adults across the board (e.g., Arnon & Christiansen, 2017; Christiansen & Arnon, 2017), the two populations showed a comparable pattern of FL processing. Both populations exhibited a processing advantage for FL over novel language, across the original and modified forms; and with the increase of modification degree, the magnitude of the processing advantage was not consistent but diminished across all forms in these two populations. This finding suggests that although FL is entrenched in learners' mental lexicon, FL advantage can be weakened with the increase of modification degree in both L1 and L2.

Second, while the processing advantage for modified FL manifested itself in both short- and long-insertion forms in lower proficiency L2 learners, the FL advantage was observed in short- but not long-insertion form in younger L1 learners.

Third, with the development of language proficiency, L1 and L2 learners showed a different trend for the magnitude of FL advantage across the original and modified forms. For L1 children, as their age increased, the magnitude of the processing advantage increased. This finding is not surprising (e.g., Bannard & Matthews, 2008; Jiang et al., 2020). However, this was not the case in L2 adults. With the development of language proficiency, although the general processing speed increased, the magnitude of L2 FL advantage did not increase, but by and large, remained consistent. This finding is somewhat different from what was observed in Matlock and Heredia (2012) and Siyanova-Chanturia et al. (2011b), in which the magnitude of the processing advantage was larger in higher-proficiency learners than in lower-proficiency learners. One possible explanation is that the lower proficiency L2 learners in the current investigation might be more experienced than those in the existing studies (Matlock & Heredia, 2012; Siyanova-Chanturia et al., 2011b), such that their sensitivity to FL and phrase frequencies could be similar to the higher proficiency learners. However, this explanation is tentative and requires further exploration in L2 learners with a wider range of language proficiency. Another possible reason is that different measures of L2 proficiency were used in the current and existing research, such that the performances of the lower/higher proficiency learners in the current and existing studies might not be directly comparable.

8.3 Theoretical account that can explain the current findings

In line with existing research (e.g., Arnon & Snider, 2010; Bannard & Matthews, 2008; Hernández et al., 2016; Jiang et al., 2020; Kyriacou et al., 2020; Siyanova-Chanturia et al., 2011b; Sonbul & El-Dakhs, 2020; Vilkaitė, 2016a; Yi et al., 2017), I take the processing advantage for Chinese FSs over novel phrases in both original and modified forms in L1 adults, L2 learners, and L1 children, to reflect the role of frequency of linguistic units above the word level, viz. phrase frequency. The phrase frequency effects observed in the current thesis can be explained by the usage-based theory (e.g., Bybee, 1998, 2006, 2011; Bybee & Beckner, 2012; Bybee & Hopper, 2001; Ellis, 2002a, 2017; Ellis & Ogden, 2017; Ellis et al., 2016; Goldberg, 2006; Langacker, 1987; Tomasello, 2003).

The usage-based theory puts emphasis on the role of frequency in the processing of language of all shapes and sizes (e.g., morphemes, words, phrases, and even longer stretches of language), regardless of whether the language unit is analysable or not. This theory holds that frequent occurrence entrenches a linguistic unit in our memory and facilitates its relative learning, processing, and use. Language learners at young ages or of lower proficiency are able to identify and collect distributional information of an expression, such as frequency, despite the fact that their cognitive skills or linguistic knowledge are relatively limited (e.g., Arnon, 2011; Arnon & Clark, 2011; Christiansen, 2019; Lieven et al., 1992; Locke, 1997; Nelson, 1981; Saffran, 2003; Tomasello, 2003; Wray, 2002). Learners automatically register each exposure to an expression in their memory (e.g., Arnon, 2019; Christiansen, 2019; Frost et al., 2015; Jurafsky, 1996, 2003; McDonald & Shillcock, 2003a, 2003b; Seidenberg & MacDonald, 1999), while frequent encounters make linguistic units stored in learners' memory (e.g., Bybee & Beckner, 2012; Ellis, 2002a, 2017; Goldberg, 2006; Langacker, 1987; Tomasello, 2003). Subsequent exposures to these stored units further entrench them in one's memory and strengthen the connections between their components.

According to the usage-based theory, FL (e.g., *白衬衫 (white shirt)*), whose components co-occur more frequently, is processed faster and requires less cognitive effort than novel language (e.g., *黄衬衫 (yellow shirt)*), whose components co-occur less frequently. This is because compared with novel language, FL is more entrenched in one's memory and the connections between its components are stronger. The greater entrenchment and stronger connections help language users and learners access FL and its components at a faster speed, with less effort to select, compute, or analyse each component, relative to novel language and its components. In this case, the processing of FL is facilitated and is more efficient than that of novel language.

In a similar vein, although FL can be modified and stretched, it is still processed faster than novel language. In a modified FS (e.g., *白色长衬衫 (long white shirt); 白格子长袖衬衫 (white-plaid long-sleeve shirt)*), components may be forced apart, with formulaicity of the sequence jeopardised. In this case, when FL is being processed in its modified form, syntactic analysis and computation can not be circumvented, which is somewhat inconsistent with Bateson (1975), Jespersen (1976), and Wray (2002). This means the processing of modified FL may require more cognitive effort, such as selecting, computing, and analysing, which may render it similar to novel language processing. Despite the inevitable syntactic analysis, modified FL is processed faster than novel language processing (e.g., Conklin & Carrol, 2020; Kyriacou et al., 2020; Vilkaitė, 2016a). This is because compared with novel language, FL is more frequent and is thus more entrenched in our memory, with stronger connections between its components. The entrenchment of FL helps us detect the similarities between the modified and original forms and map the modified form onto the canonical form that is already stored in our memory. Furthermore, the strong connections between FL's components provide us with ready and rapid access to them, even though they are not adjacent.

In comparison with FL, novel language is relatively infrequent and is thus less entrenched in our memory and the connections between its components are relatively weak and loose. Unlike FL processing, when processing novel language either in its original (e.g., *黄衬衫 (yellow shirt)*) or modified forms (e.g., *黄色长衬衫 (long yellow shirt); 黄格子长袖衬衫 (yellow-plaid long-sleeve shirt)*), language users and learners have to go through a process of analysis, selection, and computation in each encounter. In this process, a phrase is regarded as a series of slots to be filled in with words (e.g., Sinclair, 1991). The filling is open to an infinite number of available choices. One can only seek help from grammar rules to decide which word could be the most appropriate one to choose for the filling. Nevertheless, this process is serial and progressive, thus, always laborious.

On the contrary, the processing of FL does not require the entire selection process. This is because the entrenchment of FL in memory and strong connections between its components make “all the slot-by-slot choices massively reduced in scope or even, in some cases, pre-empted” (Sinclair, 1991, p. 110; see also Vespignani et al., 2009). Even though FL is stretched with its components knocked apart, the connections persist (e.g., Bolinger, 1979). The strong connections further render the selection of FL’s components, at least of the final component, to be largely accelerated and facilitated, or even, unnecessary (e.g., Bybee & Beckner, 2012; Siyanova-Chanturia & Pellicer-Sánchez, 2019). As a consequence, the processing of FL, whether in its canonical or modified form, is always more efficient than the processing of novel language.

8.4 Contributions to the current literature

The present thesis has the potential to contribute to the current literature and body of knowledge in several ways.

First, this thesis reaffirms the psychological reality of FL in language users’ and learners’ mental lexicon and the sensitivity to frequencies above the word level. Crucially, this is the case even when FL is modified. While FL in its original form has been consistently shown to exhibit a significant processing advantage over novel language in L1 adults (e.g., Arnon & Snider, 2010; Carrol & Conklin, 2020; Millar, 2011; Siyanova-Chanturia et al., 2017; Tremblay et al., 2011), L2 learners (e.g., Hernández et al., 2016; Jiang & Nekrasova, 2007; Siyanova-Chanturia et al., 2011b; Matlock & Heredia, 2002; Yi et al., 2017), and L1 children (e.g., Bannard & Matthew, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020; Skarabela et al., 2021), very little evidence exists with regard to modified FL. The present studies extend the current literature by showing that, akin to FL in its original form, modified FL enjoys a significant processing advantage over novel language in L1 adults, L2 learners, as well as L1 children. These findings once again point to the crucial role of frequency and exposure (experience) in L1 and L2 learning, processing, and use.

Second, the current studies are the first in the literature to explore how different modification degrees affect FL processing. Before the current studies, existing research focused on modified FL with only one level of modification degree (Geeraert et al., 2017; Haeuser et al., 2020; Kyriacou et al., 2020; McGlone et al., 1994; Molinaro et al., 2013; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019). The current studies sought to address this gap by comparing the processing of modified FL with two levels of modification degree: the short-insertion condition versus the long-insertion condition. The current studies revealed that although FL advantage remained in the two modified forms, the magnitude of FL advantage was not invariant but diminished with the increase of modification degree. Despite this, the processing advantage for FL over novel language persisted in the two modified forms not only in L1 adults but also in children and L2 learners.

Third, the current investigation is the first in the literature to explore the processing of modified FL in lower proficiency L2 learners. Existing research has largely looked at modified FL processing in highly proficient learners, with those at lower proficiency disregarded (Vilkaitė & Schmitt, 2019). The current study sought to address this gap by observing a wider range of L2 learners. Results of the current study indicate that not only higher-proficiency learners but also those of lower-proficiency exhibited a processing advantage for modified FL and sensitivity to phrase frequencies.

Fourth, the current investigation is one of the very few studies on FL processing in children, and the first in the literature to explore the processing of modified FL in children. While children have been found to show a processing advantage for FL in its original form (Bannard & Matthew, 2008; Hallin & Van Lancker, 2017; Jiang et al., 2020; Skarabela et al., 2021), whether or not the FL advantage remained in its modified form had not been studied. The current study sought to address this gap by observing third and sixth graders when reading sentences embedded with modified FSs versus novel phrases. The results show that both age groups exhibited a processing advantage for modified FL and sensitivity to phrase frequencies.

Fifth, the current studies contribute to the body of knowledge of frequency effects in Chinese. Earlier research suggested that L1 users are sensitive to Chinese character frequencies (e.g., Lin, Angele, Hua, Shen, Zhou, & Li, 2018; Ma et al., 2015; Yan, Tian, Bai, & Rayner, 2006) and word frequencies (e.g., Liu, Li, & Han, 2015; Liu, Wang, Yan, Paterson, & Pagán, 2020; Rayner, Li, Juhasz, & Yan, 2005). Recent research has extended frequency effects to Chinese phrasal processing (e.g., Kong et al., 2016; Liu et al., 2010; Shen et al., 2017; Yi et al., 2017; Yu et al., 2016; Zhang et al., 2013; Zhou et al., 2004). However, the recent research has largely focused on figurative Chinese FSs, which are relatively infrequent. The current studies used literal FSs, which are more frequent than figurative sequences, and reconfirmed the phrase frequency effect. Crucially, the current studies extended this finding to phrases in their modified forms, with up to four Chinese characters intervening between the component words; and with the increase of inserted characters, the magnitude of the phrase frequency effect did not entirely disappear but diminished.

Finally, from the methodological standpoint, the current studies evidence how the use of eye-tracking and self-paced reading methodologies can help us deepen and broaden our understanding of the mechanisms behind FL processing. In terms of the eye-tracking methodology, the early and late measures, along with different interest areas (the whole phrase and its final word), enabled me to paint a more detailed and complex picture of different reading patterns during FL processing. While FSs were read significantly faster than relative controls across the analyses of both interest areas, this processing advantage was largely confined to late measures. This finding again supports the argument that late measures, which indicate the reanalysis of text information and more controlled and strategic processing, may be more sensitive to phrase frequencies and the potential processing differences between FL and novel language than early measures, which indicate initial lexical access and automatic processing (e.g., Jiang et al., 2020; Siyanova-Chanturia et al., 2011a). Further, although the processing advantage for FSs over novel phrases was observed across the original and modified forms in late measures in the whole phrase analysis, this was never the case in the phrase-final word. This finding again underlines the necessity for a variety of interest areas to be observed in FL processing investigations, in that the processing of different target areas could vary.

With respect to the self-paced reading methodology, although it may be less sensitive than the eye-tracking technique, it nonetheless helped me detect the processing differences between FSs and novel phrases in their original and modified forms. Moreover, self-paced reading helped me pinpoint the differences in FL processing between L1 users at different ages, as well as between L2 learners from distinct L1 backgrounds and of different language proficiencies. It is worth noting that the current studies are the first in the literature to use the self-paced reading methodology to explore modified FL processing in children and L2 learners from distinct L1 backgrounds and of different L2 proficiencies. Interestingly, if we compare the reaction times of L1 adults in the current self-paced reading studies and the first-pass reading times for the whole phrase of L1 adults in the present eye-tracking study, we see that the reaction times are approximately 1.50 times longer than the first-pass reading times. This illustrates that the reading rate in self-paced reading is slower than in eye-tracking. This is because the task of self-paced reading (i.e., pressing a button during reading) slows down the natural reading time. This finding offers support to Rayner's (1998) argument that the reading rate of self-paced reading technique is "about half as fast as it is normally" (p. 391).

