Are Chinese Correlative Conjunctions Psychologically Real? An Investigation of the Combination Frequency Effect

Psychological Reports
2016, Vol. 119(1) 106–123

DOI: 10.1177/0033294116658631
prx.sagepub.com

Lingyue Kong
Peking University, China

John X. Zhang
Fudan University, China

Yongwei Zhang
The Chinese Academy of Social Sciences, China

Abstract

The present study used an online grammaticality judgment task to examine whether Chinese discontinuous correlative conjunctions are psychologically real in mental lexicon. High- and low-frequency discontinuous correlative conjunctions were compared with random combinations differing in combination frequencies but matched for constituent word frequency. Forty graduate students participated in the study. Results showed that responses were faster and more accurate for high-frequency correlative conjunctions than low-frequency ones, but the effects were absent for random combinations. The results indicate that Chinese discontinuous correlative conjunctions have psychological reality in mental lexicon in addition to the representation of their constituent words, and that grammatical functions of correlative conjunctions may be a critical factor for the formation of such holistic representations.

Keywords

formulaic language, correlative conjunction, combination frequency, psychological reality

Corresponding Author:
Lingyue Kong, School of Chinese as a Second Language, Peking University, Beijing 100871, China.
Email: lingyue@pku.edu.cn

Introduction

Lexical representation and lexical processing is a central topic in word recognition research. However, attention had been paid mainly to single words. One basic research finding is that there are lexical representations of words in one's mental lexicon and word frequency affects lexical processing. The greater the frequency, the faster the processing of a word (Balota & Chumbley, 1990; Seidenberg, 1985). Driven by the latest findings of formulaic languages ubiquity in natural speech revealed by corpus linguistics (Conklin & Schmitt, 2012), and the development of linguistic theories tied to language acquisition and representation like the emergentist models including exemplar- or usage-based models of linguistic knowledge and representations (Bod, 1998, 2006; Bybee, 1998; Goldberg, 1995, 2006; Langacker, 1988; Pierrehumbert, 2001; Tomasello, 2003), recent years has seen increased interest on multi-word units. These units are generally termed formulaic languages including binomials (i.e., *bride and groom*), lexical bundles (i.e., *don't worry about that*), or compositional phrases (i.e., *don't have to worry*).

Emergentist models proposed that language consists of a set of constructions varying in size and level of abstractness and that the basic unit of language acquisition is construction. Language acquisition is a process in which variety of constructions are acquired (Goldberg, 2006; Tomasello, 2003). Language learners are sensitive to the statistical properties of a linguistic unit. The representation and processing of all linguistic units are subject to the effect of their statistical properties such as frequency. These findings and related theories are of importance for understanding the nature of linguistic processing and language acquisition.

As far as multi-word units are concerned, one critical issue is whether multi-word units have holistic representations (stored and retrieved as a single unit) in mental lexicon besides their constituent words as single word or whether the statistical property like frequency of a multi-word unit affects its processing.

At present, there are only a few studies on multi-word units such as two-word combinations (e.g., Bell et al., 2003; De Cat, Klepousniotou, & Baayen, 2015; Durrant & Doherty, 2010; Kapatsinski & Radicke, 2008; Mondini, Jarema, Luzzatti, Burani, & Semenza, 2002; Sosa & MacFarlane, 2002; Wolter & Yamashita, 2015) and three- to five-word sequences (Arcara et al., 2012; Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Bannard & Matthews, 2008; Cacciari, 2014; Columbus, 2010; Holsinger, 2013; Reali & Christiansen, 2007; Siyanova-Chanturia, Conklin, & van Heuven, 2011; Tremblay & Baayen, 2010; Zhang, Yang, Gu, & Ji, 2013). These studies in general offer some limited evidence that multi-word units have holistic representations in the mental lexicon and the processing of such units is affected by phrasal frequency. For example, Sosa and MacFarlane (2002) asked participants to monitor the word *of* in two-word collocations with high frequency (e.g., *sort of*) or low frequencies (e.g., *colleague of*) when hearing sentences. Their hypothesis was that if high-frequency, two-word sequences were stored holistically like single words, no composition processing would occur in retrieving the sequence, and thus access to their constituent words would be hindered by the holistic processing in the monitoring task. But such interference effects should not be present in monitoring low-frequency, two-word sequences without holistic representation. The results supported their hypothesis. Sosa and MacFarlane found that participants' reaction time to *of* in high-frequency sequences was significantly slower than that in low-frequency sequences. With the same task, Kapatsinski and Radicke (2008) obtained similar results on sequences such as *give up*.

