Journal of Neurolinguistics 66 (2023) 101127

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

Frontotemporal effective connectivity revealed a language-general syntactic network for Mandarin Chinese

Luyao Chen[a,d,1], Chenyang Gao[b,1], Zhongshan Li[c], Emiliano Zaccarella[d], Angela D. Friederici[d,**], Liping Feng[a,*]

[a] *Max Planck Partner Group, School of International Chinese Language Education, Beijing Normal University, Beijing, China*
[b] *School of Global Education and Development, University of Chinese Academy of Social Sciences, Beijing, China*
[c] *School of Foreign Languages and Literature, Beijing Normal University, Beijing, China*
[d] *Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany*

ARTICLE INFO

Keywords:
Chinese syntax
Function words
Hierarchical syntactic processing
fMRI
Effective connectivity

ABSTRACT

Human language is proposed to be hierarchically constructed according to syntactic information. Studies on languages with overt morphosyntactic markers (e.g., German) have found a key frontotemporal syntactic network that includes Broca's area (Brodmann Area, BA 44/45) and the posterior temporal cortex (pTC). Whether this syntactic network is language-general is still unspecified. Mandarin Chinese is a suggestive empirical test case, lacking morphosyntax and relying heavily on function words to guide syntactic hierarchy construction. By developing the jabberwocky sentence paradigm, we created sets of visually-presented Chinese structures formed by function words and pseudo-words (the structure condition), and contrasted the structures with comparable word lists (the word-list condition) in healthy Chinese-speaking adults in a functional magnetic resonance imaging (fMRI) experiment. Participants were required to identify the syntactic category of each structure by merging its constituents into syntactic hierarchies, guided by function words. Compared with the word-list condition, the structure condition (a) elicited higher involvement of left BA 44, and (b) recruited a language-general syntactic network as revealed by the effective connectivity between BA 44, precentral gyrus, and pTC. These findings specified the neural basis for Chinese syntax and further corroborated the unique human language faculty across languages in a neurobiologically ubiquitous fashion.

1. Introduction

The human language is proposed to be hierarchical in nature, distinctive in the human species (Berwick & Chomsky, 2016; Friederici, 2017; Hauser et al., 2002). Hierarchical syntactic processing refers to using the syntactic categories (content words like nouns and verbs, and function words such as auxiliaries and prepositions) to build up syntactic hierarchies on the basis of a fundamental mechanism—linguistically known as *Merge* (e.g., Chomsky, 1995, 2013, 2015; Fujita, 2014; Goucha et al., 2017; Hoshi, 2018, 2019; Miyagawa et al., 2013; Zaccarella, Schell, & Friederici, 2017). Merge denotes a binary syntactic operation which takes two

---

[*] Corresponding author. Prof. Dr. Liping Feng. ; No.19, Xinjiekouwai Str., Haidian Dist., 100875 Beijing, China.
[**] Corresponding author. Prof. Dr. Angela D. Friederici. ; Stephanstraße 1a, 04103 Leipzig, Germany.
E-mail addresses: friederici@cbs.mpg.de (A.D. Friederici), fengliping@bnu.edu.cn (L. Feng).
[1] These authors contributed equally to the current work.

https://doi.org/10.1016/j.jneuroling.2023.101127
Received 3 October 2022; Received in revised form 22 December 2022; Accepted 4 February 2023

elements and combines them together as a new constituent, and entails the steps: (a) identifying the syntactic categories of the elements to be merged (e.g., “the”, a determiner, D; “apple”, a noun, N); (b) searching the Head whose syntactic category reflects the syntactic identity of the whole phrase [e.g., in “the” and “apple”, “the” (D) is the Head] (c) labeling the newly-merged phrase (e.g., labeling “the apple” as a determiner phrase, DP) (see also Goucha et al., 2017).; The phrase (“the apple”) itself can be merged with a new element (e.g., eat, a verb, V), and such recursive merge processes will build up hierarchical syntactic structures (e.g., “{VP {V eat} {DP the apple}}“).

Recent data converge on the fact that Broca’s area, especially the left BA (Brodmann Area) 44 of Broca’s area in the posterior inferior frontal gyrus (pIFG), is consistently involved during hierarchical syntactic processing in language (e.g., Goucha & Friederici, 2015; Makuuchi et al., 2009; Ohta et al., 2013; Wang et al., 2021; Zaccarella & Friederici, 2015). A number of studies support the hypothesis that the connection between the posterior inferior frontal gyrus and the posterior superior and middle temporal cortex (pTC) appears to function as a selective left-hemispheric frontotemporal network engaged during complex syntactic processes (e.g., den Ouden et al., 2012; Matchin et al., 2017; Pallier et al., 2011; Santi & Grodzinsky, 2010; Schell et al., 2017; Zaccarella, Meyer, et al., 2017; see also Hagoort & Indefrey, 2014 and Zaccarella, Schell, & Friederici, 2017 for meta-analyses). These studies generally adopted the following comparisons.

(a) Sentences/phrases (i.e., structured language sequences) vs. word lists (Goucha & Friederici, 2015; Matchin et al., 2017; Ohta et al., 2013; Schell et al., 2017; Tyler et al., 2010; Zaccarella, Meyer, et al., 2017; Zaccarella & Friederici, 2015). When compared with word lists, natural sentences/phrases elicit higher activation in the posterior inferior frontal gyrus (esp., BA 44) and the posterior temporal lobe (Matchin et al., 2017; Schell et al., 2017; Zaccarella, Meyer, et al., 2017). Activation in BA 44 is also found when only the function words, as well as the inflectional affixes, are retained in grammatical *jabberwocky* sentences (Goucha & Friederici, 2015; Matchin et al., 2017; Ohta et al., 2013; Pallier et al., 2011; Tyler et al., 2010). For example, Goucha and Friederici (2015) found that when the German *jabberwocky* sentences only had function words and inflectional affixes, BA 44 was still consistently involved. The result was also evidenced by using Japanese jabberwocky sentences in Ohta et al. (2013). When studies narrowed the sentences down to simpler structures (such as two-word phrases) at the basic level of Merge, BA 44 as well as the pTC were also activated (Schell et al., 2017; Zaccarella, Meyer, et al., 2017).

(b) Complex structures (with deeper hierarchical embedding or more mergeable elements) vs. simpler structures (e.g., den Ouden et al., 2012; Makuuchi et al., 2009; Pallier et al., 2011; Santi & Grodzinsky, 2010; Wang et al., 2021). For instance, Makuuchi et al. (2009) demonstrated that complex German sentences with more center-embedded structures (i.e., deeper hierarchical embedding) could highly activate BA 44 when compared with simpler sentences. Santi and Grodzinsky (2010) further specified that BA 44 was sensitive to the hierarchical syntactic structure types during English complex structure processing. A recent language training study on complex German structures (Wang et al., 2021) identified that BA 44 and its functional connection to pTC could predict training performance. Besides, complex sentences with more mergeable elements also provided evidence for the neural substrates for hierarchical syntactic processing, as evidenced by Pallier et al. (2011) that the activation of both IFG and pTC would increase with the size of the French constituent. This was replicated by using English structures in Matchin et al. (2017).

(c) Hierarchy-based artificial grammars vs. adjacency-based artificial grammars (e.g., Bahlmann et al., 2008, 2009; Chen, Goucha, et al., 2021; Friederici et al., 2006; Opitz & Friederici, 2003, 2004, 2007; Uddén et al., 2017). Previous studies showed that the processing of sequences following hierarchical rules (see also Fitch & Hauser, 2004) lead to a significant activation of BA 44 (e.g., Bahlmann et al., 2008, 2009; Friederici et al., 2006), compared to sequences following simpler linear rules. A recent artificial grammar study (Chen, Goucha, et al., 2021) using improved hierarchical rules in which building up syntactic hierarchies was mandatory, demonstrated that BA 44 as well as its effective connectivity to the pTC were critically engaged in hierarchical syntactic processing regardless of meaning.

Overall, studies using natural languages mainly focused on languages with overt morphosyntactic cues, which are necessary to the construction of syntactic hierarchies. However, Mandarin Chinese (hereafter, Chinese for short) is “impoverished” of morphosyntactic information, and its syntax heavily relies on function words and word order information (Chao, 1968; Zhu, 1985). While word order is seen as an outcome of the externalization system which translates hierarchical structures into linear sequences, the syntactic rules applying over lexical items constitute the core component of the human language faculty, according to a prominent linguistic model (Berwick et al., 2013). In syntactic processing, we hereby specifically focused on the role of Chinese function words, a critical syntactic device for Chinese syntax (Chao, 1968; Guo, 2018; Huang & Liao, 2011; Tang, 2019; Zhu, 1985), which can help to better understand how the human brain can efficiently use the available syntactic information in a given language to form hierarchical syntactic structures. For instance, if the Chinese word “高兴/kɑʊ1 ɕiŋ4/” is combined with “的/tə/” (a structural auxiliary of noun phrases), then it forms the noun modifier “happy”, “高兴的”; conversely if it is combined with “地/tə/” (a structural auxiliary of verb phrases), then it forms the verb modifier “happily”, “高兴地”. Nevertheless, this by no means implies that the function words are the only syntactic device for Chinese syntax, or the only root for syntactic hierarchy in Chinese. Rather, the function words are able to provide us with a proper window to investigate the neurobiology of Chinese syntax, and to make it comparable with the findings of previous studies on other languages.

Emerging evidence from natural Chinese sentence processing revealed a dominant wide-range activation pattern (including both syntactic and semantic regions) in the left frontotemporal cortex, including but not limited to, the inferior frontal gyrus (IFG), the middle frontal gyrus (MFG), the anterior temporal lobe (aTL), the posterior temporal cortex (pTC), and part of the parietal lobe (PL) (Chang et al., 2020; Chee et al., 1999; Chou et al., 2012; Luke et al., 2002; Sun et al., 2021; Wang et al., 2008; Wu et al., 2018). Such a wide range of activation might be attributed to the fact that Chinese syntactic processing highly interacts with semantic processing, as revealed by a number of event-related potentials (ERP) studies previously. ERP studies on Indo-European languages (esp., German) identified an initial syntactic-structure building effect as indexed by the early left anterior negativity (ELAN) within 120–200 ms elicited by the syntactic violation condition (Friederici, 2017; Friederici et al., 1993; Friederici & Weissenborn, 2007). Moreover, when a sentence contains both syntactic and semantic violations—"double violations", only ELAN appeared (without the semantic-related N400), indicating that the early syntactic violation might *block* the subsequent semantic processing, resulting in a "semantic block effect", that is, when participants realized the syntactic violation, they would give up processing the meaning of the target sentence (e.g., Friederici et al., 2004; Friederici & Weissenborn, 2007; Gunter et al., 1999; Hahne & Friederici, 2002; Hahne & Jescheniak, 2001; Isel et al., 2007). Both the existence of ELAN and the block effect for semantic processing support the assumption that in Indo-European languages (esp., German), syntactic processing might be in advance of semantic processing, and function in an independent manner, thus in support of a syntax-first model. On the contrary, a number of Chinese studies failed to identify such effects (e.g., Wang et al., 2013; Wang et al., 2015; Yu & Zhang, 2008; Zhang et al., 2010). For instance, Yu and Zhang (2008) found that semantic integration proceeded (as a semantic-related N400 appeared in the double violation condition) even when the syntactic information was violated, and no ELAN could be identified. Therefore, these researchers demonstrated that unlike in Indo-European languages, syntactic and semantic processing in Mandarin Chinese might be parallel and highly interactive (see also Wu et al., 2018 for a similar comment).

The related fMRI studies on Chinese syntactic processing in NATURAL contexts (containing both syntactic and semantic information), comparing (a) Chinese syntactic structure processing with other languages (Chee et al., 1999; Luke et al., 2002), (b) syntactic structures of different complexities (Feng et al., 2015; Sun et al., 2021), (c) structures that contained syntactic violations with normal structures (Chou et al., 2012; Wang et al., 2008), and (d) phrases/sentences with word-list conditions (Bulut et al., 2017; Chang et al., 2020; Malik-Moraleda et al., 2022; Wu et al., 2018), are unlikely to systematically disentangle syntactic processing from semantic processing in Chinese. For example, by initially introducing the basic Merge paradigm into the study of Chinese syntactic processing, Wu et al. (2018) found that Chinese basic structure building could activate the posterior inferior frontal gyrus (BA 45) as well as the pTC (including both the posterior superior temporal and middle gyri, that is, pSTG and pMTG). Their analyses via the approach of dynamic causal modeling revealed that the effective connectivity between BA 44 and pMTG might play a critical role in building up the phrase structures. Since the semantic information was not excluded in that study, the syntactic network for Chinese was unspecified yet. Furthermore, to the best of our knowledge, a latest study using the natural Chinese language, demonstrated that syntactic processing is dissociable from semantic processing within the left IFG via the high-density electrocorticography (ECoG) technique (Zhu et al., 2022). However, the syntactic network still awaits to be specified, preferably in a noninvasive design.