In this case, although the results obtained from self-paced reading and eye-tracking are comparable, self-paced reading is less accurate than eye-tracking.

8.5 Limitations, future directions, and conclusions

While the present thesis contributes to the literature, it has several limitations. For example, due to the unavailability of an eye-tracker, Study 2 (modified FL processing in L2 learners) and Study 3 (modified FL processing in children) employed the self-paced reading technique to collect data. This research method might be considered to be less sensitive than eye-tracking (e.g., Godfroid et al., 2020; Rayner, 1998). Despite this, the current thesis is still the first in the literature to explore modified FL processing and confirm the processing advantage for modified FL in children and L2 learners.

In addition, in the current investigation into L2 learners, due to the unavailability of a reliable proficiency test, I used only one measure, self-rating, to define their language proficiency, which might be deemed subjective. It would have better to use more objective proficiency measures, such as vocabulary size tests, to define proficiency in L2 research.

Importantly, the present work has highlighted a number of directions for future research. First, the current thesis confirmed the processing advantage for modified FL by examining FSs with only one type of modification (i.e., insertion) at two levels of modification degree (i.e., short-insertion vs long-insertion). As suggested in earlier studies (e.g., Geeraert et al., 2017; Haeuser et al., 2020; Kyriacou et al., 2020; McGlone et al., 1994; Molinaro et al., 2013; Vilkaitė, 2016a; Vilkaitė & Schmitt, 2019), different modification types can lead to different results for FL processing. For example, in two recent studies on modified L1 idiom processing, Haeuser et al. (2020) focused on idioms with insertion (e.g., *bite the bullet* vs *bite the iron bullet*), while Kyriacou et al. (2020) looked at idioms with inflection (e.g., *kick the bucket* vs *the bucket was kicked*). Although both studies showed a processing advantage for idioms over novel phrases in their original form, the processing advantage extended to modified idioms with inflection (Kyriacou et al., 2020), not to modified idioms with insertion (Haeuser, et al., 2020). In addition, different modification degrees can lead to different results for FL processing. For example, in the current investigation into third graders, the processing advantage for FL over novel language was found in the original and short-insertion forms (with two meaningful Chinese characters inserted in the middle of the phrase) but disappeared in the long-insertion form (with four meaningful Chinese characters inserted in the middle of the phrase). Therefore, it could be problematic to extend the current findings to FL with other modification types (such as with inflection, with negation) or to FL with other modification degrees (e.g., inserted with three characters vs inserted with five characters), without attestation. Future research may thus consider the processing of FL with other types of modification, such as inflection, negation, transposition, or truncation, and explore FL with a wider range of modification degrees, such as inserted with three characters versus inserted with five characters.

Second, in the current investigation into children, in order to reduce the complexity of the experimental design, I employed three age groups, third graders (eight-nine years old), sixth graders (11–12 years old), and university students (19–30 years old, as a reference group). According to Wray (2002), the sensitivity to FL is a dynamic developmental process, and FL processing in L1 could be different among various age groups. In this case, it could be problematic to generalise the results obtained from the current investigation to any other age groups, without verification. In order to forge a more complete understanding of the developmental trajectory of L1 FL processing, future research could employ a wider range of age, for example, pre-schoolers versus school-age children versus adolescents versus the elderly (e.g., Haeuser et al., 2020), to capture the developmental trajectory across the lifespan.

Third, although the current thesis has confirmed the processing advantage for modified FL in L2 learners of different language proficiencies, it is still important to know how modified L2 FL is processed in learners in a wider range of language proficiency. This is because, as noted by Wray (2002), the sensitivity to L2 FL is not invariant across proficiency and the developmental trajectory of L2 FL processing can be distinctly different from it in L1 FL processing. For example, in Matlock and Heredia (2012) and Siyanova-Chanturia et al. (2011b), higher-proficiency learners were faster, and more importantly, more sensitive to FL and phrase frequencies than lower-proficiency learners. In addition, as shown in the current thesis, although the sensitivity to FL increased greatly with age in L1 children, it developed at a relatively slow pace with the increase of L2 proficiency. Therefore, in order to unpack the complete developmental trajectory of L2 FL processing in a fine-grained way, it is worthwhile for future research to observe learners in a wider proficiency range, especially those at lower proficiency levels.

Fourth, in the current thesis, I examined modified FL processing in healthy populations. Whether and how people with disorders, such as Alzheimer's Disease, Parkinson's Disease, Autism, Agenesis of the Corpus Callosum (a brain disorder), Schizophrenia (a severe mental illness), Tourette's Syndrome (vocalizations of taboo words and phrases indigenous to a culture; see e.g., Van Lancker & Cummings, 1999), Dysprosody (a speech disorder) process modified FL is worthy of exploration (e.g., Van Lancker & Sidtis, 2012, 2018a, 2018b; Van Lancker-Sidtis, Choi, Alken, & Sidtis, 2015; Van Lancker-Sidtis, Pachana, Cummings, & Sidtis, 2006). According to Van Lancker and colleagues' pioneering work (e.g., Van Lancker & Sidtis, 2012, 2018a, 2018b; Van Lancker-Sidtis et al., 2006, 2015), although people with disorders have difficulties producing novel language, these populations can still produce a high proportion of formulaic expressions. This is because the production of FL and novel language is dependent on distinct brain areas, with FL production relying on the right hemisphere, whereas novel language production relying largely on the left hemisphere. Therefore, it is warranted for future research to study modified FL processing in people with disorders. Future research could adopt neurolinguistic methodologies, such as Positron Emission Tomography (PET), functional Magnetic Resonance Imaging (fMRI), to observe the brain of people with disorders when they are processing modified FL versus novel language.

Finally, in the current studies, I only focused on collocations. As noted in Jiang et al. (2020), different kinds of FSs "are likely to differ in how reading development happens in childhood" (p. 26). In fact, recent research has demonstrated that different kinds of FSs are not always processed equally and are driven by distinct linguistic factors. For example, in a recent eye-tracking study, Carrol and Conklin (2020) compared the processing of three types of FSs – idioms versus binomials versus collocations. Although a processing advantage emerged in all FS types relative to their infrequent novel phrases, Carrol and Conklin (2020) found different additional factors that contributed to the processing advantage for the three FS types. While idiom processing was driven by familiarity and decomposability, binomial processing tended to be subjective to predictability and semantic association, and collocation processing was more dependent on mutual information than the other two types. Therefore, future research could attempt to compare the processing of modified FL between distinct kinds, such as modified collocations versus n-grams versus idioms.

In conclusion, the results of the current investigations have reaffirmed the processing advantage for FL over novel language and sensitivity to frequencies in Chinese phrasal processing in L1 adults, third and sixth graders, as well as L2 learners from distinct L1 backgrounds and of different language proficiencies. Crucially, the current investigations have confirmed that the processing advantage for FL in its original form extends to their modified forms, with two and even four Chinese characters intervening between their components. Of note is that the magnitude of the processing advantage is not consistent but decreases as the modification degree increases. These findings further affirm that the psychological reality of FL in the mental lexicon of L1 adults, children, and L2 learners is tenable even when FL is modified.

The use of eye-tracking and self-paced reading methodologies has allowed me to tap into the mechanisms behind phrasal processing. The eye-tracking methodology has further enabled me to investigate the differences between the whole phrase and its final word in the early versus late stage of FL processing.

Although the findings of the current investigations need further interrogation and replication, I believe they have the potential to contribute to the existing literature and body of knowledge on exposure (experience)-based learning and frequency effects in language processing. I also hope that the current investigations will pave the way for future research on the modification of FL and its role in language acquisition, processing, and use.

References

Abbot-Smit, K., & Tomasello, M. (2006). Exemplar-learning and schematization in a usage-based account of syntactic acquisition. *The Linguistic Review, 23*, 275-290.

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

Ambridge, B. (2020a). Against stored abstractions: A radical exemplar model of language acquisition. *First Language, 40 (5-6)*, 509-559.

Ambridge, B. (2020b). Abstractions made of exemplars or 'You're all right, and I've changed my mind': Response to commentators. *First Language, 40 (5-6)*, 640-659.

Anglin, J. M., Miller, G. A., & Wakefield, P. C. (1993). Vocabulary development: A morphological analysis. *Monographs of the Society for Research in Child Development, 58 (10)*, 1-186.

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Benincà, P., & Semenza, C. (2012). Is "hit and run" a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology, 3*. Retrieved from https://doi.org/10.3389/fpsyg.2012.00011

Arnon, I. (2011). Units of learning in language acquisition. In I. Arnon., & E. V. Clark (Eds.), *Experience, Variation, and Generalization: Learning a First Language* (pp. 167-180). Retrieved from https://doi.org/10.1075/tilar.7

Arnon, I. (2015). What can frequency effects tell us about the building blocks and mechanisms of language learning? *Journal of Child Language, 42*, 274-277.

Arnon, I. (2019). Statistical learning, implicit learning, and first language acquisition: A critical evaluation of two developmental predictions. *Topics in Cognitive Science, 11*, 504-59.

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

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

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

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

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

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modelling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

Babaei, S., Najafabadi, F. T., & Fotovatnia, Z. (2015). Processing of lexical bundles by Persian speaking learners of English. *The Journal of Teaching Language Skills, 6 (4)*, 1-18.

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.

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.

Barlow, M. (2000). Usage, blends and grammar. In M. Barlow and S. Kemmer (Eds.)., *Usage-Based Models of Language* (pp. 315-345). Stanford, USA: CSLI Publications.

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

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). *Parsimonious mixed models*. Ithaca: Cornell University Library. Retrieved from: http://arxiv.org/abs/1506.04967

Bates, E., Bretherton, I., & Snyder, L. (1988). *From First Words to Grammar: Individual Differences and Dissociable Mechanisms*. New York, USA: Cambridge University Press.

Bateson, M. C. (1975). Linguistic models in the study of joint performances. In M. D. Kinkade., K. L. Hale., & O. Werner (Eds.), *Linguistics & Anthropology, in Honor of C. F. Voeglin* (pp. 53-66). Amsterdam, The Netherlands: John Benjamins Publishing.

Beck, S. D., & Weber, A. (2016a). Bilingual and monolingual idiom processing is cut from the same cloth: The role of the L1 in literal and figurative meaning activation. *Frontiers in Psychology, 7*, 1-16.

Beck, S. D., & Weber, A. (2016b). *L2 idiom processing: Figurative attunement in highly idiomatic contexts*. The 38th Annual Conference of the Cognitive Science Society. Philadelphia, USA.

Becker, J. (1975). The phrasal lexicon. *Bolt Beranek & Newman Report no. 3081, AI Report no. 28*. Retrieved from https://www.aclweb.org/anthology/volumes/T75-2/

Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., & Gildea, D. (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *The Journal of the Acoustical Society of America, 113 (2),* 1001-1024.

Biber, D., Conrad, S., & Cortes, V. (2004). If you look at…: Lexical bundles in university teaching and textbooks. *Applied Linguistics, 25 (3)*, 371-405.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Harlow, UK: Longman.

Blais, M-J., & Gonnerman, L. M. (2013). Explicit and implicit semantic processing of verb-particle constructions by French-English bilinguals. *Bilingualism: Language and Cognition, 16 (4)*, 829-846.

BLCU Corpus Center. (2016). Retrieved from http://bcc.blcu.edu.cn

Bod, R. (1998). *Beyond Grammar: An Experience-Based Theory of Language*. Stanford, USA: Center for the Study of Language and Information Publications.

Bod, R. (2006). Exemplar-based syntax: How to get productivity from examples. *The Linguistic Review, 23 (3)*, 291-320.

Bod, R., Hay, J., & Jannedy, S. (2003). *Probabilistic Linguistics*. Cambridge, Massachusetts, USA: The MIT Press.

Boers, F., Eyckmans, J., Kappel, J., Stengers, H., & Demecheleer, M. (2006). Formulaic sequences and perceived oral proficiency: putting a lexical approach to the test. *Language Teaching Research, 10 (3)*, 245-261.

Bolinger, D. (1979). Meaning and memory. In G. G. Haydu (Ed.) *Experience Forms: Their Cultural and Individual Place and Function* (pp. 95-112). Hague, The Netherlands; Paris, France; New York, USA: Mouton Publishers.

Bonatti, L. (2020). PsyScope X. Retrieved from http://psy.ck.sissa.it/

Buerki, A. (2020). (How) is formulaic language universal? Insights from Korean, German and English. In E. Piirainen., N. Filatkina., S. Stumpf., and C. Pfeiffer (Eds.), *Formulaic Language and New Data: Theoretical and Methodological Implications, Formulaic Language, Volume 2* (pp. 103-134). Berlin, Germany: De Gruyter.

Bunt, J. (2003). *Oxford Japanese Grammar & Verbs*. Oxford, UK: Oxford University Press.

Bybee, J. (1998). The emergent lexicon. *Chicago Linguistic Society, 34*, 421-435.