Several researchers examined compounds (e.g., chalkboard and blueberry), but the findings were inconsistent. For example, Juhasz (2007) argued that compounds are firstly decomposed into their individual components (i.e., *chalk* and *board* in *chalkboard*) during the initial stage of word recognition and then retrieved based on the access of constituents, suggesting that only the representations of their individual words are stored in mental lexicon. Other researchers proposed that both compounds and their constituents can be stored in mental lexicon. Compounds can be retrieved and accessed by two parallel routes, the online computation via individual words and the direct retrieval via the holistic representation. For example, Mondini et al. (2002) examined the processing of Italian two-word compounds (*natura morta /still life*) and novel combinations *(natura bella / beautiful nature*) by two non-fluent aphasic patients. They found that the patients performed significantly better on compounds than on novel combinations. The findings indicated that for novel combinations, their constituents were first retrieved separately and then assembled online via the agreement rules by patients. However, for compounds, they were retrieved as a whole without any morphosyntactic operations. Such parallel processing in compounds was also supported by eye-tracking studies (Kuperman, Schreuder, Bertram, & Baayen, 2009).

In recent years, researchers have used more rigorous experimental designs to examine the representation issues of four-word sequences in children and adults (Arnon & Snider, 2010; Bannard & Matthews, 2008). For example, Bannard and Matthews (2008) investigated how two- and three-year-old children repeated the frequent four-word sequences. In their study, the frequent (e.g., *sit in your chair*) and infrequent (e.g., *sit in your truck*) sequences were different only at the final word. That is, they differed on phrase frequencies but were identical in other aspects. Researchers let children hear the utterance of four-word sequences and then asked them to repeat what they heard. They found that compared with the processing of low-frequency sequences, children's responses to high-frequency sequences were faster and more accurate. The finding indicated that the four-word sequences were stored as a unit, even in two-year-old children. Arnon and Snider (2010) further investigated the processing of four-word sequences in adults and replicated the findings in children.

Stronger evidence for the holistic representations of multi-word units came from Tremblay and Baayen (2010) who examined the event-related potential (ERP) effect in processing four-word bundles presented all at once (e.g., *in the middle of*). The key finding was that there were ERP effects for bundles within 110–150 ms after stimuli onset and the effects on P1, N1, and P2 were modulated by the phrase frequency of four-word bundles. P1, N1, and P2 were three early ERPs with different peaks within 200 ms after stimuli onset, which are usually associated with attentional processes (and the P2 has also been associated with stimulus probability, expectancy, and frequency). Because lexical access occurred in the initial 200 ms period of word recognition, and the ERP effects on N1, P1, and P2 were within this time window, it appeared that the four-word bundles were stored holistically. Otherwise, no such early ERP effects would have been observed.

Some researchers also used eye-tracking paradigm to explore the processing of three-word binomial phrases in both L1 and L2 learners. Siyanova-Chanturia et al. (2011) investigated the effects of binomial phrase frequency on their processing when both native and nonnative learners read sentences containing binomials (e.g., *bride and groom*) and their reversal form (*groom and bride*). The semantics and syntactic were the same in both binominal and the reversal form. Results showed that both native and nonnative readers could read high-frequency binominals more quickly. No matter what the phrase frequency was, the processing of binominals was significantly faster than that of the reversal form. Moreover, the processing advantage was argued not to be attributed to predictability (e.g., the first two words *bride and* would predict the last word *groom* in the *bride and groom*, but *groom and* would not predict *bride* in the reversal form *groom and bride*). These results suggested that binominals have been entrenched in memory and that both phrase frequency and the configuration of binominals affect processing.

Evidently, all the aforementioned studies focused on continuous multi-word units such as collocations, binominals, or lexical bundles. It is unclear whether findings from continuous multi-word units are able to apply to discontinuous multi-word units with slots or frames (e.g., *as far as . . . concerned, 因为 . . . 所以 . . . /because . . . therefore . . .*). Are they psychologically real? That is, are there holistic representations for the discontinuous multi-word units in mental lexicon in addition to the representations of the constituent words? Some early work seemed to imply that there might be strong association between the constituents of discontinuous word sequences, although it could not directly answer the issue of holistic representations. For example, Schcaneveldt and Meyer (1973) conducted a lexical decision study in which they interposed an unrelated category between the related prime sequence and target (e.g., DOCTOR PAPER NURSE) and had participants make a "yes" response if all three stimuli were words and a "no" response otherwise. With this procedure, priming still occurred even though the participants were sequentially processing the three simultaneously presented items. The present study was intended to provide direct evidence regarding this issue using discontinuous Chinese correlative conjunctions.