In parallel, a recent artificial grammar learning/processing fMRI study (Chen et al., 2019) investigated the neural substrates for the long-distance dependency processing in *Chinese-like* structures, and found a significant involvement of Broca's area and the superior temporal gyrus (STG) in the left hemisphere. A follow-up effective connectivity analysis revealed that the left-hemispheric functional connection between Broca's area (BA 44) and STG might be critical to successfully processing the Chinese-like artificial grammar (Chen, Wu, et al., 2021). Nevertheless, a major concern on artificial grammar studies is the fact that they solely employ artificially generated words and structures, which only loosely reflect real words (esp., function words) and syntactic structures found in the Chinese language.

In light of these considerations, we followed the design of jabberwocky sentences as in Goucha and Friederici (2015) and Pallier et al. (2011), and developed a function-word-guided syntactic processing paradigm with meaning impoverished in the Chinese language—in which content words were replaced by pseudo-words, while highly abstract real Chinese function words with mere syntactic information (Guo, 2018; Huang & Liao, 2011; Ma, 2015; Tang, 2019; Zhu, 1982) were retained. By making use of real Chinese function words, syntactic rules, and structures, we highlighted the syntactic features of Chinese, while suppressing semantic processes, thus synthesized the advantages of both the natural language and the artificial grammar experiments. Moreover, the jabberwocky sentence design is beneficial for making our imaging results comparable with previous findings in the languages with overt morphosyntactic cues. Healthy adult Chinese native speakers underwent an fMRI (functional magnetic resonance imaging) experiment, during which they were required to identify the syntactic categories of the structures they read by merging the elements according to the syntactic information carried by the function words. Notably, the syntactic category identification task required participants to identify the specific types (NP or VP) of the target structures, thus ensuring that participants would build up higher syntactic nodes step-by-step following the Merge processes on the basis of function words. As for the imaging analyses, in addition to the standard whole-brain analysis, given the specific focus on brain activation patterns for syntactic processing within the language network, the present study employed the functional left hemispheric atlas derived from more than 200 participants processing natural language stimuli (extended from Fedorenko et al., 2010; http://web.mit.edu/evlab//funcloc/; see also Fedorenko & Blank, 2020).

In summary, the present study aims to.

(a) Specify the neural network for hierarchical syntactic processing in Chinese by developing a meaning-impoverished paradigm in which real function words, syntactic rules, and structures were retained.
(b) Evaluate the neurobiological comparability of syntactic processing with previous findings, by taking Chinese as an extreme case (relative to the languages with morphosyntactic information).

As such, we expected to verify that Chinese sentences with real function words would recruit a similar syntactic network, composed of pIFG (esp., BA 44) and pTC, identified by the previous studies on the languages with morphosyntactic information (e.g., den Ouden et al., 2012; Matchin et al., 2017; Pallier et al., 2011; Schell et al., 2017; Wang et al., 2021; Zaccarella, Meyer, et al., 2017), thus following some neurobiological universal principle of syntactic network for language (see also Wu et al., 2018).

2. Materials and methods

2.1. Participants

Twenty-two native Chinese-speaking adults were recruited for the study. All participants with normal or corrected-to-normal vision reported to be right-handed as confirmed by the Edinburgh handedness inventory (Oldfield, 1971). Participants also reported no history of psychiatric or neurological diseases. Besides, they were non-musicians, non-early-bilinguals, and did not major in linguistics, psychology, or neuroscience. Of note, the pseudo-words also contained Japanese Katakana, which are similar to Chinese characters, and therefore, Japanese learners were also excluded so as to avoid both the familiarity effect and the false processing of the Experimental materials (see *Materials*). The participants also had no experience in using the Chinese Phonetic alphabet (i.e., the Taiwanese Bopomofo, Zhuyin). All of them signed the informed consent before the experiment and received 120 Chinese Yuan (CNY) remuneration for participation. Twenty-one participants' data were analyzed (male:11, female:10, age: $M = 23.86$ years, $SD = 3.01$ years), while one participant's data were excluded because of excessive head motion artifacts (>2 mm in translation and >2° in rotation). Informed consent was obtained in a manner approved by the ethics committee of Beijing Normal University, Beijing, China.

2.2. Materials

In order to investigate the neural substrates for hierarchical syntactic processing on the basis of function words in Chinese, we selected typical Chinese function words (Tang, 2019), including dynamic and structural auxiliaries, prepositions, and quantifiers (see Table 1 for explanations). These typical function words are highly abstract without concrete meanings (Guo, 2018; Huang & Liao, 2011; Ma, 2015; Tang, 2019; Zhu, 1982), but they can combine with other content words to generate well-formed hierarchical syntactic structures, either verb phrases or noun phrases (see Table 1 for the example of basic structures). Pseudo-words were generated as content words by using unfamiliar symbols, including Japanese Katakana and Taiwanese Bopomofo (Zhuyin), which are similar to Chinese orthography, but unpronounceable and semantic-free for the participants (see also Supporting *Information Section 1*). To clarify, the Taiwanese Bopomofo (Zhuyin) was originally designed according to the Japanese Katakana, and both systems of symbols were close to real Chinese characters, and therefore, as also confirmed by the participants' oral reports, these symbols composed a unified set of "unfamiliar Chinese characters", instead of being treated as two language systems.

These pseudo-words were classified *a priori* into five syntactic categories (four-word tokens for each category), including verb, noun, adjective, numeral, and adverb (Table 2). Due to the fact that the content words contained both semantic and syntactic information (Goucha et al., 2017), we hoped to retain the syntactic information of the content words while removing the meanings. And this was helpful for the current experiment to set ungrammatical violations (see below). Moreover, for each word category, the pseudo-words were designed to share some visual similarities so as to facilitate the accessibility to their word category information, and it was further confirmed by all the participants that they memorized the word categories according to the visual similarities/cues. The rationale is that components of the Chinese characters are often able to reveal the syntactic/semantic categories of the corresponding characters (Chen & Yeh, 2015; Feldman & Siok, 1999; Wang & Zhang, 2016; Yeh et al., 2017; Zhang et al., 2020). For example, Chinese characters with a component "扌 " (hand) are always verbs, such as "打" (hit), "找" (look for), "摇" (shake), and so on. Therefore, we reasoned that this should be comparable to participants' natural language competence to relate the visual information to the word (syntactic) categories.

During the preparation of the word tokens and word categories, 24 Chinese linguists (also native speakers) were invited to evaluate the difficulty of relating the pseudo-content words *semantically* to the assigned syntactic categories by considering the closest real Chinese word/characters of these pseudo-words, with a 5-point Likert scale (1: extremely easy; 5: extremely difficult) for a set of 20 target pairs of the pseudo-content word and its syntactic category, randomly mixed up with 20 foils. This was to check whether, for the Chinese linguists, a semantic strategy to relate the pseudo-words to the real words (so as to identify their syntactic categories for ease of memorization) could be easily utilized via an explicit and active rating task. The descriptive statistics (mean scores with their

Table 1
Experimental materials: function-word categories, word tokens, and sample basic structures.

Function-Word Categories	Typical Word Tokens	Sample Basic Structures
dyn-Aux.	着 了 过 /tʂə//lə//kuo4/	$[_{\mathrm{VP}}$ V着/了/过N]
str-Aux.'	地 /tə/	$[_{\mathrm{VP}}$ Adv.地V]
Prep.	把 被 /pa3//peɪ4/	$[_{\mathrm{VP}}$ 把/被N V]
Q	只 件 个 /tʂʅ1//tɕiɛn4//kə4/	$[_{\mathrm{NP}}$ Num.只/件/个N]
str-Aux.	的 /tə/	$[_{\mathrm{NP}}$ Adj.的N]

Abbreviations: dyn-Aux.: dynamic auxiliary; str-Aux'.: structural auxiliary (for verb modifiers); Prep.: preposition; Q: quantifier; str-Aux.: structural auxiliary (for noun modifiers); V: verb; N: noun; Adj.: adjective; Num.: number; Adv.: adverb.

Table 2
Experimental materials: content-word categories and pseudo-content-word tokens.

Content-Word categories	Pseudo-Content-Word Tokens	Semantic Relating Mean Scores [95% CI]
V	ㄉ ㄞ ㄙ ㄊ	3.90 [3.59, 4.20]
N	ラ ウ ㄜ ㄛ	3.80 [3.34, 4.26]
Adj.	ㄓ 屮 ヨ 王	3.89 [3.49, 4.28]
Num.	ㄩ ㄇ ユ ㄈ	3.66 [3.19, 4.12]
Adv.	ㄍ リ 彳 ㄐ	4.04 [3.59, 4.49]

Abbreviations: V: verb; N: noun; Adj.: adjective; Num.: numeral; Adv.: adverb; CI: confidence intervals.

confidence intervals) were shown in Table 2. A one-way ANOVA was performed for the 5 pseudo-word syntactic categories, and the result showed no significant difficulty differences among the categories ($F(4,92) = 0.25, p = .62, \eta_p^2 = 0.01$). Moreover, the difficulty of each category was significantly ($ts(23) \geq 2.94, ps < .01, ds \geq 0.60$) higher than the intermediate level (= 3), and presented no statistical differences with the level of "very difficult" (= 4) ($-1.54 \leq ts(23) \leq 1.92, ps \geq .14$), as revealed by two separate one-sample t-tests performed for each category. These results indicated that even for the Chinese linguists under an explicit and active rating condition, the semantic relations between the pseudo-words and their corresponding syntactic categories were very difficult to detect. Thus, we assumed that for the non-linguistic background participants it should be quite difficult to memorize the word tokens and their corresponding syntactic categories via semantic strategies. Indeed, in the post-test interviews, all participants reported that they categorized the word tokens according to the visual similarities rather than semantics, which was deliberately designed to reduce the memory load of relating the word tokens to their categories.

The "complex structures" (see Table 3 and Fig. 1 for illustration), including grammatical verb phrases and noun phrases, were generated for the task during scanning, in which each structure was composed of three function words and four pseudo-content words (see Table S2 for details in Supporting *Information*). For instance, in the sequence of "ㄉ了ㄇ只ㄓ的ㄜ", "ㄉ ㄇ ㄓ ㄜ" were pseudo-content words, and "了 只 的" were real Chinese function words. This sequence corresponded to the syntactic structure "V dyn-Aux. Num. Q Adj. str-Aux.' N", and a natural language example could be "点了三只贵的蟹" [/tiɛn3 lə san1 tʂʅ1 kuei4 tə ɕiɛ4/; literal translation in English: order-(ed) three (Q) expensive (str-Aux.') crab; English translation: ordered three expensive crabs].

As for the reason to adopt complex structures in this study, the basic structures containing only one function word might not be optimal for the current experiment. Besides, the types of the basic structures were limited and relatively simple (see Table 1), and this might trigger the risk of using a mechanical memorization strategy from the participants. Moreover, given that each basic structure contained a single function word, which might also induce the keyword strategy by which participants could just focus on the function word of each basic structure to complete the task. Therefore, the complex structures are *complex* enough to maximize the hierarchical syntactic processing effect within a reasonable time-window, and the number of the function words and pseudo-words could ensure the structure variability so as to prevent the non-syntactic strategies.

Furthermore, in case participants could identify the phrase category by merely focusing on certain function words or their combinations without processing the whole structure, the function word categories varied in positions, and the occurrence frequency of the function word tokens was controlled. "Ungrammatical structures" were also designed (Table 3 and Fig. 1, see also Table S3 for illustration in Supporting *Information*). For example, in "ㄉ了ㄇ只ㄍ的ㄜ", "的/tə/" an adjective was required to generate a noun modifier like "[$_{\text{Adj.}}$ ㄓ的]" as aforementioned. It should be stressed that, in the complex structures, the violation was designed to occur at either the second or the third pseudo-word position to avoid the "edge-effect" (e.g., Endress et al., 2010; Sonnweber et al., 2015). Nevertheless, the *local* syntactic hierarchy build-up was still possible in these structures. For instance, in the abovementioned example, "ㄉ了ㄇ只 … ㄜ" could still be hierarchically processed by the participants (see also Fig. 1). Thus, even in the ungrammatical structures, the hierarchical syntactic processing mechanism should still be in operation. The word category violations were set in accordance with their occurrence frequency (<1%) in the BCC Corpus (Xun et al., 2016), thus ensuring these violations were grammatically unacceptable. Besides, the frequency of occurrence of each pseudo-word token from each category was also balanced across grammatical and ungrammatical structures, guaranteeing that participants had to process the whole structure carefully before giving a response. As subsidiary evidence, non-task strategies were not reported during the post-hoc interview after both the pilot and actual experiments. Therefore, due to the present manipulation of the materials, taking non-task strategies would merely result in chance-level performances (See also Supporting *Information Section 1* for more about the materials of the structure condition).