Bybee, J. (2000). The phonology of the lexicon. In M. Barlow, & S. Kemmer (Eds.), *Usage-based Models of Language* (pp. 65–85). Stanford, USA: CSLI Publications.

Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language Variation and Change, 14*, 261-290.

Bybee, J. (2006). From usage to grammar: The mind’s response to repetition. *Language, 82 (4)*, 711-733.

Bybee, J. L. (2011). Usage-based theory and grammaticalization. In B. Heine and H. Narrog (Eds.), *The Oxford Handbook of Grammaticalization* (pp. 225-238). Oxford, UK: Oxford University Press.

Bybee, J., & Beckner, C. (2012). Usage-based theory. In B. Heine and H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (First Edition, pp. 828-856). Oxford, UK: Oxford University Press.

Bybee, J., & Hopper, P. (2001). *Frequency and the Emergence of Linguistic Structure*. Amsterdam, The Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Bybee, J., & Scheibman, J. (1999). The effect of usage on degrees of constituency: The reduction of don’t in American English. *Linguistics, 37*, 575–596.

Cacciari, C., & Glucksberg, S. (1995). Understanding idioms: Do visual images reflect figurative meanings? *European Journal of Cognitive Psychology, 7 (3)*, 283-305.

Cacciari, C., & Levorato, M. C. (1998). The effect of semantic analyzability of idioms in metalinguistic tasks. *Metaphor and Symbol, 13 (3)*, 159-177.

Cacciari, C., & Tabossi, P. (1988). The comprehension of idioms. *Journal of Memory and Language, 27,* 668-683.

Caillies, S., & Butcher, K. (2007). Processing of idiomatic expressions: Evidence for a new hybrid view. *Metaphor and Symbol, 22 (1)*, 79-108.

Caillies, S., & Declercq, C. (2010). Kill the song – steal the show: What does distinguish predicative metaphors from decomposable idioms? *Journal of Psycholinguistic Research, 40*, 205–223.

Caillies, S., & Sourn-Bissaoui, S. L. (2005). Idiom comprehension in French children: A cock-and-bull story. *European Journal of Developmental Psychology*, 189-206.

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

Carrol, G., & Conklin, K. (2014b). Eye-tracking multi-word units: Some methodological questions. *Journal of Eye Movement Research*, *7 (5)*, 1-11.

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 (2)*, 299-317.

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

Carrol, G., & Littlemore, J. (2020). Resolving figurative expressions during reading: The role of familiarity, transparency, and context. *Discourse Processes*, *57 (7)*, 609-626.

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*, 403-443.

Chen, L., Li, X., & Yang, Y. (2012). Focus, newness and their combination: Processing of information structure in discourse. *PLOS ONE*, *7 (8)*, 1-9.

Chinese Testing International. (2012). Chinese proficiency test. Retrieved from HSK https://www.chinesetest.cn

Christiansen, M. H. (2019). Implicit statistical learning: A tale of two literatures. *Topics in Cognitive Science*, *11*, 468-481.

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

Christiansen, M. H., & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science*, *23 (2)*, 157-205.

Chung, K. K. H., & McBride-Chang, C. (2011). Executive functioning skills uniquely predict Chinese word reading. *Journal of Educational Psychology*, *103 (4)*, 909-921.

Cieślicka, A. (2006). Literal salience in on-line processing of idiomatic expressions by second language learners. *Second Language Research*, *22 (2)*, 115-144.

Cieślicka, A. (2010). Formulaic language in L2: Storage, retrieval and production of idioms by second language learners. In M. Pütz., & L. Sicola (Eds.), *Cognitive Processing in*

Second Language Acquisition (pp. 149-168). Amsterdam, The Netherlands; Philadelphia, USA: John Benjamins Publishing Company.

Cieślicka, A. (2013). Second language learners' processing of idiomatic expressions: Does compositionality matter? In K. Droździał-Szelest., & M. Pawlak (Eds.), *Psycholinguistic and Sociolinguistic Perspectives on Second Language Learning and Teaching: Studies in Honor of Waldemar Marton* (pp. 115-136). Heidelberg, Germany; New York, USA; Dordrecht, The Netherlands; London, UK: Springer.

Cieślicka, A., & Heredia, R. R. (2011). Hemispheric asymmetries in processing L1 and L2 idioms: Effects of salience and context. *Brain & Language, 116*, 136-150.

Cieślicka, A., & Heredia, R. R., & Olivares, M. (2014). It's all in the eyes: How language dominance, salience, and context affect eye movements during idiomatic language processing. In M. Pawlak., & L. Aronin (Eds.), *Essential Topics in Applied Linguistics and Multilingualism: Studies in Honor of David Singleton* (pp. 21-41). Zurich, Switzerland: Springer.

Clark, R. (1974). Performing without competence. *Journal of Child Language, 1 (1)*, 1-10.

Colombo, L. (1993). The comprehension of ambiguous idioms in context. In C. Cacciari., & P. Tabossi (Eds.), *Idioms: Processing, Structure, and Interpretation* (pp. 163-200). New Jersey, USA: Lawrence Erlbaum Associates, Inc.

Columbus, G. (2010). Processing MWUs: Are MWU subtypes psycholinguistically real? In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 194-210). London, UK: Continuum.

Columbus, G. (2013). In support of multiword unit classifications: Corpus and human rating data validate phraseological classifications of three different multiword unit types. *Yearbook of Phraseology, 4*, 23-43.

Conklin, K. (2020). Processing single-word and multiword items. In S. Webb (Ed.), *The Routledge Handbook of Vocabulary Studies* (pp. 174-188). Oxon, UK; New York, USA: Routledge.

Conklin, K., & Carrol, G. (2019). First language influence on the processing of formulaic language in a second language. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 62-77). New York, USA: Routledge.

Conklin, K., & Carrol, G. (2020). Words go together like 'bread and butter': The rapid, automatic acquisition of lexical patterns. *Applied Linguistics*. Advanced online. https://doi.org/10.1093/applin/amaa034

Conklin, K., & Pellicer-Sánchez, A. (2016). Using eye-tracking in applied linguistics and second language research. *Second Language Research, 32 (3)*, 453-467.

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.

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A Guide for Applied Linguistics Research*. New York, USA: Cambridge University Press.

Cronk, B. C., Lima, S. D., Schweigert, W. A. (1993). Idioms in sentences: Effects of frequency, literalness, and familiarity. *Journal of Psycholinguistic Research, 22 (1),* 59-82.

Culicover, P. W., Jackendoff, R., & Audring, J. (2017). Multiword constructions in the grammar. *Topics in Cognitive Science, 9,* 552-568.

Cutter, M. G., Driedge, D., & Liversedge, S. P. (2014). Preview benefit in English spaced compounds. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40 (6),* 1778-1786.

Cutting, J. C., & Bock, K. (1997). That's the way the cookie bounces: Syntactic and semantic components of experimentally elicited idiom blends. *Memory and Cognition, 25 (1),* 57-71.

Doughty, C. J., & Long, M. H. (2003). *The Handbook of Second Language Acquisition*. Malden, USA; Oxford, UK; Melbourne, Australia; Berlin, Germany: Blackwell Publishing Ltd.

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

Duyck, W., Van Assche, E., Drieghe, D., & Hartsuiker, R. R. (2007). Visual word recognition by bilinguals in a sentence context: Evidence for nonselective lexical access. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33 (4),* 663-679.

Edmonds, A. (2014). Conventional expressions: Investigating pragmatics and processing. *Studies in Second Language Acquisition, 36,* 69-99.

Ellis, N. C. (1996). Sequencing in SLA: Phonological memory, chunking, and points of order. *Studies in Second Language Acquisition, 18,* 91-126.

Ellis, N. C. (2001). Memory for language. In P. Robinson (Ed.), *Cognition and Second Language Instruction* (pp. 32-68). Cambridge, UK: Cambridge University Press.

Ellis, N. C. (2002a). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies of Second Language Acquisition, 24*, 143-188.

Ellis, N. C. (2002b). Reflections on frequency effects in language processing. *Studies of Second Language Acquisition, 24,* 297-339.

Ellis, N. C. (2003). Constructions, chunking, and connectionism: The emergence of second language structure. In C. J. Doughty and M. H. Long (Eds.), *The Handbook of Second Language Acquisition* (pp. 63-103). Oxford, UK: Blackwell.

Ellis, N. C. (2017). Cognition, corpora and computing: Triangulating research in usage-based language learning. *Language Learning, 67 (S1)*, 40-65.

Ellis, N. C., & Ogden, D. C. (2017). Thinking about multiword construction: Usage-based approaches to acquisition and processing. *Topics in Cognitive Science, 9*, 604-620.

Ellis, N. C., Frey, E., & Jalkanen, I. (2009). The psycholinguistic reality of collocation and semantic prosody (1): Lexical access. In U. Römer & R. Schulze (Eds.), *Exploring the Lexis-Grammar Interface* (pp. 89–114). Amsterdam, The Netherlands; Philadelphia, USA: John Benjamins.

Ellis, N. C., Römer, U. & O'Donnell, M. B. (2016). *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Language Learning Monograph Series. New Jersey, USA: Wiley-Blackwell.

Ellis, N. C., Simpson-Vlach, R., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL QUARTERLY, 42 (3)*, 375-396.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning, 7,* 195-225.

Erman, B., & Warren, B. (2000). The idiom principle and the open choice principle. *Text, 20 (1)*, 29-62.

Estill, R. B., & Kemper, S. (1982). Interpreting idioms. *Journal of Psycholinguistic Research, 11 (6)*, 559-568.

Fanari, R., Cacciari, C., & Tabossi, P. (2010). The role of idiom length and context in spoken idiom comprehension. *European Journal of Cognitive Psychology, 22 (3)*, 321-334.

Fernández, B. G., & Schmitt, N. (2015). How much collocation knowledge do L2 learners have? The effects of frequency and amount of exposure. *International Journal of Applied Linguistics, 166 (1)*, 94-126.

Fillmore, C. J. (1979). On fluency. In C. J. Fillmore., D. Kempler., & S-Y. W. Wang (Eds.), *Individual Differences in Language Ability & Language Behavior* (pp. 85-101). New York, USA: Academic Press.

Firth, J. R. (1937). *The Tongues of Men and Speech*. London, UK: Oxford University Press.

Firth, J. R. (1957). A synopsis of linguistic theory. In *Studies in Linguistic Analysis: Special Volume of the Philological Society* (pp. 1-32). Oxford, UK: Blackwell.

Forster, K. (1976). Accessing the mental lexicon. In R. Wales & E. Walker (Eds.), *New Approaches to Language Mechanisms: A Collection of Psycholinguistic Studies* (pp. 257-284). Amsterdam, The Netherlands: North-Holland Publishing.

Forster, K. (2002). DMDX Display Software. Retrieved from https://www.u.arizona.edu/~kforster/dmdx/dmdx.htm

Foster, P. (2001). Rules & routines: A consideration of their role in the task-based language production of native and non-native speakers. In M. Bygate, P. Skehan & M. Swain (Eds.), *Researching Pedagogic Tasks: Second Language Learning, Teaching and Testing* (pp. 75-94). Harlow, UK: Longman.

Foster-Cohen, S. H. (1999). *An Introduction to Child Language Development*. New York, USA: Longman.

Fowler, C. A. (1988). Differential shortening of repeated content words produced in various communicative contexts. *Language and Speech, 31 (4)*, 307-319.

Frost, R., Armstrong, B. C., Siegelman, N., & Christiansen, M. H. (2015). Domain generality versus modality specificity: The paradox of statistical learning. *Trends in Cognitive Sciences, 19 (3)*, 117-125.

Garnier, M., & Schmitt, N. (2016). Picking up polysemous phrasal verbs: How many do learners know and what facilitates this knowledge? *System, 59*, 29-44.

Geeraert, K., Baayen, R. H., & Newman, J. (2017, April 4). *Understanding idiomatic variation* (pp. 80-90). Proceedings of the 13th Workshop on Multiword Expressions. Valencia, Spain.

Gibbs, R. W. (1980). Spilling the beans on understanding and memory for idioms in conversation. *Memory & Cognition, 8 (2)*, 149-156.

Gibbs, R. W., & Gonzales, G. P. (1985). Syntactic frozenness in processing and remembering idioms. *Cognition, 20*, 243-259.

Gibbs, R. W., Bogdanovich, J. M., Sykes, J. R., & Barr, D. J. (1997). Metaphor in idiom comprehension. *Journal of Memory and Language, 37*, 141-154.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language, 28*, 576-593.

Glass, A. L. (1983). The comprehension of idioms. *Journal of Psycholinguistic Research, 12 (4)*, 429-442.

Glucksberg, S. (1989). Metaphors in conversation: How are they understood? Why are they used? *Metaphor and Symbolic Activity, 4 (3)*, 125-143.

Godfroid, A., Winke, P., & Conklin, K. (2020). Exploring the depths of second language processing with eye tracking: An introduction. *Second Language Research, 36 (3)*, 243-255.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford, UK: Oxford University Press.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22 (5)*, 1166-1183.