Historically, the Chinese writing system has been viewed as ideographic because a character, the basic writing unit, maps onto a morpheme. Chinese is unique in several aspects compared with other languages such as English. For example, homophonic characters abound in Chinese and above 85% of Chinese characters are compounds, with a phonetic and semantic radical providing cue for the pronunciation or meanings of the whole character. Nevertheless, researchers argued that it may be more appropriate to view the Chinese writing system as morphemic (e.g., Leong, 1973) or morphosyllabic (e.g., DeFrancis, 1989; Mattingly, 1992), because a character often maps onto a single syllable morpheme in the spoken language. In other words, the Chinese writing system is a combination of sounds and meaning. Furthermore, studies have shown that both Chinese compound words and their constituents can be stored in mental lexicon (Taft, Liu, & Zhu, 1999). The present study used discontinuous Chinese correlative conjunctions as materials to further explore the issues on lexical processing and representations of Chinese multi-word units in addition to compound words.

The materials used were high- and low-frequency Chinese correlative conjunctions and random conjunction combinations in discontinuous form. In Chinese, correlative conjunction is one of the types of conjunctions that are quite familiar for native speakers (Chao, 1970, p. 792). Such conjunctions are pairs of correlative words which serve to bind clauses together into compound or complex sentences (e.g., *因为 . . . 所以 . . . /because . . . therefore . . .*; *虽然 . . . 也 . . . /although . . . yet . . .*). In other words, correlative conjunctions possess clear grammatical function.

In most cases, the correlatives are adverbial conjunctions and can either precede or follow the subject. Sometimes, either the first or the second of the correlatives is adverb only and never precedes the subject, or is conjunction only and never follows the subject. For example, in a complex sentence bound by *既不 . . . 又不 . . . /neither . . . nor . . .* (他既不聪明，又不勇敢，但是他总能在危险的时候帮助主角脱离险境./*He is neither smart nor brave, but he is always there to help out in times of danger*), the two adverbs never precede the subject. Contrarily, in a complex sentence bound by *尚且 . . . 何况 . . . /even . . . how much more will- . . .* (大人尚且举不起来,何况小孩子/*Even adults still cannot lift it, how much more will a child do it*), the second word, the conjunction "何况," cannot follow the subject.

In Chinese, frequent correlative conjunctions are considered formulaic sequences assumed to be stored holistically, although direct evidence is still lacking (Xue & Shi, 2013). Random conjunction combinations here refer to two-word pairs of conjunctions, which are formed arbitrarily by two conjunction words and do not have grammatical functions like correlative conjunctions (e.g., *虽然 . . . 还是 . . . /although . . . still . . .*). In the corpus, a random combination may occur in the same sentence and thus their total occurrences can be counted by a certain extraction method (e.g., *虽然 . . . 还是* . . . in the sentence "虽然人还是有的，但成为人物的机会就没有了/ *although* there are *still* some people, (but) there is no chance for them to become a big character").

To obtain the frequency information, we extracted all two-word conjunction pairs from the corpus, which included the correlative conjunctions and random conjunction combinations. Then, the total occurrences of a conjunction pair in the corpus would be counted as its combination frequency. Specifically, a pair of conjunctions that occurred in a sentence that ended with a period, a question mark, or an exclamation mark would be extracted and counted. Note that corpus-extracted conjunction pairs definitely contained grammatical correct conjunctions and random conjunction combinations without correct grammatical function. Thus, to avoid any misunderstanding of the corpus-extracted pairs, the term *random combinations* here refers in particular to the entire corpus-extracted, arbitrary, two-word conjunction pairs except the grammatical correlative conjunctions. Such random combinations are not familiar to native speakers compared with correlative conjunctions. The reason for the inclusion of random conjunction combinations was that it was unclear if uncommon, corpus-extracted, multi-word units were psychologically real (Durrant & Doherty, 2010; Schmitt, Grandage, & Adolphs, 2004). Examining discontinuous random combinations would provide more insights into the representation issue.

In the present design, combination frequencies (i.e., the total occurrences of a conjunction pair in corpus) differed but the frequencies of their constituents were matched between high- and low-frequency grammatical and random combinations. In the experimental task, participants were required to judge whether or not a conjunction pair was a correct correlative conjunction that can bind two causes into a complex sentences. The difference between the reaction time of high- and low-frequency conjunction pairs would be the key result to examine.