Given that, according to Table 3, in all the three sub-structure-types: grammatical verb phrase structure, grammatical noun phrase

Table 3
Experimental materials: complex structures and word lists.

Conditions	Sample Structures/Word lists	Identities/Categories
Structure	ㄉ了ㄇ只ㄓ的ㄜ	**V** (simplified for verb phrase)
	ㄩ只被ㄜㄉ的ㄛ	**N** (simplified for noun phrase)
	ㄉ了ㄇ只ㄍ的ㄜ	※ (word category violation)
Word list	ㄉㄞㄉㄞㄞㄊㄊ	**A** (simplified for artificial word list)
	件过的过件的的	**C** (simplified for Chinese word list)
	ㄉㄞ#ㄉㄞㄊㄊ or 只只着过#过着	※ (outlier member violation)

Note: An underline was shown in this table for illustrating the violation.

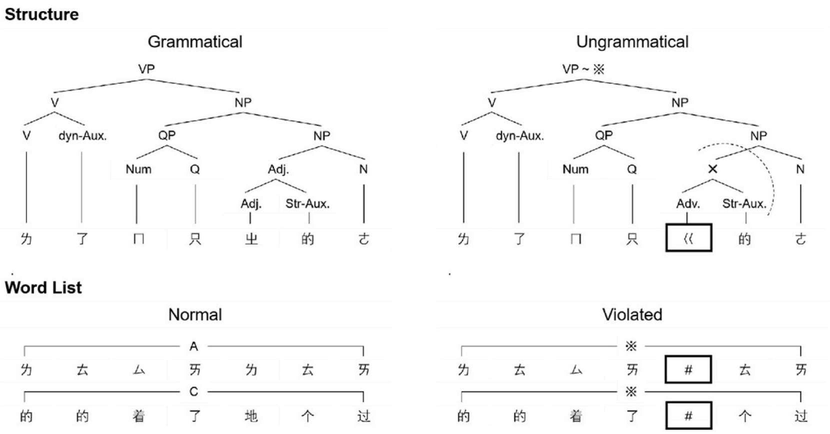


Fig. 1. Experimental material examples for each condition. In the structure condition, both grammatical and ungrammatical verb phrase (VP) trials were shown with tree diagrams. As for the ungrammatical VP trial, the black box marked the word category violation in this example (here, the violated word category is the Adv.), and the dashed curve denoted to truncate the violated part while keeping the rest of elements still combinable (here, resulting in a *problematic* VP; participants should choose ※ for response). Normal and violated word-list trials were also presented here for comparison, and the black box marked the problematic element in the violated sequence. Abbreviations: str-Aux.: structural auxiliary (for noun modifiers); dyn-Aux.: dynamic auxiliary; Q: quantifier; V: verb; N: noun; Adj.: adjective; Num.: number; Adv.: adverb; P: phrase (e.g., VP: verb phrase); A: artificial word list; C: Chinese word list; ※: structure/word-list violation.

structure, and ungrammatical structure in which local merge is still permitted, the hierarchical syntactic processing on the basis of the function words in Mandarin is at work. We synthesized these three kinds of materials together to form a single condition — “structure”.

In order to isolate the hierarchical syntactic processing effect, word lists were generated as a control condition, in which the words could not be hierarchically combined (see Table 3 and Fig. 1). According to a meta-analysis (Zaccarella, Schell, & Friederici, 2017), a word list containing both content and function words might inevitably trigger the syntactic processes, and resulted in the reduction of the syntactic effects under the contrast “structure > word list”. Therefore, two types of “*mono*-word-category word lists” were developed: one was composed of the pseudo-content-word tokens from a given category (such as a list of pseudo-nouns), and the other was composed of real function words. The same words and pseudo-words were used to create both the structures and the word lists. The materials within each block were randomized for each condition. The block types were also pseudo-randomized, which means that a structure block may occur before or after a word-list block. Therefore, the current design could avoid the potential repetition inhibition effect induced by the same words from different conditions. Rather, using the same words to create structures along with word lists might be more beneficial for subtracting the lexical effects from the structure condition.

In each word list, an outlier symbol (not belonging to any word categories) would appear in the same position as that in the ungrammatical structure. To note, all the structures or word lists were of the same length (7 words, with each character or symbol as one word). Therefore, by carefully developing the word-list condition, we were able to have the low-level processing effects removed under the “structure > word list” contrast. It is also noteworthy that, the unfamiliar pseudo-words were unpronounceable for the participants. This was set to mimic a common situation in Chinese language reading: while the pronunciation of unfamiliar words (or characters) in Chinese may not be accessible due to the presence of pictophonetic characters or ideographic characters with no direct phonetic mapping (esp., single characters, as used in this study: Peng, 1997; Wang, 2000; Zhou et al., 2001), the syntactic sense can still be perceived (i.e., realizing the structures being processed are NPs or VPs). This is because the syntactic categories of these pseudo-content words as well as the critical function words are all accessible (see also *2.3 Procedures*), and Merge, by definition (see also *Introduction*), operates on the syntactic category information, instead of the sounds (acoustic features) of the words. As actual Chinese grammatical structures were retained, the sequences would be parsed as well-formed structures just with some *unfamiliar* content words. Participants could only accurately and systematically pronounce the real function words in the structure condition as well as the word-list sub-types (the Chinese word list and the abnormal Chinese word list such as “只只着过#过着”). The same words (either pseudo-words or real function words) were used for both the structure and the word-list conditions. Thus, the contrast “structure > word list” was able to reduce the articulatory effects, especially of the function words. In detail, there were 72 structure trials with $72 \times 3 = 216$ pronounceable function words in total (to note, each structure contained 3 function words), and 36 Chinese pronounceable function word lists, including 18 normal trials with $18 \times 7 = 126$ function words and 18 abnormal trials with $18 \times 6 = 108$ function words (to note, in a 7-word abnormal function word list, there was an outlier and 6 function words). Therefore, structure condition had 216 pronounceable function words, and word-list condition had 234 ones. The contrast of “structure > word list” should thus cancel out the effects of pronunciation at least at the lexical level.

Materials were visually presented by E-prime 2.0 (Psychology Software Tools, Inc., Pittsburgh, PA, USA).

2.3. Procedures

The experimental procedures were illustrated in Fig. 2. Before the fMRI scanning session, a behavioral adaptation session was conducted to help the participants familiarize with the paradigm. Three days before the behavioral adaptation session, a pool of participants received the pseudo-content word vocabulary list, in which the pseudo-content words as well as their corresponding word categories were explicitly presented. The participants were required to memorize the word categories of these pseudo-content words, and to participate in an online word category test via Sojump (https://www.wjx.cn) to judge whether the presented words belong to the target word categories. Participants who reached the accuracy of 90% for the first run were recruited for the behavioral adaptation session, which guaranteed that the participant well acquired the word category knowledge of the pseudo-words, and thus improved the pseudo-word familiarity.

The behavioral adaptation session was composed of three phases, as illustrated in Fig. 2A. In the first phase, participants underwent vocabulary testing, in which they were required to pick up the target word category (expressed in Chinese: "动" for verb, "名" for noun, "形" for adjective, "副" for adverb, and "数" for numeral; the order of the choices were randomized for each trial) listed below the presented pseudo-content word through a multiple-choice task, with feedback presented after the response. This phase contained 20 trials per block (8 blocks in total), and if a participant reached the criterion — two-successive-block-accuracy above 90%, he/she would enter into the next phase. This vocabulary testing phase was set up to help participants re-familiarize with the pseudo-content words to their corresponding syntactic categories, thus ensuring that the participants could access the categories of the pseudo-content words successfully. All participants reported to utilize the visual similarities of the words to easily access their corresponding syntactic categories in a relatively efficient manner.

The subsequent basic-structure phase included two sub-phases in order: one was a grammaticality judgement sub-phase, and the other was a sub-phase of category identification. In accordance with the definition of *Merge*, hierarchical syntactic processing generally requires the categories of the elements to be identified first, and then licenses the newly-generated component with a new categorical label (Chomsky, 1995; Goucha et al., 2017; Zaccarella & Friederici, 2015). Therefore, the grammaticality judgement sub-phase guided

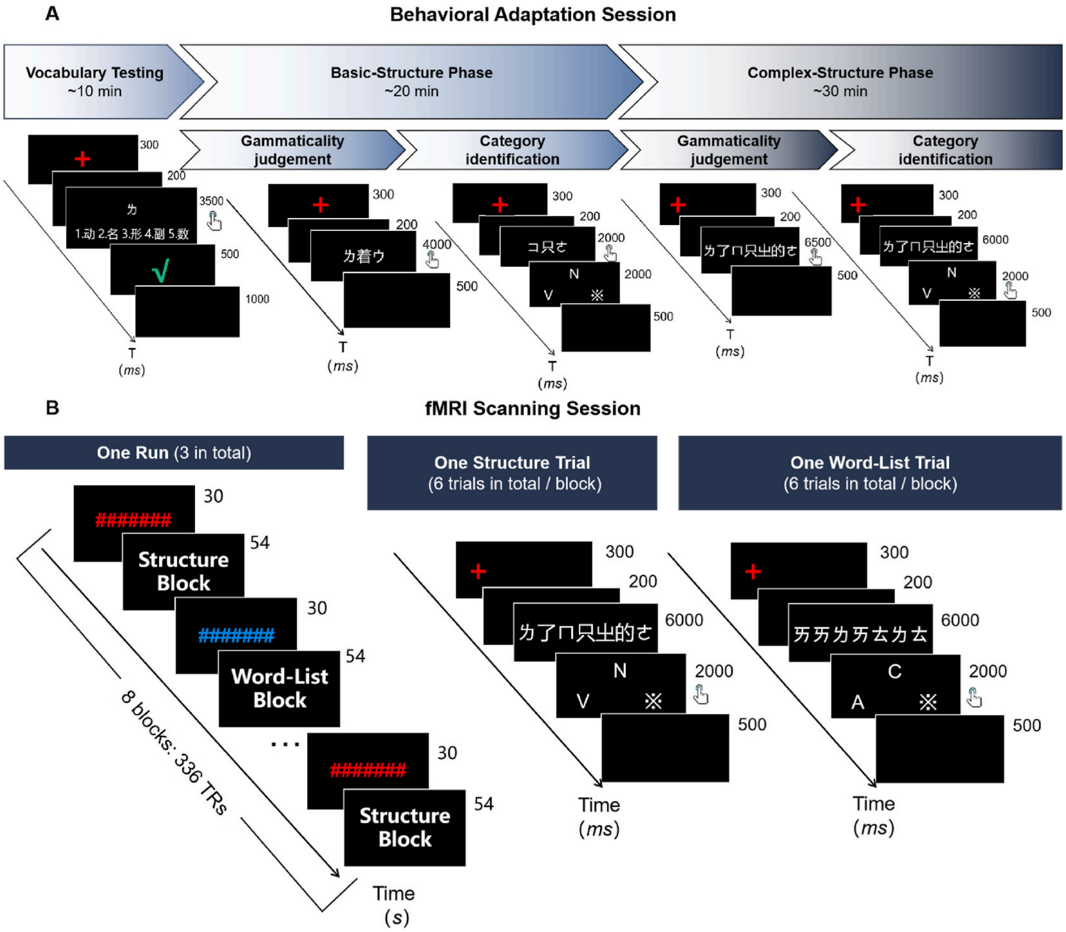


Fig. 2. The procedures of the experiment. (A) The behavioral adaptation session. Firstly, participants underwent vocabulary testing by identifying the syntactic categories of the target pseudo-words. When their accuracy of the 2-successive-block reached 90%, they could enter into the next phase. Secondly, they underwent a "basic-structure phase" to judge the grammaticality of the basic structures and to identify the syntactic categories of the target basic structures in separate blocks. Last, participants completed both the grammaticality judgement and the category identification tasks for the complex structure in the "complex-structure phase". (B) The fMRI scanning session. One run and the presentation of the trials with the timing parameters were shown. See the Main Text for details.

the participants to access the syntactic category information of each (pseudo-content) word, and the category identification sub-phase then required the participants to combine the words, including labeling the final outcome with its syntactic identity. Note that each sub-phase was composed of 2 practicing blocks and several testing blocks: (a) The grammaticality judgement sub-phase contained 8 blocks for testing with 8 trials per block, and each trial lasted for 5.5s; (b) The category identification sub-phase contained 6 testing blocks with 8 trials per block, and each trial lasted for 5s. In the behavioral adaptation session, participants were required to press the buttons—F, J, and 6—on the QWERTY keyboard to complete the category identification tasks. These three buttons form a "triangle", corresponding to the labels presented in the response screen (see Fig. 2). Participants used the index finger of one hand, and the index as well as the middle fingers of the other hand to press these buttons. The response hands were counter-balanced across the participants.