Gonnerman, L, M., & Hayes, C. R. (2005). *The professor chewed the students…out: Effects of dependency, length, and adjacency on word order preferences in sentences with verb particle constructions*. Proceedings of the Annual Meeting of the Cognitive Science Society, 27. Retrieved from https://escholarship.org/uc/item/1j27v4qr

Granger, S. (1998). Prefabricated patterns in advanced EFL writing: Collocations and formulae. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 145-160). Oxford, UK: Oxford University Press.

Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., & Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society, 35*, 151–166.

Gu, J., & Li, X. (2015). The effects of character transposition within and across words in Chinese reading. *Attention, Perception and Psychophysics, 77*, 272-281.

Gu, J., Li, X., & Liversedge, S. P. (2015). Character order processing in Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 41 (1)*, 127-137.

Gyllstad, H., & Wolter, B. (2016). Collocational processing in light of the phraseological continuum model: Does semantic transparency matter? *Language Learning, 66 (2),* 296-323.

Haeuser, K. I., Baum, S., & Titione, D. (2020). Effects of aging and noncanonical form presentation on idiom processing: Evidence from eye tracking. *Applied Psycholinguistics*. Advance online publication. http://doi.org/10.1017/S0142716420000612

Haiman, J. (1994). Ritualization and the development of language. In W. Pagliuca (Ed.), *Perspectives on Grammaticalization* (pp. 3-28). Amsterdam, The Netherlands: John Benjamins Publishing.

Hakuta, K. (1974). Prefabricated patterns and the emergence of structure in second language acquisition. *Language Learning, 24 (2)*, 287-297.

Hallin, A. E., & Van Lancker Sidtis, D. (2017). A closer look at formulaic language: Prosodic characteristics of Swedish proverbs. *Applied Linguistics, 38 (1)*, 68-89.

Havron, N., & Arnon, I. (2020). Starting big: The effect of unit size on language learning in children and adults. *Journal of Child Language*. Advance online publication. https://doi.org/10.1017/S0305000920000264

Henderson, J. M., & Ferreira, F. (1990). Effects of foveal processing difficulty on the perceptual span in reading: Implications for attention and eye movement control. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16 (3)*, 417-429.

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

Ho, C. S-H., Chan, D. W-O., Lee, S-H., Tsang, S-M., & Luan, V. H. (2004). Cognitive profiling and preliminary subtyping in Chinese developmental dyslexia. *Cognition, 91*, 43-75.

Hoey, M. (2005). *Lexical Priming: A New Theory of Words and Language*. London, UK: Routledge.

Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: Theory, Analysis and Applications* (pp. 161–186). Oxford, UK: Oxford University Press.

Huey, E. B. (1908). *The Psychology and Pedagogy of Reading*. New York, USA: Macmillan.

Hunston, S. (2002). *Corpora in Applied Linguistics*. Cambridge, UK: Cambridge University Press.

Hymes, D. H. (1968). The ethnography of speaking. In J. A. Fishman (Ed.), *Readings in the Sociology of Language* (pp. 99-138). Hague, The Netherlands; Paris, France: Mouton.

Inhoff, A. M., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception* (pp. 29-54). Amsterdam, The Netherlands: Elsevier.

Institute of Linguistics at the Chinese Academy of Social Sciences. (2002). *Modern Chinese Dictionary*. Beijing, China: The Commercial Press.

Isobe, Y. (2011). Representation and processing of formulaic sequences in L2 mental lexicon: How do Japanese EFL learners process multi-word expressions? *JACET Kansai Journal, 13,* 38-49.

Iwasaki, S., & Ingkaphirom, P. (2005). *A Reference Grammar of Thai*. Cambridge, UK: Cambridge University Press.

Janssen, N., & Barber, H. A. (2012). Phrase frequency effects in language production. *PLOS One*, *7 (3),* 1-11.

Jescheniak, J. D., & Levelt, W. J. M. (1994). Word frequency effects in speech production: Retrieval of syntactic information and phonological form. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20 (4)*, 824-843.

Jespersen, O. (1924). *The Philosophy of Grammar*. London, UK: George Allen & Unwin Ltd.

Jespersen, O. (1976). Living grammar. In D. D. Bornstein (Ed.), *Readings in the Theory of Grammar* (pp. 82-93). Cambridge, Massachusetts, USA: Winthrop Publishers.

Jiang, N., & Nekrasova, T. M. (2007). The processing of formulaic sequences by second language speakers. *The Modern Language Journal, 91 (iii)*, 433-445.

Jiang, S., Jiang, X., & Siyanova-Chanturia, A. (2020). The processing of multi-word expressions in children and adults: An eye-tracking study of Chinese. *Applied Psycholinguistics, 41 (4)*, 901-931.

Jolsvai, H., McCauley, S. M., & Christiansen, M. H. (2013, January). *Meaning overrides frequency in idiomatic and compositional multiword chunks*. Proceedings of the 35th Annual Conference of the Cognitive Science Society. Retrieved from http://cogsci.mindmodeling.org/2013/papers/0146/index.html

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science, 20*, 137-194.

Jurafsky, D. (2003). Probabilistic modelling in psycholinguistics: Linguistic comprehension and production. In R. Bod., J. Hay., & S. Jannedy (Eds.), *Probabilistic Linguistics* (pp. 39-96). Cambridge, Massachusetts, USA; London, UK: The MIT Press.

Jurafsky, D., Bell, A., Gregory, M. L., & Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In J. L. Bybee, & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 229–254). Amsterdam, The Netherlands: John Benjamins.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87 (4)*, 329-354.

Kapatsinski, V., & Radicke, J. (2009). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan., E. A. Moravcsik., H. Ouali., & K. Wheatley (Eds.), *Formulaic Language, Volume II: Acquisition, Loss, Psychological Reality, Functional Explanations* (pp. 499–520). Amsterdam, Holland: John Benjamins.

Keating, G. D., & Jegerski, J. (2015). Experimental designs in sentence processing research: A methodological review and user's guide. *Studies in Second Language Acquisition, 37*, 1-32.

Kemper, S. (1986). Inferential processing and the comprehension of idioms. *Metaphor and Symbolic Activity, 1 (1)*, 43-55.

Kempler, D., Van Lancker, D., Marchman, V., & Bates, E. (1999). Idiom comprehension in children and adults with unilateral brain damage. *Developmental Neuropsychology, 15 (3)*, 327-349.

Kerz, E., Wiechmann, D., Frinsel, F. F., & Christiansen, M. H. (2020). Tuning to real-life language statistics: Online processing of multiword sequences in native and non-native speakers across language registers. *PsyArXiv*. Preprint retrieved from http://doi.org/10.31234/osf.io/v4tbu

Kim, S. H., & Kim, J. H. (2012). Frequency effects in L2 multiword unit processing: Evidence from self-paced reading. *TESOL QUARTERLY, 46 (4)*, 831-841.

Kong, L., Zhang, J. X., & Zhang, Y. (2016). Are Chinese correlative conjunctions psychologically real? An investigation of the combination frequency effect. *Psychological Reports, 119 (1)*, 106-123.

Konopka, A. E., & Bock, K. (2009). Lexical or syntactic control of sentence formulation? Structural generalizations from idiom production. *Cognition Psychology, 58*, 68-101.

Krishnamurthy, R., Sinclair, J., Jones, S., & Daley, R. (2004). *English Collocation Studies: The OSTI Report*. London, UK; New York, USA: Continuum.

Krummes, C., & Ensslin, A. (2015). Formulaic language and collocations in German essays: from corpus-driven data to corpus-based materials. *The Language Learning Journal, 43 (1)*, 110-127.

Kyriacou, M., Conklin, K., & Thompson, D. (2020). Passivizability of idioms: Has the wrong tree been barked up? *Language and Speech, 63 (2),* 404-435.

Lamb, S. M. (1999). *Pathways of the Brain: The Neurocognitive Basis of Language*. Amsterdam, The Netherlands: John Benjamins Publishing Company.

Landerl, K., & Wimmer, H. (2008). Development of word reading fluency and spelling in a consistent orthography: An 8-year follow-up. *Journal of Educational Psychology, 100 (1),* 150–161.

Langacker, R. (1987). *Foundations of Cognitive Grammar*. Stanford, California, USA: Stanford University Press.

Laurent, J-P., Denhières, G., Passerieux, C., Iakiomva, G., Hardy-Baylé, M-C. (2006). On understanding idiomatic language: The salience hypothesis assessed by ERPs. *Brain Research, 1068,* 151-160.

Lee, Y-S., & Chen, H-H. (1997). Metadata for integrating Chinese text and speech documents in a multimedia retrieval system. *The Association for the Advancement of Artificial Intelligence Technical Report*, 64-69.

Levorato, M. C., & Cacciari, C. (1999). Idiom comprehension in children: Are the effects of semantic analyzability and context separable? *European Journal of Cognitive Psychology, 11 (1),* 51-66.

Lewis, M. (2000). *Teaching Collocations: Further Developments in the Lexical Approach*. Boston, Massachusetts, USA: Thomson.

Li, L., Liu, J., Chen, F., Feng, L., Li, H., Tian, J., & Lee, K. (2013). *Neuropsychologia, 51,* 1571-1583.

Li, S., & Guo, S. (2016). Collocation analysis tools for Chinese collocation studies. *Journal of Technology and Chinese Language Teaching, 7 (1),* 56-77.

Li, T., Wang, Y., Tong, X., & McBride, C. (2017). A developmental study of Chinese children’s word and character reading. *Journal of Psycholinguistic Research, 46,* 141-155.

Li, X., & Logan, G. D. (2008). Object-based attention in Chinese readers of Chinese words: Beyond Gestalt principles. *Psychonomic Bulletin & Review, 15 (5),* 945-949.

Li, X., & Pollatsek, A. (2020). An integrated model of word processing and eye-movement control during Chinese reading. *Psychological Review, 127 (6),* 1139-1162.

Libben, M. R., & Titone, D. A. (2008). The multidetermined nature of idiom processing. *Memory & Cognition, 36 (6),* 1103-1121.

Lieven, E, V. M., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language*, *19*, 287-310.

Lieven, E. V. M., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: A usage-based approach. *Journal of Child Language, 30*, 333-370.

Lieven, E. V. M., Pine, J. M., & Barnes, H. D. (1992). Individual differences in early vocabulary development: Redefining the referential-expressive distinction. *Journal of Child Language*, *19*, 287-310.

Lieven, E. V. M., Salomo, D., & Tomasselo, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics, 20 (3)*, 481-507.

Lieven, E., Ferry, A., Theakston, A., & Twomey, K. E. (2020). Similarity, analogy and development in radical exemplar theory: A commentary on Ambridge (2020). *First Language, 40 (5-6)*, 600-603.

Lin, N., Angele, B., Hua, H., Shen, W., Zhou, J., & Li, X. (2018). Skipping of Chinese characters does not rely on word-based processing. *Attention, Perception, & Psychophysics, 80*, 600-607.

Liu, P., Li, X., & Han, B. (2015). Additive effects of stimulus quality and word frequency on eye movements during Chinese reading. *Reading and Writing*, *28*, 199-215.

Liu, Y., Li, P., Shu, H., Zhang, Q., & Chen, L. (2010). Structure and meaning in Chinese: An ERP study of idioms. *Journal of Neurolinguistics*, *23*, 615-630.

Liu, N., Wang, X., Yan, G., Paterson, K. B., & Pagán, A. (2020). Eye movements of developing Chinese readers: Effects of word frequency and predictability. *Scientific Studies of Reading*. Advance online publication. https://doi.org/10.1080/10888438.2020.1759074

Liversedge, S. P., Drieghe, D., Li, X., Yan, G., Bai, X., & Hyönä, J. (2016). Universality in eye movements and reading: A trilingual investigation. *Cognition*, *147*, 1-20.

Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.) *Eye Guidance in Reading and Scene Perception* (pp. 55-76). Amsterdam, The Netherlands: Elsevier.

Locke, J. L. (1997). A theory of neurolinguistic development. *Brain and Language*, *58*, 265-326.

López, B. G., & Vaid, J. (2017). Fácil or a piece of cake: Does variability in bilingual language brokering experience affect idiom comprehension? *Bilingualism: Language and Cognition, 21 (2),* 340-354.

López, B. G., Vaid, J., Tosun, S., & Rao, C. (2017). Bilinguals' plausibility judgments for phrases with a literal vs. non-literal meaning: The influence of language brokering experience. *Frontiers in Psychology, 8 (1661)*. Retrieved from https://doi.org/10.3389/fpsyg.2017.01661

Ma, G., & Li, X. (2015). How character complexity modulates eye movement control in Chinese reading. *Reading and Writing: An Interdisciplinary Journal, 28,* 747-761.

Ma, G., Li, X., & Rayner, K. (2014). Word segmentation of overlapping ambiguous strings during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 40 (3),* 1046-1059.

Ma, G., Li, X., & Rayner, K. (2015). Readers extract character frequency information from nonfixated-target word at long pre-target fixations during Chinese reading. *Journal of Experimental Psychology: Human Perception and Performance, 41 (5),* 1409-1419.