Participants can judge whether or not a conjunction pair is a correlative conjunction in several ways (Pollatsek, Hyona, & Bertram, 2000). One way is to access individual components first and then assemble them by grammatical rules. If the online computation is able to complete, a *Yes* response would be elicited. Otherwise, a *No* response would be elicited. For example, in judging the correlative conjunction "*因为 . . . 所以 . . . /because . . . therefore . . .*," *因为* and *所以* would be activated and then can be assembled via the rule of *cause and effect*. Participants would judge it as a correct correlative conjunction and make a *Yes* response by pressing the corresponding key. Another way is to activate the conjunction pair as a single unit. Successful retrieval would elicit a *Yes* response, and otherwise, a *No* response. The third way is that the above two ways could be used in parallel, although the direct way would be faster and more effective than the computed one. Effect of combination frequency in processing correlative conjunctions would still be present for such parallel processing. Based on these considerations, we proposed the following hypothesis.

Hypothesis 1. If only online computation exists, there would be no difference in processing time between the processing of combinations with high and low combination frequency.

Hypothesis 2. If combination frequency effect was observed, it would indicate that correlative conjunctions or random conjunction combinations may have holistic representations in mental lexicon and they can be accessed rapidly.

Method

Participants

Forty graduate students (M age = 23.2 years, SD = 1.4; 32 females) took part in the study. All were students at Peking university in Beijing. Students were informed of the experimental information in several classes by the authors and then they could contact the authors directly by email or phone if they were willing to attend. All participants were paid a nominal sum for their participation, after the experiment.

Materials

All materials were selected from Modern Chinese Corpus of State Language Commission (totaling 20 million characters) free for use from Corpus online (Institute of Applied Linguistics of Ministry of Education, 2015). As noted above, combination frequency of a random conjunction pair is the total count that the random conjunction combination occurs in the corpus. The same method was applied to count the combination frequency of correlative conjunctions.

In total, there were four groups of conjunction combinations, 24 in each group (condition). Groups 1 and 2 were administered grammatical correlative conjunctions, one with high combination frequency and the other with low combination frequency; Groups 3 and 4 were random conjunction combinations, one with high combination frequency and the other with low combination frequency. Analyses showed that for the four groups of materials, their combination frequencies differed between the high- and low-frequency combinations both in grammatical and random combination conditions (ps < .005), but the frequencies of constituent words were systematically matched across four groups (ps > .3). Table 1 shows the material exemplar in each group. The full set of materials is shown in the Appendix 1.

Table 1. Material exemplars in four groups.

	Exemplar	Mean combination and mean constituent frequencies in corpus	Mean frequencies in each million character
CCHF	不但…而且… (not only…but also)	571.7	28.6
	不但(not only)	2372.0	118.6
	而且(but also)	18803.0	940.2
CCLF	不但…并且… (not only…but also)	22.2	1.1
	不但(not only)	1998.4	99.9
	并且(moreover)	14080.5	704.0
RCHF	比如…就… (such as…then)	196.1	9.8
	比如(such as)	2686.5	134.3
	就(then)	18333.5	916.7
RCLF	不但…又… (not only…again)	18.5	0.9
	不但(not only)	2285.3	114.3
	又(again)	17898.4	894.9

CCHF: correlative conjunctions with high frequency; CCLF: correlative conjunctions with low frequency; RCHF: random combination with high frequency; RCLF: random combination with low frequency.

Procedure

Materials were divided into four blocks by Latin square. Each block included four kinds of combinations, totally 24 with 6 for each kind. Participants were first given instructions on the computer screen and asked to respond rapidly and accurately by pressing the button (the key *F* for *No* response or *J* for *Yes* response). Participants were required to judge whether or not a combination was a correct correlative conjunction binding two causes when it appeared on the screen. In each trial, a cross was first shown on the center of the screen for 1000 ms and disappeared, and then a combination was exposed in the same position. The combination remained on the screen until participants responded. After the response, the next trial was started. Before the formal experiment, 15 practice trials were given to participants to familiarize them with the procedure.

Table 2. Mean reaction time and accuracy for four groups.

	Reaction time (ms)		Accuracy (%)	
Group	*M*	*SD*	*M*	*SD*
CCHF	1313	336	91	0.05
CCLF	1806	425	63	0.14
RCHF	2125	706	43	0.17
RCLF	2203	450	71	0.14

CCHF: correlative conjunctions with high frequency; CCLF: correlative conjunctions with low frequency; RCHF: random combination with high frequency; RCLF: random combination with low frequency.

In actual experiment, if a participant need more practices, they would be allowed to have more trials of practice.