The complex-structure phase unfolded in a similar way, in which the grammaticality judgement sub-phase contained 16 testing blocks with 6 trials per block, and each trial lasted for 7.5s, while the category identification sub-phase only contained two practicing blocks just before scanning. Of note, participants were instructed to memorize the word categories of the pseudo-content words, but not the structure types. And only half of the complex syntactic structure types were randomly selected (but keeping the same amount of VPs and NPs) in the complex-structure phase of the behavioral adaptation session (see Supporting *Information* for all the complex structure types in Tables S2 and S3), and the category identification sub-phase only contained two practicing blocks with 12 trials in total, which were unlikely to induce the mechanical memorization strategy with regard to the complex structures.

Moreover, according to the definition of Merge mentioned in *Introduction*, the category identification task, especially in the complex-structure phase, required participants to read the visually-presented structures and to (a) identify the lexical items with their syntactic categories to be merged while checking whether there are syntactic violations (practiced during the grammaticality judgement sub-phase in advance), (b) for each time of Merge, search the Head of the phrase, and to (c) label the whole structure with a specific syntactic category (NP or VP). Given the structure complexity and variety, as well as the settings of syntactic violations, non-syntactic (such as template-matching and keyword searching) strategies were avoided, and participants had to hierarchically process the structures via the Merge mechanism. This was also confirmed according to the post-test interview report that participants would access the syntactic categories of the elements and "chunk" the mergeable ones into larger constituents, and finally label the root syntactic category of the whole structure as expected.

The detailed timing parameters for each trial of these (sub-)phases can be found in Fig. 2A.

As for the scanning session (Fig. 2B), we employed a blocked-design. This is because this study aims to detect and specify the neural correlates for the hierarchical syntactic processing on the basis of function words in Mandarin Chinese. Therefore, blocked-design with a higher detection power was adopted when compared with the ER (event-related) design (Huettel et al., 2004). This was also in line with several related studies (e.g., Matchin et al., 2017), so as to make the results more comparable. Besides, each block lasted for 54 s. This is in line with Opitz and Friederici (2003, 2004) in which robust activation of Broca's area (esp., BA 44) was reported for syntactic processing during natural language-like artificial language tasks using even longer blocks (70 s per block).

We divided the scanning session into three runs. Within each run, 4 *structure* blocks and 4 *word-list* blocks were pseudo-randomly organized, each block containing 6 trials (structures: 2 grammatical verb phrases, 2 grammatical noun phrases, and 2 ungrammatical structures; word lists: 2 function-word lists, 2 pseudo-content-word lists, and 2 violation word lists; for a similar setup see Matchin et al., 2017). Thus, there were 72 trials in total for each condition. For each trial of the structure condition, a left-sided fixation "+" appeared for 300 ms, indicating the initial reading position (i.e., from left to right), and then a blank screen (200 ms) followed. Subsequently, a structure was presented for 6000 ms (see also Opitz & Friederici, 2007 for a similar presentation manner), during which participants were asked to combine the words/pseudo-words together to retrieve the resulting syntactic category. A response screen was shown afterwards, and three labels were presented separately — "V" for verb phrase, "N" for noun phrase, and "※" for violation. Participants were asked to press a button corresponding to the position of the label with the index finger of one hand, and the index finger as well as the middle finger of the other hand, as they did in the behavioral adaptation session. To note, the positions of the three labels were pseudo-randomized, and the finger assignments were also counter-balanced across participants. The participants had a maximum of 2000 ms to respond, after which there was a 500-ms inter-trial-interval.

The word-list trials were presented in the same manner as the structure trials, except that participants could be able to identify the categories of the word lists without resorting to any syntactic processing. The word list categories/identities were as following: "A" for "artificial word lists" (i.e., the pseudo-content-word lists), "C" for "Chinese word lists" (i.e., the function-word lists), and "※" for word lists containing outlier symbols. Note that these labels together with the labels of the structure condition were well instructed before the scanning session.

In between each block, there was a cue composed of 7 pound signs "#######" (to match the length of the complex structures or the control word lists) at the center of the screen, lasted for 30 s. This cue was designed to signal the subsequent condition by changing its color: red for the structure condition, and blue for the word-list condition. Thus, participants could have enough time to adapt to the target block according to the cue (see also Matchin et al., 2017). The whole scanning session for each participant was approximately 1 h, comprising of experimental instruction, preparations, and the actual scanning process.

2.4. Behavioral data analyses

Behavioral data were acquired in the scanner. We separately performed one-sample t-tests to examine whether the accuracy of both conditions and the accuracy of the ungrammatical/violation conditions could significantly surpass the chance level (i.e., 50%). Paired-sample t-tests were also performed to compare participants' accuracy and reaction time (RT) in both conditions. Unless otherwise specified, the multiple-comparison results were corrected via "false discovery rate" (FDR)-correction. All these statistical analyses were performed with SPSS 26 (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp) and jamovi (Version 1.6, https://www.jamovi.org).

2.5. Imaging data acquisition

The MR imaging data were acquired via a 3.0-T Siemens PRISMA magnetic resonance scanner (Siemens AG, Erlangen, Germany) using a 64-radiofrequency-channel head coil.

For functional data acquisition, a T2*-weighted gradient echo planar imaging (EPI) sequence was adopted with the following parameters: repetition time (TR) = 2000 ms; echo time (TE) = 30 ms; flip angle (FA) = 90°; field of view (FOV) = 208 × 208 mm^2base resolution = 104 × 104 mm^2in-plane resolution = 2 × 2 mm^2slice thickness = 2 mm; number of slices = 64; gap = 0 mm; alignment to AC-PC plane. Signals from different slices were acquired by the multi-band scanning technique (multi-band factor = 2) to efficiently minimize slice-timing effects.

Parameters for high-resolution anatomical T1-weighted images for co-registration were listed as following: TR = 2530 ms; TE = 2.27 ms; FA = 7°; FOV = 256 × 256 mm^2; base resolution = 256 × 256 mm^2; in-plane resolution = 1 × 1 mm^2; slice thickness = 1 mm; number of slices = 208.

2.6. Imaging data preprocessing

The imaging data were preprocessed by DPARSF 5.1 Advanced Edition (DPARSF: Data Processing Assistant for Resting-State fMRI, Yan et al., 2016), implemented in the environment of MATLAB R2020b. The preprocessing steps were following Yan et al. (2013), and included: (a) Removing the first 4 vol to reduce the magnetic saturation effect; (b) Slice time correction; (c) Field mapping; (d) Spatial realignment; (e) Co-registration; (f) Segmentation (New segment + DARTEL); (g) Nuisance covariates regression (polynomial trend: 1, linear detrending), also including head motion regression by using the Friston-24 model; (h) Normalization of the images to the echo planar imaging (EPI) template based on Montreal Neurological Institute (MNI) stereotactic space to minimize cerebral differences between participants, and resampled the images into 2 × 2 × 2 mm^3; (i) Smoothing the images with a 3D Gaussian kernel with full-width at half-maximum (FWHM) of 4 mm.

2.7. Whole-brain level analyses

The whole-brain level analyses were performed by using SPM 12 (https://www.fil.ion.ucl.ac.uk/spm/). At the first level, a general linear model (GLM) was set up for each participant by adding the structure and the word-list conditions as two regressors of interest, with the onset and duration (54 s) of each block modulated as a boxcar function convolved with a canonical hemodynamic response function (HRF). The data were further high-pass filtered at 128 Hz to eliminate low-frequency drifts. The "structure > word list" contrast results of each participant were entered into the second level analysis.

At the second level for group analysis, we were interested in the one-sample *t*-test against the null hypothesis that there was no difference between the structure and the word-list conditions. As previously proposed (Yan et al., 2013), we also took each individual's mean FD (framewise displacement) Jenkinson vale as a covariate to regress out the head motion artifacts at the group-level. Moreover, in case of the "time-on-task effect", we further included the reaction time as a regressor of no interest for the group-level analysis (see also Wu et al., 2018). Besides, the accuracy would be further regressed out, and this was to remove the potential effects caused by the potential difficulty discrepancies (see also Chen, Goucha, et al., 2021). The results of the whole-brain analysis were reported with a cluster-level FWE (family-wise error)-corrected threshold of $p < .05$, using the cluster defining uncorrected threshold at $p < .001$, cluster size (K_E) ≥ 35. In addition, activation under each condition was detected ($p_{uncorrected} < .00001$, $K_E \geq 50$) in contrast with the implicit baseline (i.e., 0) respectively so as to ensure the normal processing of the two conditions.

2.8. Region of interest (ROI) analyses

Since we were most interested in the "structure > word list" activation differences within the language network, a 220 participant-based functional left-hemispheric language atlas extended from Fedorenko et al. (2010) (http://web.mit.edu/evlab//funcloc/) was adopted as the language mask for Small Volume Correction (SVC) to identify the peak activity coordinates (see also Chen, Goucha, et al., 2021) (cluster-level $p_{FWE} < .05$, using the cluster defining threshold at $p_{uncorrected} < .001$, $K_E \geq 30$). The percentage of signal change was measured on the basis of the *original* BOLD (Blood Oxygenation Level Dependent) signal without regressing out the behavioral indices (i.e., accuracy and RT) in each ROI, and their correlations to the behavioral indices were tested, see Supporting *Information Section 4* for details. It is noteworthy that we controlled for cognitive load differences between the different conditions at the whole-brain level, by regressing out the behavioral scores for each participant (see also Chen et al., 2021) when comparing the hypothesized BOLD signal with the real signal for each condition of interest. When we conversely extracted the signal change percent for each ROI under analysis was instead divided the raw "task signal" (i.e., the structure condition signal) by the "mean signal" (i.e., the averaged signal of all the conditions), to examine the direct correlational relationship between signal change and behavioral scores in a certain ROI.

2.9. Effective connectivity analysis

In order to track further the information transfer among the ROIs, we adopted the unified structural equation modeling (uSEM) approach for the effective connectivity analysis (Gates et al., 2011; Kim et al., 2007). The uSEM is implemented by the MATLAB-compatible toolkit, Group Iterative Multiple Model Estimation (GIMME; Gates et al., 2011; Gates & Molenaar, 2012), an exploratory, hypothesis-free approach with unbiased parameter estimation. It can provide more time-specific information, including the lagged effect (i.e., a longitudinal effect from a previous time point to the current one), and the contemporaneous effect (i.e., effects at the same time point) (Beltz & Gates, 2017; Gates et al., 2011; Gates & Molenaar, 2012; Kim et al., 2007). Such a structural equation modeling approach for the analysis of effective connectivity has been well established in recent language learning and processing studies (Chen, Goucha, et al., 2021, 2021b; Wu et al., 2019; Yang et al., 2015).

Accordingly, we extracted the *original* structure condition signal from each ROI (without regressing out the behavioral indices) and included them in GIMME for model specification. GIMME estimates the optimal models from the group-level to the individual-level via Lagrange multiplier tests. For each level, should a connection improve the model fit significantly ($\geq$75% of the sample), it was added to the model for re-estimation. The detailed parameter setup for uSEM was following Beltz and Gates (2017), Gates et al. (2011), and Gates and Molenaar (2012).

3. Results

3.1. Behavioral results

The behavioral results were summarized in Table 4 and Fig. 3. The accuracy of the structure condition was significantly higher than the chance level ($t(20) = 9.23, p < .001, d = 2.01$), and was not statistically different from 80% ($t(20) = -1.080, p = .293$), indicating that the participants were compliant with the experiment task. Besides, the accuracy of the ungrammatical trials also exceeded the chance level ($t(20) = 5.91, p < .001, d = 1.29$), and a *post-hoc* paired-sample *t*-test further revealed that there was no statistical accuracy difference between the grammatical and the ungrammatical structures ($t(20) = 0.624, p = .54$). This indicated that both types of structure were well processed. Nevertheless, the accuracy of both types of structures did not show a ceiling effect, which indicated the difficulty of the syntactic task, and it was unlikely to be solved by easier non-syntactic strategies. Moreover, participants needed significantly longer time to process the grammatical structures ($t(20) = 5.51, p < .001, d = 1.2$). This might be due to the fact that grammatical structures should be processed to the very end so as to merge the whole structures, whereas ungrammaticality can be detected as soon as the violation occurred, during which only local merge was required.