Ma, G., Li, Z., Xu, F., & Li, X. (2018). The modulation of eye movement control by word length in reading Chinese. *Quarterly Journal of Experimental Psychology, 72 (7),* 1620-1631.

Ma, G., Pollatsek, A., Li, Y., & Li, X. (2016). Chinese readers can perceive a word even when it's composed of noncontiguous characters. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43 (1),* 158-166.

Maad, M. R. B. (2010). Holistic and analytic processing modes in non-native learners' performance of narrative tasks. *System, 38,* 591-602.

Macis, M., & Schmitt, N. (2017). Not just "small potatoes": Knowledge of the idiomatic meanings of collocations. *Language Teaching Research, 21 (3),* 321-340.

MacWhinney, B. (1998). Models of the emergence of language. *Annual Review of Psycholinguistics, 49,* 199-227.

MacWhinney, B. (2020). The role of competition and timeframes: A commentary on Ambridge (2020). *First Language, 40 (5-6),* 604-607.

Marsden, E., Thompson, S., & Plonsky, L. (2018). A methodological synthesis of self-paced reading in second language research. *Applied Psycholinguistics, 39,* 861-904.

Massachusetts Institute of Technology. (2020). Linger. Retrieved from https://www.bu.edu/linguistics/UG/hagstrom/projects/linger/

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods, 44 (2),* 314-324.

Matlock, T., & Heredia, R. R. (2002). Understanding phrasal verbs in monolinguals and bilinguals. In R. R. Heredia and J. Altarriba (Eds.), *Bilingual Sentence Processing* (pp. 251-274). Amsterdam, The Netherlands: North-Holland. Also published in *Advances in Psychology, 134*, 251-274.

McDonald, S. A., & Shillcock, R. C. (2003a). Eye movements reveal the on-line computation of lexical probabilities during reading. *Psychological Science, 14 (6)*, 648-652.

McDonald, S. A., & Shillcock, R. C. (2003b). Low-level predictive inference in reading: The influence of transitional probabilities on eye movements. *Vision Research, 43*, 1735-1751.

McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes*, *17*, 167-190.

Metzner. P., von der Malsburg. T., Vasishth. S., Rösler. F. (2017). The importance of reading naturally: Evidence from combined recordings of eye movements and electric brain potentials. *Cognitive Science, 41 (6)*, 1232-1263.

Millar, N. (2011). The processing of malformed formulaic language. *Applied Linguistics*, *32 (2)*, 129-148.

Molinaro, N., & Carreiras, M. (2010). Electrophysiological evidence of interaction between contextual expectation and semantic integration during the processing of collocations. *Biological Psychology, 83*, 176-190.

Molinaro, N., Canal, P., Vespignani, F., Pesciarelli, F., & Cacciari, C. (2013). Are complex function words processed as semantically empty strings? A reading time and ERP study of collocational complex prepositions. *Language and Cognitive Processes, 28 (6)*, 762-788.

Molinaro, N., Vespignani, F., Canal, P., Fonda, S., & Cacciari, C. (2008). Cloze probability does not only affect N400 amplitude: The case of complex prepositions. *Psychophysiology, 45*, 1008-1012.

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

Moreno, E. M., Federmeier, K. D., & Kutas, M. (2002). Switching languages, switching palabras (words): An electrophysiological study of code switching. *Brain and Language, 80*, 188-207.

Nattinger, J. R., & DeCarrico, J. S. (1992). *Lexical Phrases and Language Teaching*. Oxford, UK: Oxford University Press.

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the Society for Research in Child Development, 38 (1-2),* 1-135.

Nelson, K. (1981). Individual differences in language development: Implications for development and language. *Developmental Psychology, 17 (2),* 170-187.

Northbrook, J., & Conklin, K. (2019). Is what you put in what you get out? – Textbook-derived lexical bundle processing in beginner English learners. *Applied Linguistics, 40 (5),* 816-833.

Öksüz, D., Brezina, V., & Rebuschat, P. (2020). Collocational processing in L1 and L2: The effects of word frequency, collocational frequency, and association. *Language Learning*. Advance online publication. http://doi.org/10.1111/lang.12427

Open Science Tools Ltd. (2018). PsychoPy v3.0. Retrieved from: https://www.psychopy.org/.

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of verbal learning and verbal behaviour*, *17*, 465-477.

Paquot, M., & Granger, S. (2012). Formulaic language in learner corpora. *Annual Review of Applied Linguistics*, *32*, 130-149.

Paulmann, S., Ghareeb-Ali, Z., & Felser, C. (2015). Neurophysiological markers of phrasal verb processing: Evidence from L1 and L2 speakers. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 245-267). Cambridge, UK: Cambridge University Press.

Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. In J. C. Richards & R. W. Schmidt (Eds.), *Language and Communication* (pp. 191–226). New York, USA: Longman.

Pellicer-Sánchez, A., & Siyanova-Chanturia, A. (2018). Eye movements in vocabulary research. *International Journal of Applied Linguistics, 169 (1),* 5-29.

Peters, A. M. (1977). Language learning strategies: Does the whole equal the sum of the parts? *Language, 53 (3),* 560-573.

Peters, A. M. (1983). *Units of Language Acquisition*. Cambridge, UK: Cambridge University Press.

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

Peterson, R. R., Burgess, C., Dell, G. S., & Eberhard, K. M. (2001). Dissociation between syntactic and semantic processing during idiom comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27 (5)*, 1223-1237.

Pierrehumbert, J. B. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure* (pp. 137–157). Amsterdam, The Netherlands: John Benjamins Publishing Company.

Pine, J. M., & Lieven, E. V. M. (1993). Reanalysing rote-learned phrases: Individual differences in the transition to multi-word speech. *Journal of Child Language, 20*, 551-571.

Pollio, H. R., Barlow, J. M., Fine, H. J., & Pollio, M. R. (1977). *Psychology and the Poetics of Growth: Figurative Language in Psychology, Psychotherapy and Education*. Hillsdale, New Jersey, USA: Lawrence Erlbaum Associates, Inc.

Pollio, M. R., & Pollio, H. R. (1974). The development of figurative language in children. *Journal of Psycholinguistic Research, 3 (3)*, 185-201.

Psychology Software Tools Inc. (2020). E-Prime. Retrieved from https://pstnet.com/products/e-prime/

Rammell, C. S., Van Lancker Sidtis, D., & Pisoni, D. B. (2017). Perception of formulaic and novel expressions under acoustic degradation. *Mental Lexicon, 12 (2)*, 234-262.

Raviv, L., & Arnon, I. (2017). The developmental trajectory of children's auditory and visual statistical learning abilities: modality-based differences in the effect of age. *Developmental Science*. Retrieved from https://doi.org/10.1111/desc.12593

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

Rayner, K. (2009). The 35th Sir Frederick Bartlett Lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology, 62 (8)*, 1457-1506.

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

Rayner, K., & Pollatsek, A. (1989). *The Psychology of Reading*. Englewood Cliffs, New Jersey, USA: Prentice-Hall.

Rayner, K., Li, X., Juhasz, B. J., & Yan, G. (2005). The effect of word predictability on the eye movements of Chinese readers. *Psychonomic Bulletin & Review, 12 (6)*, 1089-1093.

Rayner, K., Pollatsek, A., Ashby, J., & Clifton, C. (2012). *Psychology of Reading (Second Edition)*. New York, USA; London, UK: Psychology Press.

Rayner, K., Sereno, S. C., & Raney, G. E. (1996). Eye movement control in reading: A comparison of two types of models. *Journal of Experimental Psychology: Human, Perception and Performance, 22 (5)*, 1188-1200.

Rayner, K., Sereno, S. C., Morris, R. K., Schmauder, A. R., & Clifton, C. (1989). Eye movements and on-line language comprehension processes. *Language and Cognitive Processes, 4 (3/4)*, SI 21-49.

Renouf, A., & Sinclair, J. (1991). Collocational frameworks in English. In K. Aijmer & B. Altenberg (Eds.), *English Corpus Linguistics* (pp. 128-144). London: Longman.

Reuterskiöld, C., & Van Lancker Sidtis, D. (2012). Retention of idioms following one-time exposure. *Child Language Teaching and Therapy, 29 (2)*, 219-231.

Robeck, M. C., & Wallace, R. R. (2018). *The Psychology of Reading: An Interdisciplinary Approach (Second Edition)*. New York, USA: Routledge.

Robert, L. (2012). Psycholinguistic techniques and resources in second language acquisition research. *Second Language Research*, *28 (1)*, 113-127.

Roberts, L., & Siyanova-Chanturia, A. (2013). Using eye-tracking to investigate topics in L2 acquisition and L2 processing. *Studies in Second Language Acquisition*, *35*, 213-235.

Rommers, J., Dijkstra, T., & Bastiaansen, M. (2013). Context-dependent semantic processing in the human brain: Evidence from idiom comprehension. *Journal of Cognitive Neuroscience*, *25(5)*, 762–776.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning the past tenses of English verbs. In D. Rumelhart & J. McClelland (Eds.), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (pp. 216–271). Massachusetts, USA: MIT Press.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12(4)*, 110–114.

Santelmann, L. M., & Jusczyk, P. W. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for limitations in processing space. *Cognition, 69*, 105-134.

Saussure, F. D. (1916). *Course in General Linguistics*. New York, USA: McGraw-Hill.

Schmitt, N. (2004). *Formulaic Sequences: Acquisition, Processing and Use*. Amsterdam, The Netherlands: John Benjamins.

Schmitt, N. (2005). Formulaic language: Fixed and varied. *Estudios de Lingüística Inglesa Aplicada, 6*, 13-39.

Schmitt, N., & Underwood, G. (2004). Exploring the processing of formulaic sequences through a self-paced reading task. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 127-152). Amsterdam, The Netherlands: John Benjamins.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 172-190). Amsterdam, The Netherlands: John Benjamins.

Schweigert, W. A., & Moates, D. R. (1988). Familiar idiom comprehension. *Journal of Psycholinguistic Research, 17 (4)*, 281-296.

Seidenberg, M. S. (1994). Language and connectionism: The developing interface. *Cognition, 50*, 385-401.

Seidenberg, M. S., & MacDonald, M. C. (1999). A probabilistic constraints approach to language acquisition and processing. *Cognitive Science*, *23 (4)*, 569-588.

Shantz, K. (2017). Phrase frequency, proficiency and grammaticality interact in non-native processing: Implications for theories of SLA. *Second Language Research, 33 (1)*, 91-118.

Shen, H., Wang, Y., & Liu, W. (2017). Familiarity in Chinese and English idiom comprehension: An event-related potential study from Chinese English learners. *Chinese Journal of Applied Linguistics, 40 (1)*, 74-92.

Shen, W., Li, X., & Pollatsek, A. (2018). The processing of Chinese compound words with ambiguous morphemes in sentence context. *Quarterly Journal of Experimental Psychology, 71 (1)*, 131-139.

Shu, H., Chen, X., Anderson, R. C., Wu, N., & Xuan, Y. (2003). Properties of school Chinese: Implications for learning to read. *Child Development, 74 (1)*, 27-47.

Shufaniya, A., & Arnon, I. (2018). Statistical learning is not age-invariant during childhood: Performance improves with age across modality. *Cognitive Science, 42,* 3100-3115.

Siegelman, N., & Arnon, I. (2015). The advantage of starting big: Learning from unsegmented input facilitates mastery of grammatical gender in an artificial language. *Journal of Memory and Language*, 60-75.

Siewierska, A., Xu, J., & Xiao, R. (2010). *Bang-le yi ge da mang* (offered a big helping hand): A corpus study of the splittable compounds in spoken and written Chinese. *Language Sciences, 32*, 464-487.

Sinclair, J. M. (1991). *Corpus, Concordance, Collocation*. Oxford, UK: Oxford University Press.

Sinclair, J. M. (1998). The lexical item. In E. Weigand (Ed.), *Contrastive Lexical Semantics* (pp. 1-24). Amsterdam, The Netherlands: John Benjamins.

Sinclair, J. M., & Renouf, A. (1988). A lexical syllabus for language learning. In R. Carter & M. McCarthy (Eds.), *Vocabulary and Language Teaching* (pp. 140-160). London, UK; New York, USA: Longman.

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.

Siyanova-Chantruai, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175*, 111-122.

Siyanova-Chanturia, A. (2013). Eye-tracking and ERPs in multi-word expression research: A state-of-the-art review of the method and findings. *The Mental Lexicon, 8 (2)*, 245-268.

Siyanova-Chanturia, A. (2015). On the “holistic” nature of formulaic language. *Corpus Linguistics and Linguistic Theory, 11 (2)*, 285-301.

Siyanova-Chanturia, A., & Janssen, N. (2018). Production of familiar phrases: Frequency effects in native speakers and second language learners. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 44 (12), 2009-2018.*

Siyanova-Chanturia, A., & Lin, P. (2018). Production of ambiguous idioms in English: A reading aloud study. *International Journal of Applied Linguistics*, *28*, 58-70.