Results

All reaction time analyses were based on the data for correct responses. The erroneous data and outliers (shorter than 400 ms and larger than 4000 ms) were excluded from the analyses. The missing data accounted for 28.1% of the total data in grammatical condition (540 erroneous data and 95 outlier data) and for 52.3% in ungrammatical condition (824 erroneous data and 181 outlier data). Participants' reaction time and accuracy rates are shown in Table 2.[1]

Two-way repeated analyses of variance (ANOVA) were conducted to analyze the reaction times and accuracy data. The two within-subject factors were grammaticality (grammatical vs. random) and combination frequency (high vs. low frequency). Both participant analysis and item analysis were conducted in the ANOVA with *F1* (participant analysis) and *F2* (item analysis) to be reported.

Analyses on reaction times showed that the main effect of combination frequency was significant by participant and item analysis, $F1(1, 39) = 34.51$, $p < .001$ and $F2(1, 23) = 20.81$, $p < .001$, showing that responses were faster to high combination frequency conjunction pairs than low combination frequency ones. The main effect of grammaticality was also significant, $F1(1, 39) = 101.13$, $p < .001$ and $F2(1, 23) = 51.74$, $p < .001$, showing that responses were faster to grammatical correlative conjunctions than random conjunction combinations. There was also significant interaction effect between grammaticality and combination frequency, $F1(1, 39) = 19.23$, $p < .001$ and $F2(1, 23) = 4.77$, $p < .05$.

Further analyses showed that for grammatical correlative conjunctions, participants' reaction times to high-frequency conjunctions were significantly shorter than to low-frequency conjunctions, *F1*: 1313 ms versus 1806 ms, $p < .001$ and *F2*: 1356 ms versus 1837 ms, $p < .001$. However, for random correlative conjunctions, there was no significant difference between the reaction times of high- and low-frequency combinations, *F1*: 2125 ms versus 2203 ms, $p = .38$ and *F2*: 2048 ms versus 2187 ms, $p = .16$.

The same ANOVA were also carried out to analyze the accuracy data. Results showed that the main effect of combination frequency was not significant, $F1(1, 39) = 0.01$, $p = .93$ and $F2(1, 23) = .001$, $p = .97$. But the main effect of grammaticality was significant, $F1(1, 39) = 45.44$, $p < .001$ and $F2(1, 23) = 23.55$, $p < .001$. Moreover, there were significant interaction effects between the two factors, $F1(1, 39) = 348.64$, $p < .001$ and $F2(1, 23) = 29.19$, $p < .001$.

Further analyses showed that for grammatical correlative conjunctions, accuracy of high-frequency combinations was significantly higher than that of low-frequency combinations (91.1% vs. 62.5%, $p < .001$). But, for random combinations, accuracy of high-frequency combinations was significantly lower than that of low-frequency combinations (42.9% vs.71.2%, $p < .005$). It was clear that there were combination frequency effects in processing correlative conjunctions. Contrarily, there were no such effects for random conjunction combinations in reaction times.

Discussion

The present study examined the processing of correlative conjunctions and corpus-extracted random conjunction combinations by using an online grammatical judgment task. Two major findings were obtained. First, there were combination frequency effects in judging correlative conjunctions. Specifically, responses to correlative conjunctions with high combination frequency were faster and mostly more accurate than those for correlatives with low combination frequency. Moreover, responses to correlative conjunctions were faster and more accurate than those for random combinations.

Second, such effect was observed only for the accuracy rate data but not the reaction times in processing random conjunction combinations. Results showed that the accuracy rates for random combinations with low combination frequency were different from those for combinations with high combination frequency. The different patterns of combination frequency effects indicated that correlative conjunctions are psychologically real, but random conjunction combinations are not. The findings are consistent with existing studies on multi-word units (e.g., Arnon & Cohen Priva, 2013; Arnon & Snider, 2010; Bannard &

Matthews, 2008; De Cat et al., 2015; Siyanova-Chanturia et al., 2011). One limitation of the present study was that overall accuracy rates were low. This may be corrected in future studies with better selection of materials or recruiting participants with natural language processing background.

Previous research showed the existence of holistic representations of continuous multi-word units in mental lexicon so that phrase frequency would affect their processing. The present findings provided further evidence for this conclusion. Due to the holistic representation, participants could access the unit in mental lexicon rapidly and accurately without online computation via grammatical rules. Thus, reaction times were significantly shorter in processing correlative conjunctions. Moreover, responses to the correlatives with high combination frequency were significantly faster than those with low combination frequency.