The word-list condition showed a similar pattern: (a) Above-chance accuracy ($ts(20) \geq 21.5, ps < .001, ds \geq 4.69$); (b) No accuracy differences between the normal and violation word lists ($t(20) = -1.54, p = .139$); (c) Higher RT for the normal word list processing ($t(20) = 2.32, p < .05, d = 0.51$).

In addition, the paired-sample t-tests showed that the structure condition had a lower accuracy than the word-list condition ($t(20) = -6.94, p < .001, d = 1.52$), but with a longer RT ($t(20) = 6.90, p < .001, d = 1.50$). This indicated that the structure condition was more difficult to process than the word-list condition. Consequently, both RT and accuracy were included as covariates to define the group-level design matrix.

3.2. Whole-brain level results

Each condition showed reliable activation when compared with the implicit baseline, guaranteeing that both structure and word-list conditions were normally processed. Moreover, under the single structure condition, a large cluster containing the left PreCG and IFG was detected.

The group-level analysis further found significant activity of the left pars opercularis of inferior frontal gyrus (IFGop), functionally corresponding to the left BA 44 at the whole-brain level (see Table 5 and Fig. 4B_1), for the contrast "structure > word-list". This indicated that compared to the mere word-list condition, the hierarchical syntactic processing mechanism underlying the structure condition might heavily rely on the left BA 44.

Table 4
Behavioral results.

Conditions	Structure				Word list			
	Mean (*SD*)	95% CI	Grammatical (*SD*)	Ungrammatical (*SD*)	Mean (*SD*)	95% CI	Normal (*SD*)	Violated (*SD*)
Accuracy	0.77 (0.13)	[0.71, 0.83]	0.78 (0.13)	0.75 (0.2)	0.91 (0.08)	[0.88, 0.95]	0.91 (0.09)	0.93 (0.09)
RT (ms)	1093.49 (157.92)	[1021.61, 1165.38]	1132.76 (170.21)	1017.95 (160.13)	962.18 (122.72)	[906.32, 1018.04]	985.76 (145.70)	918.71 (132.82)

Abbreviations: SD: standard deviation; CI: confidence intervals; RT: reaction time.

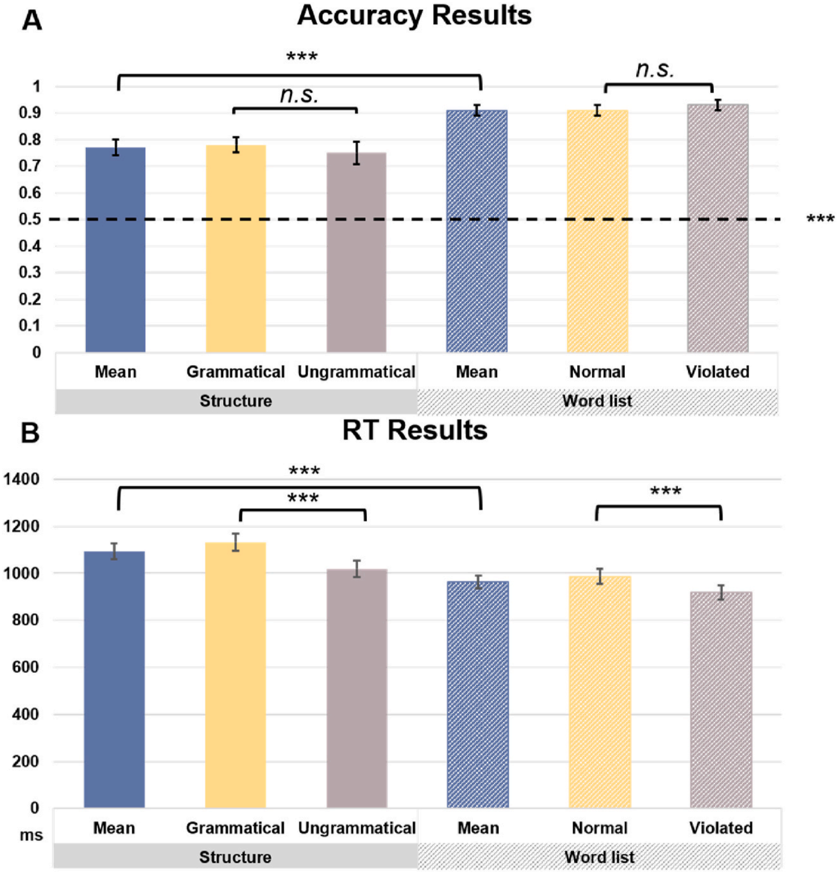


Fig. 3. Behavioral results. (A) The accuracy result for each condition. (B) The RT result for each condition. Under the structure/wordlist condition, the mean accuracy/RT of all trials, the grammatical/normal trials and the ungrammatical/violated trials were displayed. The error bar showed the standard error of the mean. Abbreviations: accuracy: accuracy rate; RT: reaction time; n. s.: non-significant, $p > .05$; ***: $p < .001$.

3.3. ROI analyses results

By employing the functional language atlas, the SVC results identified three peak-activity coordinates located in the left BA 44 (−46, 16, 26), the left precentral gyrus (PreCG, −44, 2, 46), and the left posterior temporal cortex (pTC, −52, −48, −2) respectively (see Table 5 and Fig. 4B$_2$). Notably, IFGop refers to the left opercular part of the inferior frontal gyrus, and preCG refers to the left precentral gyrus. Both labels indicate the locations corresponding to the peak coordinates, and they were automatically defined by the atlas Neuromorphometrics 1.0 (http://Neuromorphometrics.com/) implemented in SPM 12 (https://www.fil.ion.ucl.ac.uk/spm/). The peak coordinate (−52, −48, −2) in the posterior temporal cortex (pTC) could not be reliably labeled in SPM 12, and thus we labeled this coordinate as pTC in a relatively broad sense due to the fact that the cluster extended from the posterior middle temporal gyrus to the left superior temporal gyrus. (A semantic deactivation pattern was also detected within the language atlas, see Supporting *Information Section 2*.)

3.4. Effective connectivity modeling results

Model fit indices indicated that the group-level effective connectivity model was reliably estimated ($CFI = 1.00$, $NNFI = 1.00$) (see also Chen, Goucha, et al., 2021, 2021b; Gates et al., 2011; Wu et al., 2019; Yang et al., 2015). The uSEM results revealed a neural circuit, in which: (a) the left BA 44 projected a contemporaneous connection to the left PreCG; (b) the PreCG further projected both contemporaneous and lagged connections to the left pTC; (c) the pTC transferred the information back to BA 44 via a contemporaneous connection. All these results were summarized in Fig. 4C.

4. Discussion

In this study, we aimed to specify the hierarchical syntactic processing mechanism—fundamental to human language—in Mandarin Chinese. We asked whether hierarchical syntactic processing in Chinese also engaged the left posterior IFG, especially, BA 44,

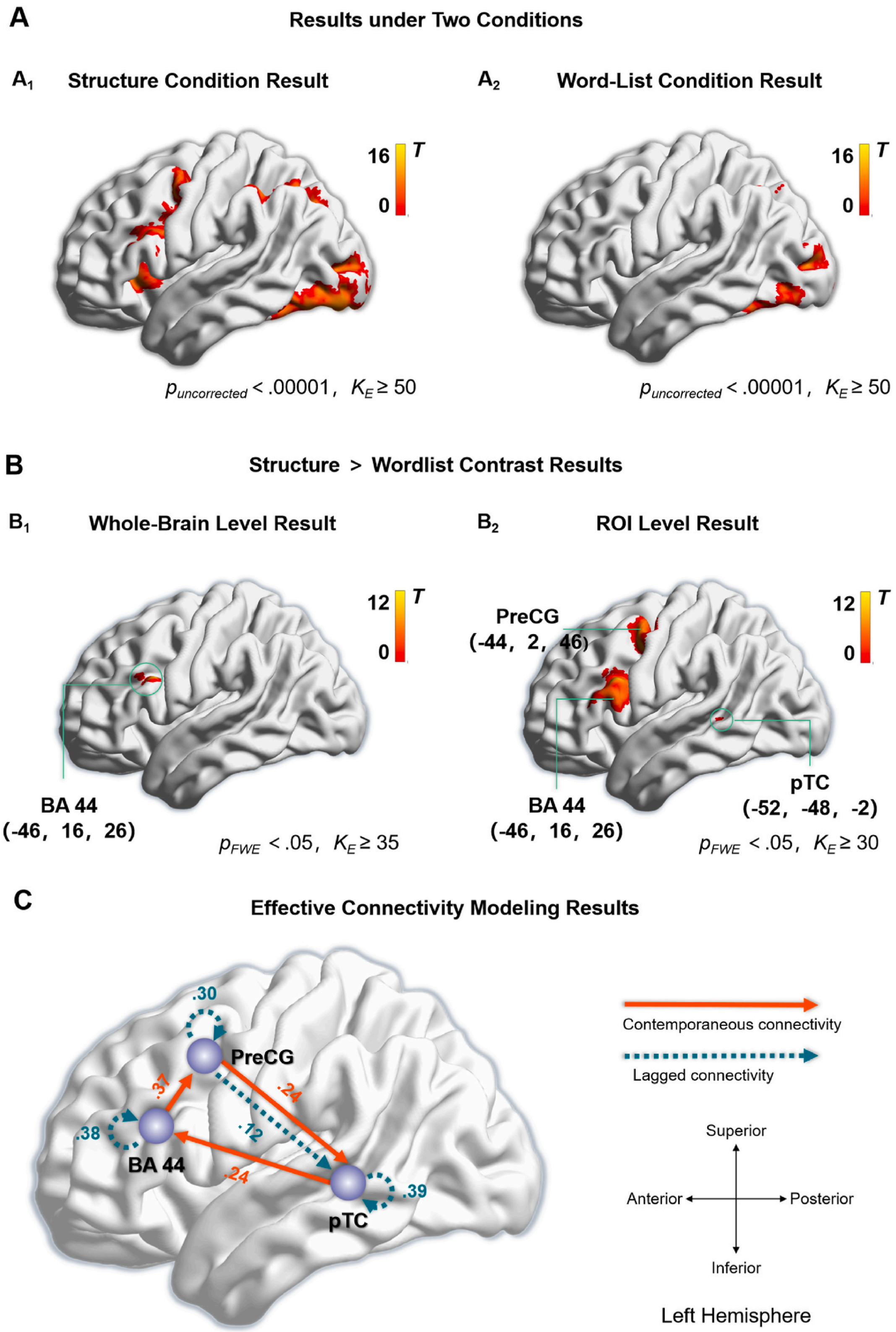
Table 5
Whole-brain level and ROI-level results.

Contrasts	Region	K_E	MNI Peak Coordinates (mm)			*T-value*
			x	y	z	
Whole-brain level ($p_{uncorrected} < .00001$, $K_E \geq 50$)						
structure >0	RIOG	6319	30	−86	−4	15.19
			16	−82	−22	14.60
			−4	−80	−6	13.28
	LAIns	136	−30	22	2	12.18
			−28	28	−4	9.56
	LPreCG/IFG	776	−44	2	30	11.46
			−50	8	32	10.60
			−44	16	26	10.42
	RSPL	384	28	−60	36	11.45
			32	−62	48	10.74
			28	−68	38	8.30
	LHippocampus	77	−22	−30	−8	11.38
			−30	−30	−6	8.14
	LSMC	362	−6	16	46	11.02
			2	18	46	9.77
			0	4	54	8.62
	LPutamen	200	−20	6	6	10.70
			−12	−18	6	8.83
			−16	−8	12	8.77
	RAIns	143	32	24	−2	10.08
			40	14	6	6.77
	RHippocampus	70	24	−26	−10	10.04
word list >0	RLiG	1965	4	−82	−2	13.59
			12	−78	−10	13.21
			6	−74	4	12.21
	ROFuG	198	32	−62	−22	11.07
			42	−62	18	9.02
			36	−48	−32	8.1
	LMOG	133	−30	−94	8	10.94
			−28	−86	12	9.92
	LAG	73	−28	−70	48	9.52
			−30	−54	40	7.2
			−34	−62	48	6.17
	LITG	213	−44	−62	−18	8.73
			−44	−72	−26	7.77
			−42	−78	−18	7.76
	RIOG	89	38	−82	−8	8.07
			46	−70	−12	7.78
			42	−78	−18	6.42
Whole-brain level ($p_{FWE} < .05$, $K_E \geq 35$)						
structure > word list	LIFGop	36	−46	16	26	11.09
ROI-level (within the language atlas) ($p_{FWE} < .05$, $K_E \geq 30$)						
structure > word list	LIFGop	540	−46	16	26	11.09
	LPreCG	274	−44	2	46	8.44
	LpTC	32	−52	−48	−2	4.45

Abbreviations: RIOG: right inferior occipital gyrus; LAIns: left anterior insula; LPreCG: left precentral gyrus; IFG: inferior frontal gyrus; RSPL: right superior parietal lobule; LHippocampus: left hippocampus; LSMC: left supplementary motor cortex; Lputamen: left putamen; RAIns: right anterior insula; RHippocampus: right hippocampus; LliG: left lingual gyrus; ROFuG: right occipital fusiform gyrus; LMOG: left middle occipital gyrus; LAG: left angular gyrus; LITG: left inferior temporal gyrus; LIFGop: left pars opercularis of inferior frontal gyrus; LpTC: left posterior temporal cortex; K_E: cluster size.

and the left posterior temporal cortex when semantic interference was controlled. To this end, we developed a function-word-based syntactic processing paradigm deprived of conceptual-semantic information. The whole-brain analysis revealed a significant activation cluster in the left BA 44. When constraining the level of analysis to the language network alone, two further regions were identified, the left PreCG and the left pTC. The three regions were highly interconnected, according to the effective connectivity analysis. It is indeed intriguing that even though the pseudo-content words are unpronounceable in Chinese, a language distinct from alphabet-spelling languages like German and English, the Chinese syntactic network identified in this study is consistent with the ones reported for those languages. These results corroborate the idea that syntactic processing relies on a specific frontotemporal syntactic network across human languages, and that the current effective connectivity model is well compatible with prominent neurocognitive models for language processing (Friederici, 2017).