Siyanova-Chanturia, A., & Pellicer-Sánchez, A. (2019). Formulaic language: Setting the scene. In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 1-15). New York, USA: Routledge.

Siyanova-Chanturia, A., & Spina, S. (2015). Investigation of native speaker and second language learner intuition of collocation frequency. *Language Learning, 65 (3),* 533-562.

Siyanova-Chanturia, A., & Van Lancker Sidtis, D. (2019). What online processing tells us about formulaic language? In A. Siyanova-Chanturia & A. Pellicer-Sánchez (Eds.), *Understanding Formulaic Language: A Second Language Acquisition Perspective* (pp. 38-61). New York, USA: Routledge.

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

Siyanova-Chanturia, A., Conklin, K., & van Heuven, W. J. B. (2011b). Seeing a phrase "time and again" matters: The role of phrasal frequency in the processing of multi-word sequences. *Journal of Experimental Psychology: Language, Memory and Cognition, 37*, 776-784.

Siyanova-Chanturia, A., Conklin, K., Caffarra, S., Kaan, E., & van Heuven, W. J. B. (2017). Representation and processing of multi-word expressions in the brain. *Brain and Language, 175*, 111-122.

Siyanova-Chanturia, A., Martinez, R. (2014). The idiom principle revisited. *Applied Linguistics, 36 (5),* 549-569.

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. *PsyArXiv*. Preprint retrieved from https://doi.org/10.31234/osf.io/jwpdc

SMI. (2020). SensoMotoric Instruments. Retrieved from http://www.smivision.com/en.html

Snider, N., & Arnon, I. (2012). A unified lexicon and grammar? Compositional and non-compositional phrases in the lexicon. In D. Divjak., & S. T. Gries (Eds.), *Frequency Effects in Language Representation* (pp. 127-163). Berlin, Germany; Boston, USA: De Gruyter Mouton.

Sonbul, S. (2014). Fatal mistake, awful mistake, or extreme mistake? Frequency effects on off-line/on-line collocational processing. *Bilingualism: Language and Cognition, 18 (3),* 419-437.

Sonbul, S., & El-Dakhs, D. (2020). Timed versus untimed recognition of L2 collocations: Does estimated proficiency modulate congruency effects? *Applied Psycholinguistics, 41*, 1197-1222.

Sosa, A. V., & MacFarlane, J. (2002). Evidence for frequency-based constituents in the mental lexicon: Collocations involving the word *of*. *Brain and Language, 83*, 227-236.

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

SR Research Ltd. (2017). *Eyelink 1000 Plus User Manual* (version 1.0.12). Retrieved from https://www.sr-research.com/support/learning-resources

SR Research Ltd. (2020). EyeLink. Retrieved from https://www.sr-research.com/

Statista. (2020). The most spoken languages worldwide in 2019 (by speakers in millions). Retrieved from http://www.statista.com/statistics/266808/the-most-spoken-languages-worldwide

Staub, A., & Rayner, K. (2007). Eye movements and on-line comprehension processes. In Gaskell (Ed.) *The Oxford Handbook of Psycholinguistics* (pp. 327-342). New York, USA: Oxford University Press.

Strandburg, R. J., Marsh, J. T., Brown, W. S., Asarnow, R. F., Guthrie, D., & Higa, J. (1993). Event-related potentials in high-functioning adult autistics: Linguistic and nonlinguistic visual information processing tasks. *Neuropsychologia, 31 (5)*, 413-434.

Stubbs, M. (1996). *Text and Corpus Analysis: Computer-Assisted Studies of Language and Culture*. Oxford, UK; Massachusetts, USA: Blackwell.

Stubbs, M. (2000). Using very large text collections to study semantic schemas: A research note. In C. Heffer & H. Saunton (Eds.), *Words in Context*. CD-ROM. Birmingham, UK: University of Birmingham.

Svartvik, J. (1990). The London Corpus of Spoken English: Description and Research. Lund Studies in English, 82. Lund, UK: Lund University Press.

Swinney, D. A., & Cutler, A. (1979). The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior*, *18*, 523-534.

Tabossi, P., Fanari, R., & Wolf, K. (2005). Spoken idiom recognition: Meaning retrieval and word expectancy. *Journal of Psycholinguistic Research*, *34 (5)*, 465-495.

Tabossi, P., Fanari, R., & Wolf, K. (2008). Processing idiomatic expressions: Effects of semantic compositionality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34 (2)*, 313-327.

Tabossi, P., Fanari, R., & Wolf, K. (2009). Why are idioms recognized fast? *Memory and Cognition, 37*, 529-540.

Taylor, I. (1981). Writing systems and reading. In G. E. MacKinnon & T. G. Waller (Eds.), *Reading Research: Advances in Theory and Practice (Volume 2)* (pp. 1-47). New York, USA: Academic Press.

The Commercial Press. (2008). *Lexicon of Common Words in Contemporary Chinese*. Beijing: The Commercial Press.

Titone, D. A., & Connine, C. M. (1994). Comprehension of idiomatic expressions: Effects of predictability and literality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20 (5)*, 1126-1138.

Titone, D.A, Columbus, G., Whitford, V., Mercier, J., & Libben, M. (2015). Contrasting bilingual and monolingual idiom processing. In R. R. Heredia & A. B. Cieślicka (Eds.), *Bilingual Figurative Language Processing* (pp. 171-207). Cambridge, UK: Cambridge University Press.

Tobii Group. (2020). Tobii. Retrieved from https://www.tobii.com/

Tomasello, M. (2003). *Constructing a Language: A Usage-based Theory of Language Acquisition*. Cambridge, Massachusetts, USA: Harvard University Press.

Tremblay, A., & Baayen, H. (2010). Holistic processing of regular four-word sequences: A behavioural and ERP study of the effects of structure, frequency, and probability on immediate free recall. In D. Wood (Ed.), *Perspectives on Formulaic Language: Acquisition and Communication* (pp. 151-173). London, UK: Continuum.

Tremblay, A., & Tucker, B. V. (2011). The effects of N-gram probabilistic measures on the recognition and production of four-word sequences. *The Mental Lexicon, 6 (2)*, 302-324.

Tremblay, A., Derwing, B., & Libben, G. (2009). Are lexical bundles stored and processed as single units? *Proceeding of the 23rd Northwest Linguistics Conference, Victoria, BC, Canada*.

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

Underwood, G., Schmitt, N., & Galpin, A. (2004). The eyes have it: An eye-movement study into the processing of formulaic sequences. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 152-172). Amsterdam, The Netherlands: John Benjamins.

Valsecchi, M., Künstler, V., Saage, S., White, B. J., Mukherjee, J., & Gegenfurtner, K. R. (2013). Advantage in reading lexical bundles is reduced in non-native speakers. *Journal of Eye Movement Research, 6 (5): 2*, 1-15.

Van Lancker, D., Canter, G. J., & Terbeek, D. (1981). Disambiguation of ditropic sentences: Acoustic and phonetic cues. *Journal of Speech and Hearing Research, 24*, 330-335.

Van Lancker, D., & Cummings, J. L. (1999). Expletives: Neurolinguistic and neurobehavioral perspectives on swearing. *Brain Research Review, 31*, 83-104.

Van Lancker-Sidtis, D. (2003). Auditory recognition of idioms by native and non-native speakers of English: It takes one to know one. *Applied Psycholinguistics, 24*, 45-57.

Van Lancker-Sidtis, D. (2012). Formulaic language and language disorders. *Annual Review of Applied Linguistics, 32*, 62-80.

Van Lancker-Sidtis, D. (2015). Formulaic language in an emergentist framework. In B. MacWhinney and W. O'Grady (Eds.), The Handbook of Language Emergence (First Edition) (pp. 578-599). Sussex, UK: John Wiley & Sons, Inc.

Van Lancker-Sidtis, D., & Sidtis, J. J. (2018a). Cortical-subcortical production of formulaic language: A review of linguistic, brain disorder, and functional imaging studies leading to a production model. *Brain and Cognition, 126*, 53-64.

Van Lancker-Sidtis, D., & Sidtis, J. J. (2018b). The affective nature of formulaic language: A right-hemisphere subcortical process. *Frontiers in Neurology, 9 (573)*, 1-6.

Van Lancker-Sidtis, D., Choi, J., Alken, A., & Sidtis, J. J. (2015). Formulaic language in Parkinson's disease and Alzheimer's disease: Complementary effects of subcortical and cortical dysfunction. *Journal of Speech, Language, and Hearing Research, 58*, 1493-1507.

Van Lancker-Sidtis, D., Pachana, N., Cummings, J. L. & Sidtis, J. J. (2006). Dysprosodic speech following basal ganglia insults: Toward a conceptual framework for the study of the cerebral representation of prosody. *Brain and Language, 97*, 135-153.

VanPattern, B., & Williams, J. (2015). *Theories in Second Language Acquisition: An Introduction (Second Edition)*. New York, USA: Routledge.

Vespignani, F., Canal, P, Molinaro, N., Fonda, S., & Cacciari, C. (2009). Predictive mechanisms in idiom comprehension. *Journal of Cognitive Neuroscience, 22 (8)*, 1682-1700.

Vilkaitė, L. (2016a). Are nonadjacent collocations processed faster? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42 (10)*, 1632-1642.

Vilkaitė, L. (2016b). Formulaic language is not all the same: comparing the frequency of idiomatic phrases, collocations, lexical bundles, and phrasal verbs. *Taikomoji Kalbotyra, 8*, 28-54.

Vilkaitė, L. (2017). Incidental acquisition of collocations in L2: Effects of adjacency and prior vocabulary knowledge. *International Journal of Applied Linguistics, 168 (2)*, 248-277.

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

Vonk, W., Radach, R., & van Rijn, H. (2000). Eye guidance and the saliency of word beginnings in reading text. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a Perceptual Process* (pp. 269–299). Amsterdam, The Netherlands: North-Holland/Elsevier Science Publishers.

Vulchanova, M., Vulchanov, V., & Stankova, M. (2011). Idiom comprehension in the first language: A developmental study. *Vigo International Journal of Applied Linguistics, 8 (1),* 207-234.

Wallace, M., & Wray, A. (2016). *Critical Reading and Writing for Postgraduates (Third Edition)*. Los Angeles, USA; London, UK; New Delhi, India; Singapore; Washington DC, USA; Melbourne, Australia: Sage.

Wang, J., Angele, B., Ma, G., & Li, X. (2020). Repetition causes confusion: Insights into word segmentation during Chinese reading. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. Advance online publication. http://doi.org/10.1037/xlm0000817

Wang, R. (2005). 留学生汉语离合词使用偏误的分析. *《语言文字应用》*, 27-29.

Wei, W., Li, X., & Pollatsek, A. (2013). Word properties of a fixated region affect outgoing saccade length in Chinese reading. *Vision Research, 80*, 1-6.

Wiechmann, D., Kerz, E., Snider, N., & Jaeger, T. F. (2013). Introduction to the special issue: Parsimony and redundancy in models of language. *Language and Speech, 56 (3)*, 257-264.

Winke, P. M., Godfroid, A., & Gass, S. M. (2013). Introduction to the special issue: Eye-movement recordings in second language research. *Studies in Second Language Acquisition, 35*, 205-212.

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

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

Wolter, B., & Yamashita, J. (2015). Processing collocations in a second language: A case of first language activation? *Applied Psycholinguistics, 36 (5)*, 1193-1221.

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*, 395-416.

Wong, H. P. (2012). *Use of formulaic sequences in task-based oral production of Chinese* (Doctoral dissertation). Durham University, Durham, UK. Available at Durham E-Theses Online https://etheses.dur.ac.uk/5947/

Wong-Fillmore, L. (1979). Individual differences in second language acquisition. In C. Fillmore., K. Kempler., & D. Wang (Eds.), *Individual Differences in Language Ability and Language Behavior* (pp. 203-228). London, UK: Academic Press.

Wood, D. (2015). *Fundamentals of Formulaic Language: An Introduction*. London, UK; New York, USA: Bloomsbury.

Wray, A. (2002). *Formulaic Language and the Lexicon*. Cambridge, UK: Cambridge University Press.

Wray, A. (2004). "Here's one I prepared earlier": Formulaic language learning on television. In N. Schmitt (Ed.), *Formulaic Sequences: Acquisition, Processing and Use* (pp. 249–268). Amsterdam, The Netherlands; Philadelphia, USA: John Benjamins.

Wray, A. (2008). *Formulaic Language: Pushing the Boundaries*. Oxford, UK: Oxford University Press.

Wu, X., Anderson, R. C., Li, W., Wu, X., Li, H., Zhang, J., Zheng, Q., Zhu, J., Shu, H., Jiang, W., Chen, X., Wang, Q., Yin, L., He, Y., Packard, J., & Gaffney, J. S. (2009). Morphological awareness and Chinese children's literacy development: An intervention study. *Scientific Studies of Reading, 13 (1)*, 26-52.