On the other hand, random conjunction combinations are unfamiliar and do not have holistic representations in the mental lexicon. They could not be retrieved directly from the mental lexicon. The only way to judge them as grammatical or not was to use grammatical rules. Such online computation processes were relatively slow and thus resulted in similar responses and relatively long reaction times to all corpus-extracted random combinations.

For ungrammatical combinations, accuracy of high-frequency combinations was significantly lower than that of low-frequency ones. Reasons for this may be that high-frequency combinations were more likely to judge as "grammatical" than low-frequency combinations when they were retrieved word by word, computed in sequence via a certain grammatical rule in the judgment task, thus lowering the accuracy rates. One example can be used to illustrate this point.

In the sentence "该厂虽然进行了认真治理，但是污染环境问题仍然存在。/ *Although the factory has carried on the earnest management, (however,) environmental pollution problems still exist*". The combination虽然 . . . 但是 . . . was correlative conjunction, but虽然 . . . 仍然 . . . was not. However, participants may be likely to judge the combination 虽然 . . . 仍然 . . . (or even仍然 . . . 但是 . . .) to be a correct correlative conjunction because the conjunction word 仍然 frequently occurs in complex sentences bound by the correlative conjunction 虽然 . . . 但是 . . . . This may explain the different patterns between reaction times and accuracy rates for random combinations with different combination frequencies. In other words, combination frequency effects for accuracy rates of random combinations did not indicate holistic representations for high-frequency random combinations.

These findings have important implications for the understanding of language representation and processing. Emergentist models have proposed that in language acquisition, cognitive process and experiences play an important role. Experiences with a certain linguistic unit would leave traces in the long-term memory. The more experiences there are, the higher degree of entrenchments of a unit there are in memory. This means that frequent units would be processed faster than less frequent ones. This should be the case for discontinuous multi-word units. With more exposures, discontinuous multi-word units would be more likely represented as a unit in mental lexicon. The present study provides empirical support for the claims of emergentist models.

On the other hand, there are no holistic representations for corpus-extracted random conjunction combinations, as indicated by the results that even high-frequency random combinations did not show any combination frequency effect. An interesting question was why high-frequency random conjunction combinations are not psychologically real. We suggest that frequency may not be the only factor in the formation of representations of discontinuous multi-word units. Besides frequency, grammatical function seems to be another critical variable for discontinuous correlative conjunctions. That is, if two conjunctions could not carry a certain grammatical function, they would not be represented as a unit, even though they have high frequency of co-occurrence in the Chinese corpus. For example, the correlative conjunction "*因为...所以.../ because...therefore...*" expresses *cause and effect*; but the reversible form or the random combination "*所以...因为.../ therefore...because...*" is not logical nor grammatical. This point can also be confirmed by the observation that a large number of conjunction combinations were observed in different sentences in the corpus used in the present study, but they cannot be treated as correlative conjunctions, even though they co-occurred frequently.

In summary, the present study showed that Chinese correlative conjunctions, but not random conjunction combinations, can be represented holistically in mental lexicon, in addition to their individual word representations. This may be attributed to the frequency and grammatical function of the correlative conjunctions. The findings enrich the understanding of the mental representation of multi-word units.

Appendix I

The full set of stimuli used in the present study.

CCHF	者...或者...	either...or...	RCHF	虽然...还是...	although...still...
	越...越...	the more...the more...		然而...与...	however...and...
	不仅...而且...	not only...but also...		由此...而...	from this...and...
	既...又...	both...and...		固然...但...	no doubt...but...
	如果...那么...	if...so...		虽然...仍然...	although...still...
	要是...就...	if only...then...		一方面...也...	on the other hand...also...
	只要...就...	as long as...then...		于是...使...	so...then...
	不仅...还...	not only...also...		即使...不...	even if...no...
	无论...还是...	no matter...or...		先...然后...	first...then...
	首先...然后...	first of all...then...		虽然...已...	although...already...
	不但...而且...	not only...but also...		比如...就...	such as...then...
	即使...也...	even if...still...		不论...都...	no matter what...all...
	一旦...就...	once...then...		然后...再...	and then...again...
	不管...都...	whatever...then...		若...则...	if...so...
	尽管...但是...	no matter if...but...		因此...就是...	so...even...
	虽然...但是...	although...yet...		当然...更...	of course...even more...
	假如...就...	if...then...		虽然...并不...	although...not at all...
	不但...还...	not only...but also...		虽然...由于...	although...due to...
	不但...反而...	not only...instead of...		虽...很...	though...very...
	无论...都...	no matter...still...		此外...还...	in addition...also...
	只有...才...	only...then...		特别...或...	specially...or...
	然而...却...	however...but...		只有...或...	only...or...
	可是...却...	but...but...		虽然...没有...	although...no...