Before assessing the functions of the specific brain regions, relative concerns about the ecological validity of the jabberwocky sentence paradigm adopted in this study are discussed (cf. Pylkkänen, 2019). First, we claim it is common that during natural language processing, novel words (esp., unfamiliar content words) cannot always be immediately processed (such as new words in a second



(caption on next page)

Fig. 4. Imaging results. (A) The whole-brain level result for each condition. (B) "structure > word list" results: B_1: at the whole-brain level; B_2: the small volume correction result at the ROI level (3 regions of interest were identified). (C) Effective connectivity modeling results via uSEM. Group-mean connectivity strength (i.e., beta value) was also presented for each connection. Abbreviations: BA 44: Brodmann Area 44; PreCG: precentral gyrus; pTC: posterior temporal cortex; K_E: cluster size.

language reading situation), but depending on the function words, the parsing system may still be able to build up hierarchical syntactic structures by assigning the syntactic categories to the words as well as the phrases. Moreover, in the present study, syntactic categories of the pseudo-content words were acquired before the actual experiment and thus accessible to ease processing difficulty (i.e., participants did not need to speculate on the syntactic categories of the content words according to the function words while reading the materials). Moreover, participants were not trained to memorize the structure types, and due to the complexity and variety of the structures used in the scanning session, participants could not complete the task via mechanical memorization. Furthermore, as shown by the imaging results under the "structure > word list" contrast (Table 5), typical memory regions (esp., the parietal lobe) were not significantly activated, backing up the claim that participants were unlikely to resort to memory strategies to cope with the task.

Besides, that the pure syntactic processing experiment utilizing the jabberwocky sentence design seems to be more difficult is considered to be caused by the lack of semantic information, whereas the syntactic information is still available. Therefore, similar to the studies utilizing syntactic violations (Schell et al., 2017; Wu et al., 2018; Zaccarella, Meyer, et al., 2017; Zaccarella & Friederici, 2015) or more complex syntactic structures (den Ouden et al., 2012; Makuuchi et al., 2009; Santi & Grodzinsky, 2010; Wang et al., 2021) in which syntactic processing became more difficult, such a syntax-specific difficulty indeed provides us with a chance to investigate the neural substrates for syntactic processing. This is also evidenced by previous studies adopting jabberwocky sentences (Goucha & Friederici, 2015; Matchin et al., 2017; Ohta et al., 2013; Pallier et al., 2011; Tyler et al., 2010). For example, Pallier et al. (2011) identified that sentences with real function words retained could still activate the syntactic network as contained in the activation pattern of natural sentences.

Hence, we suggest that the jabberwocky sentence paradigm might be tentatively acceptable for investigating the neural basis of syntactic processing, by keeping the real function words as well as the real syntactic structures, more ecologically valid than the artificial grammar designs, and by suppressing semantic processes involved in natural language designs, thus, to some extent, balancing ecological validity and the purity of syntactic processing.

4.1. *BA 44: the core region for syntactic hierarchy construction*

Recent studies on syntactic processing in Chinese identified a broad left-hemispheric activation pattern (Chang et al., 2020; Chee et al., 1999; Chou et al., 2012; Sun et al., 2021; Wang et al., 2008; Wu et al., 2018). For example, Chang et al. (2020) parametrically modulated the constituent size for both Chinese sentences and nominal phrases, and found that the nominal phrases involved the left BA 44. By comparing the basic phrase structure with the word-list condition, the dynamic causal modeling (DCM) analysis by Wu et al. (2018) suggested that BA 44 might also play a critical role in the Chinese syntactic system. Sun et al. (2021) also found activation of the left IFG by adopting a syntactic priming paradigm. Additionally, they detected the activation in the left anterior temporal lobe, a region well documented for conceptual-semantic combinatorics (see Pylkkänen, 2019 for a review). All these studies used natural materials, which might unavoidably cause semantic processing, due to the fact that Chinese is highly context-semantics dependent, and that semantic processing might highly interact with syntactic processing as evidenced by the ERP (event-related potential) findings (Wang et al. 2013, 2015; Yu & Zhang, 2008; Zhang et al., 2010).

In the present study, we highlighted the importance of the function words to carry syntactic information in Chinese. Noteworthy, the function words of the current study were typical and highly abstract, with no concrete conceptual or pragmatic meanings, and independent of the context (Guo, 2018; Huang and Liao; 2011; Ma, 2015; Tang, 2019; Zhu, 1982, 1985). Nevertheless, a concern was raised that participants may have completed the syntactic processing task by making associations based on the true word contents, in that function words might also have certain kinds of "meanings". For example, "着, 了, 过" (/tʂə/,/lə/,/kuo4/) were claimed to have the so-called "tense meanings". However, they are also implied by the inflectional morphology such as those suffixes in "he work-*s*/is work-*ing*/work-*ed*". Such tenses (e.g., past, present, and future) and aspects (e.g., progressive, perfective) of verbs are theoretically known as "grammatical meanings" (Huang & Liao, 2011; Jin, 2002; Li, 2021; Shao & Zhao, 2006), which are incomparable with the conceptual meanings conveyed by the content words. Besides, the quantifiers used in this study are also abstract without clear measurement meanings (Chen, 1998; Guo, 1987; Ma, 2015; Tang, 2019). For instance, "个" can be generally applied to various kinds of objects. In addition to the theoretical considerations, the control condition contained function-word lists, in which for the Chinese native speakers, "meanings" of the function words were assumed to be automatically accessed. Thus, under the contrast of "structure > word list", the semantic effects of the function words might be dramatically reduced. Moreover, as mentioned in the section on *Materials*, it should be very difficult for the participants to associate the word tokens to the corresponding syntactic categories by establishing the semantic associations. And the presentation time (6 s) for 7-word structures was rather limited for the participants to generate semantically plausible expressions for the whole structures. As confirmed by the post-test interviews, some rare attempts to add semantic content were present, but they were very vague and rather implausible, and meanings could not be correctly integrated into the grammatical structures (see also Tyler et al., 2010 for the description of their jabberwocky sentence condition as "no discernible meaning" could be obtained). Furthermore, as demonstrated by the "word list > structure" contrast in the Supporting *Information*, classic semantic areas such as the anterior temporal lobe and the angular gyrus were deactivated (i.e., their activation was highly constrained), which was quite similar to the findings of a recent artificial grammar processing study (Chen, Goucha, et al., 2021). Therefore, due to the vagueness of the whole structure meanings, semantic strategies were highly suppressed, unlikely to be an efficient way to complete the syntactic task.

The imaging result showed that only the left BA 44 robustly survived at the whole-brain level. A post-hoc individual activation analysis within the left IFG of the functional language atlas was also performed and demonstrated a congruent activation pattern within this region (see Supporting *Information Section 3* for details). Since the structure condition is more cognitively demanding than the word-list condition, might the activation of the left BA 44 be merely explained by the heavier general cognitive loads? We considered that the hierarchical syntactic processing mechanism should be "downward compatible" to (i.e., contain) the low-level word list processing mechanism (see also Chen, Goucha, et al., 2021 for a similar comment with regard to hierarchical syntactic processing and the associating mechanism), and that is why a number of studies adopted the word lists as the control condition (see also Fedorenko et al., 2010, 2011). Therefore, the detection of the phrase violation (i.e., the word category violation) in the structure condition should also contain the low-level word list processes (esp., detection of the outlier). Both violation detection processes contained the identification of the words as well as the detection of the violated element(s), and the key difference is that for the word-list condition, participants only needed to detect the outlier concatenated with its neighbor words without Merge; whereas for the structure condition, they should assess whether the word categories were mergeable to form grammatical phrases. Thus, the Merge process would be experimentally highlighted by the comparison between two types of violations. Moreover, even though this study enhanced the complexity of materials from basic phrase structures, as adopted by Schell et al. (2017), Wu et al. (2018), Zaccarella and Friederici (2015), and Zaccarella, Meyer, et al. (2017) in the basic Merge paradigm, to complex syntactic structures with multiple hierarchical embeddings in Chinese, BA 44 was the only surviving region after the whole-brain correction. This finding further confirmed that, as pinpointed by Zaccarella, Meyer, et al. (2017), BA 44 functioned as a Merge engine to support hierarchical syntactic processing regardless of the syntactic complexity. In this sense, such cognitive load might be specifically induced by hierarchical syntactic processing, and should reflect the neurobiological essence of the specific hierarchical syntactic processes. The rationale for "structure > word list" contrast is following Fedorenko et al. (2011), in which hard tasks were compared with easier ones in various domains (e.g., for language: language sentences vs. word lists, for mathematics: a hard mathematical task vs. a simpler one, and for visual working memory: a hard visual working memory task vs. a simpler one), and the results revealed distinct activation patterns for those different cognitive domains. Their findings indicated that a difficult task in a given domain should always accompany with the simple task in the same cognitive domains. For instance, the activation of the comparison between hard and easy working memory tasks should be interpreted as "working memory activation" rather than a general executive control activation possibly induced by the general task difficulty or complexity differences. It is also noteworthy that in the current activation pattern, no significant activation could be identified in the left middle frontal and parietal lobes at the whole-brain level, which were assumed to be the core regions composing the general cognitive control network (Feng et al., 2022; Uddin et al., 2019).

The imaging result that BA 44 was highly activated converged with former findings from languages employing morphosyntactic markers. For instance, in German jabberwocky sentences with inflectional suffixes retained, Goucha and Friederici (2015) found that the left BA 44 was activated for the "syntactic skeleton" processing. Similarly, Ohta et al. (2013) also identified the activation of the left BA 44 for the Japanese jabberwocky sentences only with the case markers preserved. Our findings demonstrated that Chinese function words might play a critical role in guiding the build-up of the hierarchical syntactic structures in Chinese, functionally resemble the syntactic markers in languages with rich morphosyntactic information. The present finding of the left BA 44 activation was also in line with the results from former studies using artificial grammar learning (AGL) paradigms which were independent of any specific languages (e.g., Bahlmann et al., 2008, 2009; Chen, Goucha, et al., 2021; Friederici et al., 2006; Uddén et al., 2017).

Taken together, the activation of BA 44 was unlikely to be induced by mere semantic processing or general cognitive demands. These confounders should not interact with the experiment conditions to submerge the syntactic effects in this study. The results indicated that hierarchical syntactic processing, mainly housed in the left BA 44, could be triggered regardless of specific syntactic representations, be it morphosyntactic affixes, function words, or even artificial grammar elements, further specifying the syntactic role of the function words in Chinese under the neurocognitive perspective.

4.2. *The language-general syntactic network*

The ROI analyses of the present study indicated that the PreCG and the pTC, along with the BA 44, formed a functional activation pattern within the left hemisphere. We used the uSEM approach to analyze the effective connectivity between these regions, which pointed toward the existence of a functional neural circuit that may implement syntactic hierarchies in a language-general fashion.