Xu, J. (2015). Corpus-based Chinese studies: A historical review from the 1920s to the present. *Chinese Language and Discourse, 6 (2)*, 218-244.

Xun, E., Rao, G., Xiao, X., & Zang, J. (2016). The construction of the BCC corpus in the age of big data. *Corpus Linguistics, 3 (1)*, 93-118.

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.

Yan, G., Tian, H., Bai, X., & Rayner, K. (2006). The effect of word and character frequency on the eye movements of Chinese readers. *British Journal of Psychology, 97*, 259–268.

Yan, M., Zhou, W., Shu, H., & Kliegl, R. (2015). Perceptual span depends on font size during the reading of Chinese sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition, 41 (1)*, 209-219.

Yang, S-Y., Ahn, J. S., & Van Lancker Sidtis, D. (2015). The perceptual and acoustic characteristics of Korean idiomatic and literal sentences. *Speech, Language and Hearing, 18 (3)*, 166-178.

Yeganehjoo, M., & Thai, Y. N. (2012). Lexical access in production of idioms by proficient L2 learners. *The Southeast Asian Journal of English Language Studies, 18 (1)*, 87-104.

Yi, W. (2018). Statistical sensitivity, cognitive attitudes, and processing of collocations. *Studies in Second Language Acquisition, 40 (4)*, 831-856.

Yi, W., Lu, S., & Ma, G. (2017). Frequency, contingency and online processing of multiword sequences: An eye-tracking study. *Second Language Research, 33 (4)*, 519-549.

Yu, L., Cutter, M., Yan, G., Bai, X., Fu, Y., Drieghe, D., & Liversedge, S. P. (2016). Word n+2 preview effects in three-character Chinese idioms and phrases. *Language, Cognition and Neuroscience, 31 (9)*, 1130-1149.

Zhang, H., Yang, Y., Gu, J., & Ji, F. (2013). ERP correlates of compositionality in Chinese idiom comprehension. *Journal of Neurolinguistics*, *26*, 89-112.

Zhang, X. (2016). Computational Linguistics. In S-W. Chan., F. Li Wing Yee., & J. Minett (Eds.), *The Routledge Encyclopedia of the Chinese Language* (pp. 420-437). Oxon, UK; New York, USA: Routledge.

Zhao, L., Yasunaga, D., & Kojima, H. (2019). Similarities and differences between native and non-native speakers' processing of formulaic sequences: A functional Near-Infrared Spectroscopy (fNIRS) study. *Journal of Psycholinguistics Research*. Retrieved from https://doi.org/10.1007/s10936-019-09655-w

Zhou, J., Ma, G., Li, X., & Taft, M. (2017). The time course of incremental word processing during Chinese reading. *Reading and Writing, 31 (3)*, 607-625.

Zhou, S., Zhou, W., & Chen, X. (2004). Spatiotemporal analysis of ERP during Chinese idiom comprehension. *Brain Topography, 17 (1)*, 27-37.

Zhou, X., & Marslen-Wilson, W. D. (2000). Lexical representation of compound words: Cross-linguistic evidence. *Psychologia*, *43*, 47-66.

现代汉语常用词表课题组 "Research Group for the Lexicon of Common Words in Contemporary Chinese". (2008). *《现代汉语常用词表》Lexicon of Common Words in Contemporary Chinese*. Beijing, China: The Commercial Press.

Appendices:

Appendix 1: Examples of target items

Item	Phrase type	Part of speech	Insertion length	Phrase length	Phrase frequency	Word 1 frequency	Final-word frequency	Association strength	Predictability
白衬衫	Collocation	A+N	Original	3	1411	85388	17441	0.00	0.30
黄衬衫	Control	A+N	Original	3	26	47373	17441	0.00	0.10
白色长衬衫	Collocation	A+N	Short insertion	5		85388	17441		0.00
黄色长衬衫	Control	A+N	Short insertion	5		47373	17441		0.00
白格子长袖衬衫	Collocation	A+N	Long insertion	7		85388	17441		0.10
黄格子长袖衬衫	Control	A+N	Long insertion	7		47373	17441		0.30
复杂的问题	Collocation	A+N	Original	5	1129	31311	135743	0.00	0.00
特别的问题	Control	A+N	Original	5	14	20567	135743	0.00	0.00
复杂的环境问题	Collocation	A+N	Short insertion	7		31311	135743		0.10
特别的环境问题	Control	A+N	Short insertion	7		20567	135743		0.00
复杂的社会安全问题	Collocation	A+N	Long insertion	9		31311	135743		0.80
特别的社会安全问题	Control	A+N	Long insertion	9		20567	135743		0.80
科学的方法	Collocation	A+N	Original	5	1022	36703	444926	0.00	0.00
奇怪的方法	Control	A+N	Original	5	6	26301	444926	0.00	0.00
科学的解题方法	Collocation	A+N	Short insertion	7		36703	444926		0.80
奇怪的解题方法	Control	A+N	Short insertion	7		26301	444926		1.00
科学而又合适的方法	Collocation	A+N	Long insertion	9		36703	444926		0.30
奇怪而又合适的方法	Control	A+N	Long insertion	9		26301	444926		0.20
可爱的孩子	Collocation	A+N	Original	5	1390	50684	347541	0.10	0.20
美丽的孩子	Control	A+N	Original	5	40	47224	347541	0.10	0.10
可爱活泼的孩子	Collocation	A+N	Short insertion	7		50684	347541		0.30
美丽活泼的孩子	Control	A+N	Short insertion	7		47224	347541		0.00
可爱而又懂事的孩子	Collocation	A+N	Long insertion	9		50684	347541		0.80
美丽而又懂事的孩子	Control	A+N	Long insertion	9		47224	347541		0.70
难得的机会	Collocation	A+N	Original	5	735	12968	197944	0.20	0.20
完美的机会	Control	A+N	Original	5	2	17674	197944	0.00	0.00
难得的进球机会	Collocation	A+N	Short insertion	7		12968	197944		0.90
完美的进球机会	Control	A+N	Short insertion	7		17674	197944		1.00
难得而又重要的机会	Collocation	A+N	Long insertion	9		12968	197944		0.10
完美而又重要的机会	Control	A+N	Long insertion	9		17674	197944		0.00
漂亮的女人	Collocation	A+N	Original	5	1180	34155	266307	0.10	0.10
热情的女人	Control	A+N	Original	5	33	78651	266307	0.10	0.00
漂亮的中国女人	Collocation	A+N	Short insertion	7		34155	266307		0.10
热情的中国女人	Control	A+N	Short insertion	7		78651	266307		0.00
漂亮而又聪明的女人	Collocation	A+N	Long insertion	9		34155	266307		0.00
热情而又聪明的女人	Control	A+N	Long insertion	9		78651	266307		0.10
幸福的生活	Collocation	A+N	Original	5	1393	43175	16717	0.10	0.10
安全的生活	Control	A+N	Original	5	27	24365	16717	0.00	0.00
幸福的童年生活	Collocation	A+N	Short insertion	7		43175	16717		0.10
安全的童年生活	Control	A+N	Short insertion	7		24365	16717		0.30
幸福而又安静的生活	Collocation	A+N	Long insertion	9		43175	16717		0.20
安全而又安静的生活	Control	A+N	Long insertion	9		24365	16717		0.00
严格的要求	Collocation	A+N	Original	5	864	22154	38423	0.10	0.00
一般的要求	Control	A+N	Original	5	22	36043	38423	0.10	0.00
严格的技术要求	Collocation	A+N	Short insertion	7		22154	38423		0.60
一般的技术要求	Control	A+N	Short insertion	7		36043	38423		0.40
严格而又合理的要求	Collocation	A+N	Long insertion	9		22154	38423		0.30
一般而又合理的要求	Control	A+N	Long insertion	9		36043	38423		0.30

Item	Phrase type	Part of speech	Insertion length	Phrase length	Phrase frequency	Word 1 frequency	Final-word frequency	Association strength	Predictability
有利的条件	Collocation	A+N	Original	5	1636	10264	316178	0.10	0.30
方便的条件	Control	A+N	Original	5	39	6483	316178	0.30	0.50
有利的工作条件	Collocation	A+N	Short insertion	7		10264	316178		0.50
方便的工作条件	Control	A+N	Short insertion	7		6483	316178		0.20
有利的学习生活条件	Collocation	A+N	Long insertion	9		10264	316178		0.50
方便的学习生活条件	Control	A+N	Long insertion	9		6483	316178		0.30
周围的环境	Collocation	A+N	Original	5	756	19971	381580	0.30	0.30
正常的环境	Control	A+N	Original	5	49	34443	381580	0.00	0.00
周围的社会环境	Collocation	A+N	Short insertion	7		19971	381580		0.60
正常的社会环境	Control	A+N	Short insertion	7		34443	381580		0.20
周围的学习生活环境	Collocation	A+N	Long insertion	9		19971	381580		0.80
正常的学习生活环境	Control	A+N	Long insertion	9		34443	381580		0.80
包饺子	Collocation	V+N	Original	3	2262	46509	17405	0.10	0.30
卖饺子	Control	V+N	Original	3	30	48613	17405	0.00	0.00
包虾仁饺子	Collocation	V+N	Short insertion	5		46509	17405		1.00
卖虾仁饺子	Control	V+N	Short insertion	5		48613	17405		1.00
包肉馅儿的饺子	Collocation	V+N	Long insertion	7		46509	17405		0.90
卖肉馅儿的饺子	Control	V+N	Long insertion	7		48613	17405		0.90
保护动物	Collocation	V+N	Original	4	1610	30578	120985	0.20	0.10
了解动物	Control	V+N	Original	4	9	54646	120985	0.00	0.00
保护这些动物	Collocation	V+N	Short insertion	6		30578	120985		0.50
了解这些动物	Control	V+N	Short insertion	6		54646	120985		0.10
保护这里的小动物	Collocation	V+N	Long insertion	8		30578	120985		0.90
了解这里的小动物	Control	V+N	Long insertion	8		54646	120985		0.60
参加会议	Collocation	V+N	Original	4	3693	75426	312908	0.00	0.00
准备会议	Control	V+N	Original	4	5	67322	312908	0.10	0.00
参加国际会议	Collocation	V+N	Short insertion	6		75426	312908		0.60
准备国际会议	Control	V+N	Short insertion	6		67322	312908		0.30
参加一个重要会议	Collocation	V+N	Long insertion	8		75426	312908		0.40
准备一个重要会议	Control	V+N	Long insertion	8		67322	312908		0.30
吃糖	Collocation	V+N	Original	2	2629	279736	124117	0.00	0.00
分糖	Control	V+N	Original	2	50	73557	124117	0.00	0.00
吃块奶糖	Collocation	V+N	Short insertion	4		279736	124117		0.10
分块奶糖	Control	V+N	Short insertion	4		73557	124117		0.00
吃这些水果糖	Collocation	V+N	Long insertion	6		279736	124117		0.00
分这些水果糖	Control	V+N	Long insertion	6		73557	124117		0.00
戴眼镜	Collocation	V+N	Original	3	3407	113172	32457	0.00	0.00
做眼镜	Control	V+N	Original	3	2	379365	32457	0.00	0.00
戴黑色眼镜	Collocation	V+N	Short insertion	5		113172	32457		0.00
做黑色眼镜	Control	V+N	Short insertion	5		379365	32457		0.00
戴黑框近视眼镜	Collocation	V+N	Long insertion	7		113172	32457		1.00
做黑框近视眼镜	Control	V+N	Long insertion	7		379365	32457		1.00
发出声音	Collocation	V+N	Original	4	1434	25583	224843	0.00	0.30
改变声音	Control	V+N	Original	4	21	38996	224843	0.00	0.10
发出一些声音	Collocation	V+N	Short insertion	6		25583	224843		0.30
改变一些声音	Control	V+N	Short insertion	6		38996	224843		0.00
发出这奇怪的声音	Collocation	V+N	Long insertion	8		25583	224843		0.60
改变这奇怪的声音	Control	V+N	Long insertion	8		38996	224843		0.20
拍照片	Collocation	V+N	Original	3	1196	47533	126048	0.00	0.00
送照片	Control	V+N	Original	3	23	96094	126048	0.00	0.00
拍一张照片	Collocation	V+N	Short insertion	5		47533	126048		0.80
送一张照片	Control	V+N	Short insertion	5		96094	126048		0.20
拍些好看的照片	Collocation	V+N	Long insertion	7		47533	126048		0.10
送些好看的照片	Control	V+N	Long insertion	7		96094	126048		0.00

Item	Phrase type	Part of speech	Insertion length	Phrase length	Phrase frequency	Word 1 frequency	Final-word frequency	Association strength	Predictability
借书	Collocation	V+N	Original	2	1577	24132	651055	0.00	0.10
扔书	Control	V+N	Original	2	21	48029	651055	0.00	0.00
借语文书	Collocation	V+N	Short insertion	4		24132	651055		0.10
扔语文书	Control	V+N	Short insertion	4		48029	651055		0.20
借英语写作书	Collocation	V+N	Long insertion	6		24132	651055		0.20
扔英语写作书	Control	V+N	Long insertion	6		48029	651055		0.50
看电影	Collocation	V+N	Original	3	22245	350890	202788	0.00	0.10
讲电影	Control	V+N	Original	3	18	505347	202788	0.00	0.00
看一部电影	Collocation	V+N	Short insertion	5		350890	202788		0.70
讲一部电影	Control	V+N	Short insertion	5		505347	202788		0.70
看了一上午电影	Collocation	V+N	Long insertion	7		350890	202788		0.20
讲了一上午电影	Control	V+N	Long insertion	7		505347	202788		0.00
坐飞机	Collocation	V+N	Original	3	2582	444030	80329	0.00	0.00
修飞机	Control	V+N	Original	3	17	358173	80329	0.00	0.00
坐一次飞机	Collocation	V+N	Short insertion	5		444030	80329		0.00
修一次飞机	Control	V+N	Short insertion	5		358173	80329		0.00
坐那架真的飞机	Collocation	V+N	Long insertion	7		444030	80329		0.60
修那架真的飞机	Control	V+N	Long insertion	7		358173	80329		0.90

Notes. "A+N" refers to an Adjective+Noun phrase; "V+N" refers to a Verb+Noun phrase. Phrase length is calculated by the number of Chinese characters. All frequency information is retrieved from the BCC corpus (BLCU Corpus Center, 2016). The maximum association strength is 1.00; the maximum predictability is 1.00.