(continued)

Continued.

	一边…一边…	while…as in…		一方面…又…	on the other hand…again…
CCLF	即使…总是…	even…always…	RCLF	尽管…就…	no matter if…then…
	固然…更…	no doubt…even more…		才能…都…	in order to…then…
	由于…是以…	due to…therefore…		于是…所以…	then…so…
	不是…就…	not…then…		不仅…则…	not only…so…
	即使…但…	even if…but…		还是…就…	still…then…
	既然…于是…	inasmuch as…so…		可是…但…	however…but…
	虽然…倒是…	Although…actually…		要是…或…	if…or…
	由于…因之…	due tovfor this reason…		越…才…	the more…then…
	不是…就是…	either…or…		不论…还…	no matter what…still…
	然后…终于…	and then…finally…		无论…却…	no matter…but…
	尽管…可是…	no matter if…yet…		立即…都…	immediatly…then…
	不但…并且…	not only…but also…		不但…又…	not only…again…
	或者…或者说…	or…or…		从而…所以…	so as to…therefore…
	要不是…也…	if not…still…		于是…但…	so…but…
	不仅…就是…	not only…then…		不仅…再…	not only…again…
	即使…还是…	even if…still…		既然…因此…	inasmuchas…for this reason…
	就算…也…	even if…still…		然后…因为…	and then…because…
	即使…但是…	even if…ut……		既然…还…	inasmuchas…still…
	不但…甚至…	not only…even…		不仅…所以…	not only…therefore…
	不但…还是…	not only…still…		不仅…却…	not only…but…
	然后…只好…	and then…onlyv…		如果…可是…	if…but…
	之所以…就是…	the reason…then…		只有…因为…	only…because…
	既…也…	both…also…		可见…还…	thus…still…
	要不是…就…	if not…then…		要是…与…	if…and…

CCHF: correlative conjunctions with high frequency; CCLF: correlative conjunctions with low frequency; RCHF: random combination with high frequency; RCLF: random combination with low frequency.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

Note

1. The high-level missing data indicated the difficulty in judgment (e.g., more errors on low-frequency combinations and random combinations). Therefore, we also conducted the analyses with and without removing any outliers. The results showed an identical pattern with and without outliers. If outliers were included, the missing data accounted for 23.2% in grammatical condition and for 42.9% in random condition. The mean reaction times for four groups were 1448, 2215, 2750, 2681 ms (*F1*) and 1531, 2381, 2733, 2781 ms (*F2*), respectively.

References

Arcara, G., Lacaita, G., Mattaloni, E., Passarini, L., Mondini, S., Beninca, P., & Semenza, C. (2012). Is “hit and run” a single word? The processing of irreversible binomials in neglect dyslexia. *Frontiers in Psychology*, *3*, 1–11.

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

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

Balota, D. A., & Chumbley, J. I. (1990). Where are the effects of word-frequency effects in visual word-recognition tasks? Right where we said they were? Comment on Monsell, Doyle, and Haggard (1989). *Journal of Experimental Psychology: General*, *119*, 231–237.

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*, 241–248.

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*, 1001–1024.

Bod, R. (1998). *Beyond grammar: An experience-based theory of language*. Stanford, CA: Center for the Study of Language and Information.

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

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

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one?. *The Mental Lexicon*, *9*, 267–293.

Chao, Y. (1970). *A Grammar of Spoken Chinese*. University of California Press.

Columbus, G. (2010). Processing MWUs: Are MWU subtypes psycholinguistically real? In D. Wood (Ed.), *Perspectives on formulaic language: Acquisition and communication* (pp. 194–210). New York, NY: Continuum.

Conklin, K., & Schmitt, K. (2012). The processing of formulaic language. *Annual Review of Applied Linguistics*, *32*, 45–61.

De Cat, C., Klepousniotou, E., & Baayen, R. H. (2015). Representational deficit or processing effect? An electrophysiological study of noun-noun compound processing by very advanced L2 speakers of English. *Frontiers in Psychology*, *6*, 1–17.

DeFrancis, J. (1989). *Visual speech: The diverse oneness of writing systems*. Honolulu, Hawaii: University of Hawaii.

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

Goldberg, A. E. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago, IL: Chicago University Press.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language*. Oxford, England: Oxford University Press.

Holsinger, E. (2013). Representing idioms: Syntactic and contextual effects on idiom processing. *Language and Speech*, *56*, 373–394.