The pTC covers both the superior and the middle parts of the left-hemispheric posterior temporal cortex, including the superior temporal gyrus (STG), superior temporal sulcus (STS), and middle temporal gyrus (MTG). Such an activation pattern is close to the findings of Wu et al. (2018), in which the posterior MTG was activated and significantly connected to the left BA 44 for the Mandarin phrase structure building. The pSTG/STS has been proposed to play a critical role in integrating syntactic and semantic information during complex syntactic structure processing (Bornkessel et al., 2005; den Ouden et al., 2012; Friederici et al., 2009; Friederici, 2011), especially when the syntactic elements were complex or difficult to be integrated into the structures (Constable et al., 2004; Cooke et al., 2002; Friederici & Kotz, 2003). Nevertheless, studies adopting artificial grammars or jabberwocky sentences have further demonstrated that complex syntactic structure processing with semantics deprived could still drive the activation of these posterior temporal regions (e.g., Bahlmann et al., 2008; Chen, Goucha, et al., 2021; Friederici et al., 2006), thus providing extra evidence that pSTG/STS might be involved in the general integration of various types of linguistic information, be it syntactic or semantic. This is also in line with the assumption of Hagoort (2005, 2013) that the posterior temporal cortex might integrate and therefore store the linguistic elements varying from lexical items to abstract syntactic frames. Moreover, as pinpointed by Wu et al. (2018), pMTG might also send feedback to BA 44 for syntactic congruence check, and pMTG together with BA 44 might establish a pivotal connection for the build-up of phrase structures. Matchin and Hickok (2020) further pinpointed that the pMTG might play a critical role in converting an incoming linear sequence into nonlinear hierarchical structures. This is, to some extent, in agreement with the functional gradient hypothesis of Hagoort and Indefrey (2014) that there might be a syntactic gradient extending from the pSTG to pMTG for higher syntactic demands. Given the functional similarity between pSTG/STS and pMTG, we, hereby, synthesized these sub-regions together to form a complex of "pTC", and proposed an integration role of this region in a broad sense, that is, the pTC might function to integrate the syntactic elements generated from the left BA 44 into larger syntactic representations and to send them back to BA 44 for subsequent syntactic operations. Post-hoc exploratory Pearson correlation tests were performed to examine the relationship between the contemporaneous connection strength and the behavioral indices, and the results revealed that the strength of the pTC-to-BA 44 connection was significantly correlated with the accuracy of the structure condition (*see* Supporting *Information Section 4 for the correlation results*), thus further confirmed the crucial role of this connectivity in syntactic processing as reported by previous studies on other languages for hierarchical syntactic processing (Goucha & Friederici, 2015; Matchin et al., 2017; Ohta et al., 2013; Pallier et al., 2011; Tyler et al., 2010; see also Hagoort & Indefrey, 2014 and Zaccarella, Schell, & Friederici, 2017 for the meta-analysis results).

Previous studies on (complex) syntactic structure processing or syntactic priming/repetition occasionally identified the involvement of the left PreCG (e.g., Bottini et al., 1994; Crozier et al., 1999; Lee & Newman, 2010; Segaert et al., 2012, 2013; Snijders et al., 2009; Sun et al., 2021; Weber & Indefrey, 2009). This region was presumed to be a terminus of the arcuate fasciculus for sensory-motor mapping (Friederici, 2017) with plenty of studies reporting its role in speech programming (e.g., Bernal & Ardila, 2009; Buchsbaum et al., 2005; Lindenberg & Scheef, 2007), as well as in silent reading (Price et al., 1994; Kaestner et al., 2021; see also Price, 2012 for a review). Moreover, Studies on the location of the phonological loop (as part of the working memory system) proposed that BA 6 or more broadly the premotor as well as the supplementary areas might house the phonological working memory (e.g., Baddeley, 2003; Smith & Jonides, 1997; Vallar & Papagno, 2002). It is also noteworthy that, the PreCG was named according to the peak coordinate under the contrast of "structure > word list", which was detected within the left middle frontal gyrus (LMFG) from the functional language atlas (http://web.mit.edu/evlab//funcloc/, see also Blank et al., 2014; Fedorenko et al., 2010, 2011), and as Fedorenko et al. (2011) pinpointed, in addition to sentence processing tasks, the LMFG also responded to verbal working memory tasks. Here we reasoned that even though at the lexical level phonological effects on the function words might be reduced by subtracting away the word-list condition effect, function words might boost on-line maintenance structures at the sentential level in the region. Therefore, since real Chinese function words were retained in the materials, one interpretation for its activation is that the PreCG might be engaged in phonological working memory, especially for representing and maintaining the complex structures (see also Hickok et al., 2003; Hickok & Poeppel, 2007). Such a phonological working memory in PreCG might facilitate the integration process from BA 44 to pTC, and together these regions might constitute a language-general syntactic network within the overall shared language network (Malik-Moraleda et al., 2022). Nevertheless, we are unaware of systematic studies as well as detailed discussions on the role of PreCG in complex sentence/syntactic processing, and therefore would like to leave the function of this region open for future explorations.

To sum up, the current tentative neural circuit model for (recursive) hierarchical syntactic processing in Chinese seems to serve as a key frontotemporal network backing up the unique human language faculty, independent of language typological syntactic differences.

4.3. Limitations

Jabberwocky sentences help *narrow down* the neurobiological scope of syntactic processing while preserving the actual grammatical structures of the specific languages under analysis. Even though semantic processing is highly suppressed in jabberwocky sentences, some sentential meaning might be still assigned to the structures, even if very vaguely (Tyler et al., 2010). However, the deactivation patterns reported for the semantic network in our study seem to indicate that our participants made very little attempt to attach meaning to the pseudowords used in the experiment (Supporting *Information Section 2*). Overall, we suggest that the jabberwocky sentence paradigm should be deemed as a tentative trade-off between a more focused investigation on the neural substrates of syntactic processing and the maintaining of the ecological validity of natural language structures.

Moreover, the contrast between the jabberwocky sentence condition and the word-list condition in this study and in previous ones (as meta-analyzed in Zaccarella, Schell, & Friederici, 2017) might confront with the critique that processing sentences requires higher cognitive load, including effects of expectation, attention, consciousness, and context, compared to the processing of word lists (see also Maran et al., 2022). As also discussed above however, language-specific regions defined by the "sentence > word list" contrast can be differentiated from the multi-task demand regions ("hard task > simpler/easy task" contrasts in various other domains such as math, music, working memory and so on: Fedorenko et al., 2011). Thus, the cognitive-load difference, if any, should be interpreted as language-specific in the present study. As such, we suggest that the effect leading to higher language-specific cognitive load for sentences in our study essentially reflects enhanced participants' reliance to syntactic information during the processing of structured sequences. Besides, behavioral indices reflecting different cognitive loads were also regressed out at the whole-brain level, and the activation patterns were in line with previous studies (e.g., Matchin et al., 2017). Nevertheless, the way core regions of the language network may functionally interact with the domain-general executive control areas during complex syntactic processing still remains a puzzling issue (see also Chen, Goucha, et al., 2021).

5. Conclusion

The present fMRI study aimed to specify the neural basis for the hierarchical syntactic processing mechanism in Mandarin Chinese, as identified by the studies on the languages with abundant morphosyntactic information. We developed a Chinese function-word-based syntactic processing paradigm with semantic effects highly suppressed to investigate this question, thus optimizing the previous neurolinguistic experiments on Chinese syntactic processing. Our results revealed that the left BA 44 might be a key region for syntactic hierarchy construction, and that the left hemispheric neural circuit including BA 44, pTC, and PreCG, might implement a key network for the recursive syntactic system in Chinese. The present findings further reinforced the idea of a universal selective syntactic network holding the core human language faculty regardless of the typological variances across languages.

Author contributions

Luyao Chen, **Chenyang Gao**, and **Liping Feng**: Came up with the original idea of the Chinese function-word-based syntactic processing paradigm, and designed the experiment. **Luyao Chen**, **Chenyang Gao**, and **Zhongshan Li**: Conducted the pilot and the actual fMRI experiment. **Luyao Chen** and **Chenyang Gao**: Analyzed the data. All authors participated in the discussion of the results. **Luyao Chen** and **Chenyang Gao**: Completed the first draft of this manuscript, which was further revised by **Liping Feng**, **Zhongshan Li**, **Emiliano Zaccarella**, and **Angela D. Friederici**. To note, **Luyao Chen** and **Chenyang Gao** contributed equally to the current work as co-first-authors.

Funding

This work was supported by the National Social Science Fund of China (Youth Project) (No.: 22CYY017).

Declaration of competing of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors wish to thank all participants who took part in this study, and are sincerely grateful for the anonymous reviewers' insightful comments. Special thanks are extended to Liaoyuan Zhang, Yang Liu, Siying Lin, Keyi Kang, Luping Wang, Yi Song, and Lu Li for their support in the behavioral and fMRI data collection. The authors also thank Junjie Wu, Siyuan Zhou, and other colleagues for their constructive input, and Joseph Oliver Simpson for English language polishing.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jneuroling.2023.101127.

References

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*, 829–839.

Bahlmann, J., Schubotz, R. I., & Friederici, A. D. (2008). Hierarchical artificial grammar processing engages Broca's area. *NeuroImage, 42*, 525–534.

Bahlmann, J., Schubotz, R. I., Mueller, J. L., Koester, D., & Friederici, A. D. (2009). Neural circuits of hierarchical visuo-spatial sequence processing. *Brain Research, 1298*, 161–170.

Beltz, A. M., & Gates, K. M. (2017). Network mapping with GIMME. *Multivariate Behavioral Research, 52*, 789–804.

Bernal, B., & Ardila, A. (2009). The role of the arcuate fasciculus in conduction aphasia. *Brain, 132*, 2309–2316.

Berwick, R. C., & Chomsky, N. (2016). *Why only us*. Cambridge (MA): MIT Press.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences, 17*, 89.

Blank, I., Kanwisher, N., & Fedorenko, E. (2014). A functional dissociation between language and multiple-demand systems revealed in patterns of BOLD signal fluctuations. *Journal of Neurophysiology, 112*, 1105–1118.

Bornkessel, I., Zysset, S., Friederici, A. D., Von Cramon, D. Y., & Schlesewsky, M. (2005). Who did what to whom? The neural basis of argument hierarchies during language comprehension. *NeuroImage, 26*, 221–233.

Bottini, G., Corcoran, R., Sterzi, R., Paulesu, E., Schenone, P., Scarpa, P., Frackowiak, R. S. J., & Frith, D. (1994). The role of the right hemisphere in the interpretation of figurative aspects of language a positron emission tomography activation study. *Brain, 117*, 1241–1253.

Buchsbaum, B. R., Olsen, R. K., Koch, P., & Berman, K. F. (2005). Human dorsal and ventral auditory streams subserve rehearsal-based and echoic processes during verbal working memory. *Neuron, 48*, 687–697.

Bulut, T., Hung, Y. H., Tzeng, O., & Wu, D. H. (2017). Neural correlates of processing sentences and compound words in Chinese. *PLoS One, 12*, 1–18.

Chang, C. H. C., Dehaene, S., Wu, D. H., Kuo, W. J., & Pallier, C. (2020). Cortical encoding of linguistic constituent with and without morphosyntactic cues. *Cortex, 129*, 281–295.

Chao, Y. R. (1968). *A Grammar of spoken Chinese*. Berkeley (CA): University of California Press.

Chee, M. W. L., Caplan, D., Soon, C. S., Sriram, N., Tan, E. W. L., Thiel, T., & Weekes, B. (1999). Processing of visually presented sentences in Mandarin and English studied with fMRI. *Neuron, 23*, 127–137.

Chen, F. (1998). The origin of Chinese companion quantifiers and their application principles. *Applied Linguistics, 4*, 27–32.

Chen, L., Goucha, T., Männel, C., Friederici, A. D., & Zaccarella, E. (2021). Hierarchical syntactic processing is beyond mere associating: Functional magnetic resonance imaging evidence from a novel artificial grammar. *Human Brain Mapping, 42*, 3253–3268.

Chen, L., Wu, J., Fu, Y., Kang, H., & Feng, L. (2019). Neural substrates of word category information as the basis of syntactic processing. *Human Brain Mapping, 40*, 451–464.

Chen, L., Wu, J., Hartwigsen, G., Li, Z., Wang, P., & Feng, L. (2021). The role of a critical left fronto-temporal network with its right-hemispheric homologue in syntactic learning based on word category information. *Journal of Neurolinguistics, 58*, Article 100977.

Chen, Y. C., & Yeh, S. L. (2015). Binding radicals in Chinese character recognition: Evidence from repetition blindness. *Journal of Memory and Language, 78*, 47–63.

Chomsky, N. (1995). *The minimalist program*. Cambridge (MA): MIT Press.