Appendix 2: Examples of target sentences

Sentence	Phrase type	Part of speech	Insertion length	Sentence length	Naturalness
这件白衬衫是他哥哥的。	Collocation	A+N	Original	10	10.00
这件黄衬衫是他哥哥的。	Control	A+N	Original	10	10.00
这件白色长衬衫是他哥哥的。	Collocation	A+N	Short insertion	12	9.40
这件黄色长衬衫是他哥哥的。	Control	A+N	Short insertion	12	9.50
这件白格子长袖衬衫是他哥哥的。	Collocation	A+N	Long insertion	14	8.50
这件黄格子长袖衬衫是他哥哥的。	Control	A+N	Long insertion	14	9.50
这真是一个复杂的问题啊。	Collocation	A+N	Original	11	8.80
这真是一个特别的问题啊。	Control	A+N	Original	11	8.90
这真是一个复杂的环境问题啊。	Collocation	A+N	Short insertion	13	8.00
这真是一个特别的环境问题啊。	Control	A+N	Short insertion	13	7.30
这真是一个复杂的社会安全问题啊。	Collocation	A+N	Long insertion	15	8.50
这真是一个特别的社会安全问题啊。	Control	A+N	Long insertion	15	7.10
这个科学的方法应该是他想出来的吧。	Collocation	A+N	Original	16	8.90
这个奇怪的方法应该是他想出来的吧。	Control	A+N	Original	16	9.75
这个科学的解题方法应该是他想出来的吧。	Collocation	A+N	Short insertion	18	8.70
这个奇怪的解题方法应该是他想出来的吧。	Control	A+N	Short insertion	18	9.10
这个科学而又合适的方法应该是他想出来的吧。	Collocation	A+N	Long insertion	20	7.40
这个奇怪而又合适的方法应该是他想出来的吧。	Control	A+N	Long insertion	20	7.70
这个可爱的孩子是我朋友的女儿。	Collocation	A+N	Original	14	10.00
这个美丽的孩子是我朋友的女儿。	Control	A+N	Original	14	10.00
这个可爱活泼的孩子是我朋友的女儿。	Collocation	A+N	Short insertion	16	9.20
这个美丽活泼的孩子是我朋友的女儿。	Control	A+N	Short insertion	16	9.35
这个可爱而又懂事的孩子是我朋友的女儿。	Collocation	A+N	Long insertion	18	8.90
这个美丽而又懂事的孩子是我朋友的女儿。	Control	A+N	Long insertion	18	9.30
这是一次难得的机会你不要错过了。	Collocation	A+N	Original	15	9.20
这是一次完美的机会你不要错过了。	Control	A+N	Original	15	8.55
这是一次难得的进球机会你不要错过了。	Collocation	A+N	Short insertion	17	8.45
这是一次完美的进球机会你不要错过了。	Control	A+N	Short insertion	17	9.30
这是一次难得而又重要的机会你不要错过了。	Collocation	A+N	Long insertion	19	7.10
这是一次完美而又重要的机会你不要错过了。	Control	A+N	Long insertion	19	6.10
这个漂亮的女人是我的朋友。	Collocation	A+N	Original	12	10.00
这个热情的女人是我的朋友。	Control	A+N	Original	12	9.85
这个漂亮的中国女人是我的朋友。	Collocation	A+N	Short insertion	14	8.80
这个热情的中国女人是我的朋友。	Control	A+N	Short insertion	14	9.10
这个漂亮而又聪明的女人是我的朋友。	Collocation	A+N	Long insertion	16	8.30
这个热情而又聪明的女人是我的朋友。	Control	A+N	Long insertion	16	8.70
像这样幸福的生活她永远不会忘记。	Collocation	A+N	Original	15	8.20
像这样安全的生活她永远不会忘记。	Control	A+N	Original	15	7.85
像这样幸福的童年生活她永远不会忘记。	Collocation	A+N	Short insertion	17	8.50
像这样安全的童年生活她永远不会忘记。	Control	A+N	Short insertion	17	7.55
像这样幸福而又安静的生活她永远不会忘记。	Collocation	A+N	Long insertion	19	8.80
像这样安全而又安静的生活她永远不会忘记。	Control	A+N	Long insertion	19	7.70
这些严格的要求其实是非常必要的。	Collocation	A+N	Original	15	8.40
这些一般的要求其实是非常必要的。	Control	A+N	Original	15	7.65

Sentence	Phrase type	Part of speech	Insertion length	Sentence length	Naturalness
这些严格的技术要求其实是非常必要的。	Collocation	A+N	Short insertion	17	8.30
这些一般的技术要求其实是非常必要的。	Control	A+N	Short insertion	17	7.25
这些严格而又合理的要求其实是非常必要的。	Collocation	A+N	Long insertion	19	8.10
这些一般而又合理的要求其实是非常必要的。	Control	A+N	Long insertion	19	6.75
这么有利的条件你要好好利用啊。	Collocation	A+N	Original	14	8.20
这么方便的条件你要好好利用啊。	Control	A+N	Original	14	8.45
这么有利的工作条件你要好好利用啊。	Collocation	A+N	Short insertion	16	8.70
这么方便的工作条件你要好好利用啊。	Control	A+N	Short insertion	16	8.13
这么有利的学习生活条件你要好好利用啊。	Collocation	A+N	Long insertion	18	8.40
这么方便的学习生活条件你要好好利用啊。	Control	A+N	Long insertion	18	8.60
这跟我们周围的环境不太一样。	Collocation	A+N	Original	13	8.25
这跟我们正常的环境不太一样。	Control	A+N	Original	13	7.80
这跟我们周围的社会环境不太一样。	Collocation	A+N	Short insertion	15	8.70
这跟我们正常的社会环境不太一样。	Control	A+N	Short insertion	15	8.10
这跟我们周围的学习生活环境不太一样。	Collocation	A+N	Long insertion	17	8.90
这跟我们正常的学习生活环境不太一样。	Control	A+N	Long insertion	17	8.90
我们正在包饺子呢。	Collocation	V+N	Original	8	9.50
我们正在卖饺子呢。	Control	V+N	Original	8	9.15
我们正在包虾仁饺子呢。	Collocation	V+N	Short insertion	10	8.20
我们正在卖虾仁饺子呢。	Control	V+N	Short insertion	10	8.00
我们正在包肉馅儿的饺子呢。	Collocation	V+N	Long insertion	12	8.70
我们正在卖肉馅儿的饺子呢。	Control	V+N	Long insertion	12	8.10
我们应该保护动物并跟它们成为朋友。	Collocation	V+N	Original	16	9.35
我们应该了解动物并跟它们成为朋友。	Control	V+N	Original	16	9.10
我们应该保护这些动物并跟它们成为朋友。	Collocation	V+N	Short insertion	18	8.20
我们应该了解这些动物并跟它们成为朋友。	Control	V+N	Short insertion	18	8.40
我们应该保护这里的小动物并跟它们成为朋友。	Collocation	V+N	Long insertion	20	9.50
我们应该了解这里的小动物并跟它们成为朋友。	Control	V+N	Long insertion	20	8.75
这个周末我要参加会议就不去找你了。	Collocation	V+N	Original	16	8.75
这个周末我要准备会议就不去找你了。	Control	V+N	Original	16	9.25
这个周末我要参加国际会议就不去找你了。	Collocation	V+N	Short insertion	18	9.20
这个周末我要准备国际会议就不去找你了。	Control	V+N	Short insertion	18	9.45
这个周末我要参加一个重要会议就不去找你了。	Collocation	V+N	Long insertion	20	8.50
这个周末我要准备一个重要会议就不去找你了。	Control	V+N	Long insertion	20	8.30
我们一起吃糖吧。	Collocation	V+N	Original	7	9.50
我们一起分糖吧。	Control	V+N	Original	7	8.65
我们一起吃块奶糖吧。	Collocation	V+N	Short insertion	9	8.30
我们一起分块奶糖吧。	Control	V+N	Short insertion	9	8.05
我们一起吃这些水果糖吧。	Collocation	V+N	Long insertion	11	9.40
我们一起分这些水果糖吧。	Control	V+N	Long insertion	11	9.50
我认识那个戴眼镜的人。	Collocation	V+N	Original	10	9.45
我认识那个做眼镜的人。	Control	V+N	Original	10	8.20
我认识那个戴黑色眼镜的人。	Collocation	V+N	Short insertion	12	9.25
我认识那个做黑色眼镜的人。	Control	V+N	Short insertion	12	7.30
我认识那个戴黑框近视眼镜的人。	Collocation	V+N	Long insertion	14	9.00
我认识那个做黑框近视眼镜的人。	Control	V+N	Long insertion	14	8.13

Sentence	Phrase type	Part of speech	Insertion length	Sentence length	Naturalness
我猜这样做可以让它发出声音吧。	Collocation	V+N	Original	14	8.25
我猜这样做可以让它改变声音吧。	Control	V+N	Original	14	8.10
我猜这样做可以让它发出一些声音吧。	Collocation	V+N	Short insertion	16	8.20
我猜这样做可以让它改变一些声音吧。	Control	V+N	Short insertion	16	7.60
我猜这样做可以让它发出这奇怪的声音吧。	Collocation	V+N	Long insertion	18	8.50
我猜这样做可以让它改变这奇怪的声音吧。	Control	V+N	Long insertion	18	7.45
我正准备给你拍照片呢。	Collocation	V+N	Original	10	9.45
我正准备给你送照片呢。	Control	V+N	Original	10	9.20
我正准备给你拍一张照片呢。	Collocation	V+N	Short insertion	12	9.60
我正准备给你送一张照片呢。	Control	V+N	Short insertion	12	9.20
我正准备给你拍些好看的照片呢。	Collocation	V+N	Long insertion	14	8.50
我正准备给你送些好看的照片呢。	Control	V+N	Long insertion	14	8.30
那个总是借书的人又来了。	Collocation	V+N	Original	11	8.10
那个总是扔书的人又来了。	Control	V+N	Original	11	7.75
那个总是借语文书的人又来了。	Collocation	V+N	Short insertion	13	8.38
那个总是扔语文书的人又来了。	Control	V+N	Short insertion	13	7.90
那个总是借英语写作书的人又来了。	Collocation	V+N	Long insertion	15	9.40
那个总是扔英语写作书的人又来了。	Control	V+N	Long insertion	15	8.60
我们在这里看电影还没结束呢。	Collocation	V+N	Original	13	8.40
我们在这里讲电影还没结束呢。	Control	V+N	Original	13	7.85
我们在这里看一部电影还没结束呢。	Collocation	V+N	Short insertion	15	8.20
我们在这里讲一部电影还没结束呢。	Control	V+N	Short insertion	15	7.60
我们在这里看了一上午电影还没结束呢。	Collocation	V+N	Long insertion	17	8.10
我们在这里讲了一上午电影还没结束呢。	Control	V+N	Long insertion	17	7.40
我要是能坐飞机就好了。	Collocation	V+N	Original	10	9.10
我要是能修飞机就好了。	Control	V+N	Original	10	9.80
我要是能坐一次飞机就好了。	Collocation	V+N	Short insertion	12	8.30
我要是能修一次飞机就好了。	Control	V+N	Short insertion	12	7.65
我要是能坐那架真的飞机就好了。	Collocation	V+N	Long insertion	14	7.80
我要是能修那架真的飞机就好了。	Control	V+N	Long insertion	14	8.30

Notes. "A+N" refers to an Adjective+Noun phrase; "V+N" refers to a Verb+Noun phrase. Sentence length is calculated by the number of Chinese characters. The maximum naturalness is 10.00.