Institute of Applied Linguistics of Ministry of Education. (2015). *Corpus online*. Retrieved from http://www.cncorpus.org/

Juhasz, B. (2007). The influence of semantic transparency on eye movements during English compound word recognition. In R. van Gompel, M. Fischer, W. Murray & R. Hill (Eds.), *Eye movements: A window on mind and brain* (pp. 374–389). Amsterdam, the Netherlands: Elsevier Science.

Kapatsinski, V., & Radicke, J. (2008). Frequency and the emergence of prefabs: Evidence from monitoring. In R. Corrigan, E. A. Moravcsik, H. Ouali & K. M. Wheatley (Eds.), *Formulaic language: Acquisition, loss, psychological reality, and functional explanations* (pp. 499–520). Amsterdam, the Netherlands: John Benjamins.

Kuperman, V., Schreuder, R., Bertram, R., & Baayen, R. H. (2009). Reading polymorphemic Dutch compounds: Toward a multiple route model of lexical processing. *Journal of Experimental Psychology*, *35*, 876–895.

Langacker, R. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in cognitive linguistics*. Amsterdam, the Netherlands: John Benjamins.

Leong, C. K. (1973). Reading in Chinese with reference to reading practices in Hong Kong. In J. Downing (Ed.), *Comparative reading: Cross-national studies of behavior and processes in reading and writing* (pp. 383–402). New York, NY: Macmillan.

Mattingly, I. G. (1992). Linguistic awareness and orthographic form. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning* (pp. 11–26). Amsterdam, the Netherlands: Elsevier.

Mondini, S., Jarema, G., Luzzatti, C., Burani, C., & Semenza, C. (2002). Why is "red cross" different from "yellow cross"? A neurophysiological study of non-adjective agreement within Italian compounds. *Brain and Language*, *81*, 621–634.

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.

Pollatsek, A., Hyona, J., & Bertram, R. (2000). The role of morphological constituents in reading Finnish compound words. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 820–833.

Reali, F., & Christiansen, M. (2007). Processing of relative clauses is made easier by frequency of occurrence. *Journal of Memory and Language*, *57*, 1–23.

Schcaneveldt, R. W., & Meyer, D. E. (1973). Retrieval and comparison processes in semantic memory. In S. Kornblum (Ed.), *Attention and performance IV*. New York, NY: Academic Press.

Schmitt, N., Grandage, S., & Adolphs, S. (2004). Are corpus-derived recurrent clusters psycholinguistically valid? In N. Schmitt (Ed.), *Formulaic sequences* (pp. 127–151). Amsterdam, the Netherlands: John Benjamins.

Seidenberg, M. S. (1985). The time course of phonological code activation in 2 writing systems. *Cognition*, *19*(1), 1–30.

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

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

Taft, M., Liu, Y., & Zhu, X. (1999). Morpheme processing in reading Chinese. In J. Wang, A. W. Inhoff & H. -C. Chen (Eds.), *Reading Chinese script: A cognitive analysis* (pp. 91–114). Mahwah, NJ: Lawrence Erlbaum.

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

Tremblay, A., & Baayen, R. H. (2010). Holistic processing of regular four-word sequences: A behavioral 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, England: Continuum International.

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

Xue, X., & Shi, C. (2013). The nature of formulaic language and the systematical hierarchy of Chinese formulaic language [In Chinese]. *Contemporary Rhetoric*, *3*, 32–46.

Zhang, H., Yang, Y., Gu, J., & Ji, F. (2013). ERP correlates of compositionality in Chinese idiom comprehension. *Journal of Neurolinguistics*, *26*, 89–112.

Author Biographies

Lingyue Kong, an associate professor in the School of Chinese as a Second language at Peking University. His research focuses on language and cognition, lexical processing, Chinese word recognition, language acquisition and teaching.

John X. Zhang, Professor of Psychology in Fudan University, BA in physics in 1989 from University of Science and Technology of China, PhD in 1999 from Princeton University, teaching from 2006 in the Psychology Department of the Chinese University of Hong Kong as an assistant professor. Working from 2013 as full professor in Fudan University. Authoring more than 70 journal papers,

main research interest is the study of Chinese visual word recognition using brain imaging techniques.

Yongwei Zhang, PhD, is currently an assistant researcher in the Institute of Linguistics, Chinese Academy of Social Sciences (CASS), China. His research interests include corpus and computational linguistics. He is the principal programmer of the Corpus-Based Computer-Assisted Lexicography (CBCAL). His publications can be found in Journal of Contemporary Linguistics (CASS).