Chomsky, N. (2013). Problems of projection. *Lingua, 130*, 33–49.

Chomsky, N. (2015). Problems of projection: Extensions. In E. D. Domenico, C. Hamann, & S. Matteini (Eds.), *Structures, strategies and beyond: Studies in honour of Adriana belletti* (pp. 1–16). Philadelphia (PA): John Benjamins Publishing Company.

Chou, T. L., Lee, S. H., Hung, S. M., & Chen, H. C. (2012). The role of inferior frontal gyrus in processing Chinese classifiers. *Neuropsychologia, 50*, 1408–1415.

Constable, R. T., Pugh, K. R., Berroya, E., Mencl, W. E., Westerveld, M., Ni, W., & Shankweiler, D. (2004). Sentence complexity and input modality effects in sentence comprehension: An fMRI study. *NeuroImage, 22*, 11–21.

Cooke, A., Zurif, E. B., DeVita, C., Alsop, D., Koenig, P., Detre, J., Gee, J., Pinãngo, M., Balogh, J., & Grossman, M. (2002). Neural basis for sentence comprehension: Grammatical and short-term memory components. *Human Brain Mapping, 15*, 80–94.

Crozier, S., Sirigu, A., Lehéricy, S., Van De Moortele, P. F., Pillon, B., Grafman, J., Agid, Y., Dubois, B., & Lebihan, D. (1999). Distinct prefrontal activations in processing sequence at the sentence and script level: An fMRI study. *Neuropsychologia, 37*, 1469–1476.

Endress, A. D., Carden, S., Versace, E., & Hauser, M. D. (2010). The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition, 13*, 483–495.

Fedorenko, E., Behr, M. K., & Kanwisher, N. (2011). Functional specificity for high-level linguistic processing in the human brain. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 16428–16433.

Fedorenko, E., & Blank, I. A. (2020). Broca's area is not a natural kind. *Trends in Cognitive Sciences, 24*, 270–284.

Fedorenko, E., Hsieh, P. J., Nieto-Castañón, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology, 104*, 1177–1194.

Feldman, L. B., & Siok, W. W. T. (1999). Semantic radicals contribute to the visual identification of Chinese characters. *Journal of Memory and Language, 40*, 559–576.

Feng, S., Legault, J., Yang, L., Zhu, J., Shao, K., & Yang, Y. (2015). Differences in grammatical processing strategies for active and passive sentences: An fMRI study. *Journal of Neurolinguistics, 33*, 104–117.

Feng, J., Zhang, L., Chen, C., Sheng, J., Ye, Z., Feng, K., Liu, J., Cai, Y., Zhu, B., Yu, Z., & Chen, C. (2022). A cognitive neurogenetic approach to uncovering the structure of executive functions. *Nature Communications, 13*, 4588.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science, 80*(303), 377–380.

Friederici, A. D. (2011). The brain basis of language processing: From structure to function. *Physiological Reviews, 91*, 1357–1392.

Friederici, A. D. (2017). *Language in our brain: The origins of a uniquely human capacity*. Cambridge (MA): MIT Press.

Friederici, A. D., Bahlmann, J., Heim, S., Schubotz, R. I., & Anwander, A. (2006). The brain differentiates human and non-human grammars: Functional localization and structural connectivity. *Proceedings of the National Academy of Sciences of the United States of America, 103*, 2458–2463.

Friederici, A. D., Gunter, T. C., Hahne, A., & Mauth, K. (2004). The relative timing of syntactic and semantic processes in sentence comprehension. *NeuroReport, 15*, 165–169.

Friederici, A. D., & Kotz, S. A. (2003). The brain basis of syntactic processes: Functional imaging and lesion studies. *NeuroImage, 20*, 8–17.

Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *NeuroReport, 20*, 563–568.

Friederici, A. D., Pfeifer, E., & Hahne, A. (1993). Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research, 1*, 183–192.

Friederici, A. D., & Weissenborn, J. (2007). Mapping sentence form onto meaning: The syntax-semantic interface. *Brain Research, 1146*, 50–58.

Fujita, K. (2014). Recursive merge and human language evolution. In T. Roeper, & M. Speas (Eds.), *Recursion: Complexity in cognition. Studies in theoretical psycholinguistics* (pp. 243–264). Cham (CH): Springer.

Gates, K. M., & Molenaar, P. C. M. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage, 63*, 310–319.

Gates, K. M., Molenaar, P. C. M., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage, 54*, 1151–1158.

Goucha, T., & Friederici, A. D. (2015). The language skeleton after dissecting meaning: A functional segregation within Broca's area. *NeuroImage, 114*, 294–302.

Goucha, T., Zaccarella, E., & Friederici, A. D. (2017). A revival of Homo loquens as a builder of labeled structures: Neurocognitive considerations. *Neuroscience & Biobehavioral Reviews, 81*, 213–224.

Gunter, T. C., Friederici, A. D., & Hahne, A. (1999). Brain responses during sentence reading: Visual input affects central processes. *NeuroReport, 10*, 3175–3178.

Guo, X. Z. (1987). *Modern Chinese measure words dictionary (现代汉语量词手册)*. Beijing (CHN): China Peace Press.

Guo, R. (2018). *Modern Chinese parts of speech: Classification theory*. London (UK): Routledge.

Hagoort, P. (2005). On Broca, brain, and binding: A new framework. *Trends in Cognitive Sciences, 9*, 416–423.

Hagoort, P. (2013). MUC (memory, unification, control) and beyond. *Frontiers in Psychology, 4*, 1–13.

Hagoort, P., & Indefrey, P. (2014). The neurobiology of language beyond single words. *Annual Review of Neuroscience, 37*, 347–362.

Hahne, A., & Friederici, A. D. (2002). Differential task effects on semantic and syntactic processes as revealed by ERPs. *Cognitive Brain Research, 13*, 339–356.

Hahne, A., & Jescheniak, J. D. (2001). What's left if the jabberwock gets the semantics? An ERP investigation into semantic and syntactic processes during auditory sentence comprehension. *Cognitive Brain Research, 11*, 199–212.

Hauser, M. D., Chomsky, N., & Fitch, W. T. (2002). Neuroscience: The faculty of language: What is it, who has it, and how did it evolve? *Science, 298*, 1569–1579.

Hickok, G., Buchsbaum, B., Humphries, C., & Muftuler, T. (2003). Auditory-motor interaction revealed by fMRI: Speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience, 15*, 673–682.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience, 8*, 393–402.

Hoshi, K. (2018). Merge and labeling as descent with modification of categorization: A neo-lennebergian approach. *Biolinguistics, 12*, 39–54.

Hoshi, K. (2019). More on the relations among categorization, merge and labeling, and their nature. *Biolinguistics, 13*, 1–21.

Huang, B. R., & Liao, X. D. (2011). *Modern Chinese (updated 5th editions) (现代汉语)*. Beijing (CHN): Higher Education Press.

Huettel, S. A., Song, A. W., & McCarthy, G. (2004). *Functional magnetic resonance imaging* (2nd ed.). Sunderland (UK): Sinauer Associates.

Isel, F., Hahne, A., Maess, B., & Friederici, A. D. (2007). Neurodynamics of sentence interpretation: ERP evidence from French. *Biological Psychology, 74*, 337–346.

Jin, L. X. (2002). The aspectual significance of the suffix le and its syntactic conditions. *Chinese Teach World, 59*, 34–43.

Kaestner, E., Thesen, T., Devinsky, O., Doyle, W., Carlson, C., & Halgren, E. (2021). An intracranial electrophysiology study of visual language encoding: The contribution of the precentral gyrus to silent reading. *Journal of Cognitive Neuroscience, 33*, 2197–2214.

Kim, J., Zhu, W., Chang, L., Bentler, P. M., & Ernst, T. (2007). Unified structural equation modeling approach for the analysis of multisubject, multivariate functional MRI data. *Human Brain Mapping, 28*, 85–93.

Lee, D., & Newman, S. D. (2010). The effect of presentation paradigm on syntactic Processing: An event-related fMRI study. *Human Brain Mapping, 31*, 65–79.

Li, L. D. (2021). *A study on the grammatical meaning of the dynamic auxiliary word "zhe" and the syntactic conditions for its realization, 297*, 99–105.

Lindenberg, R., & Scheef, L. (2007). Supramodal language comprehension: Role of the left temporal lobe for listening and reading. *Neuropsychologia, 45*, 2407–2415.

Luke, K. K., Liu, H. L., Wai, Y. Y., Wan, Y. L., & Tan, L. H. (2002). Functional anatomy of syntactic and semantic processing in language comprehension. *Human Brain Mapping, 16*, 133–145.

Ma, A. M. (2015). *The origin and development of individual quantifiers in Chinese (汉语个体量词的产生与发展)*. Beijing (CHN): China Social Sciences Publishing House.

Makuuchi, M., Bahlmann, J., Anwander, A., & Friederici, A. D. (2009). Segregating the core computational faculty of human language from working memory. *Proceedings of the National Academy of Sciences of the United States of America, 106*, 8362–8367.

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffmann, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience, 25*, 1014–1019.

Maran, M., Friederici, A. D., & Zaccarella, E. (2022). Syntax through the looking glass: A review on two-word linguistic processing across behavioral, neuroimaging and neurostimulation studies. *Neuroscience & Biobehavioral Reviews, 142*, Article 104881.

Matchin, W., Hammerly, C., & Lau, E. (2017). The role of the IFG and pSTS in syntactic prediction: Evidence from a parametric study of hierarchical structure in fMRI. *Cortex, 88*, 106–123.

Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex, 30*, 1481–1498.

Miyagawa, S., Berwick, R. C., & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology, 4*, 1–6.

Ohta, S., Fukui, N., & Sakai, K. L. (2013). Syntactic computation in the human brain: The degree of merger as a key factor. *PLoS One, 8*, Article e56230.

Oldfield, R. C. (1971). The assessment and analysis of handedness: The Edinburgh inventory. *Neuropsychologia, 9*, 97–113.

Opitz, B., & Friederici, A. D. (2003). Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage, 19*, 1730–1737.

Opitz, B., & Friederici, A. D. (2004). Brain correlates of language learning: The neuronal dissociation of rule-based versus similarity-based learning. *Journal of Neuroscience, 24*, 8436–8440.

Opitz, B., & Friederici, A. D. (2007). Neural basis of processing sequential and hierarchical syntactic structures. *Human Brain Mapping, 28*, 585–592.

den Ouden, D. B., Saur, D., Mader, W., Schelter, B., Lukic, S., Wali, E., Timmer, J., & Thompson, C. K. (2012). Network modulation during complex syntactic processing. *NeuroImage, 59*, 815–823.

Pallier, C., Devauchellea, A. D., & Dehaenea, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 2522–2527.

Peng, D. L. (1997). *Cognitive research on Chinese language*. Jinan (CHN): Shandong Education Press.

Price, C. J. (2012). A review and synthesis of the first 20 years of PET and fMRI studies of heard speech, spoken language and reading. *NeuroImage, 62*, 816–847.

Price, C. J., Wise, R. J. S., Watson, J. D. G., Patterson, K., Howard, D., & Frackowiak, R. S. J. (1994). Brain activity during reading the effects of exposure duration and task. *Brain, 117*, 1255–1269.

Pylkkänen, L. (2019). The neural basis of combinatory syntax and semantics. *Science, 366*, 62–66.

Santi, A., & Grodzinsky, Y. (2010). fMRI adaptation dissociates syntactic complexity dimensions. *NeuroImage, 51*, 1285–1293.

Schell, M., Zaccarella, E., & Friederici, A. D. (2017). Differential cortical contribution of syntax and semantics: An fMRI study on two-word phrasal processing. *Cortex, 96*, 105–120.

Segaert, K., Kempen, G., Petersson, K. M., & Hagoort, P. (2013). Syntactic priming and the lexical boost effect during sentence production and sentence comprehension: An fMRI study. *Brain and Language, 124*, 174–183.

Segaert, K., Menenti, L., Weber, K., Petersson, K. M., & Hagoort, P. (2012). Shared syntax in language production and language comprehension-An fMRI study. *Cerebral Cortex, 22*, 1662–1670.

Shao, J. M., & Zhao, C. L. (2006). A theoretical thought on semantic categories. *Chinese Teach World, 75*, 29–40.

Smith, E. E., & Jonides, J. (1997). Working memory: A view from neuroimaging. *Cognitive Psychology, 33*, 5–42.

Snijders, T. M., Vosse, T., Kempen, G., Van Berkum, J. J. A., Petersson, K. M., & Hagoort, P. (2009). Retrieval and unification of syntactic structure in sentence comprehension: An fMRI study using word-category ambiguity. *Cerebral Cortex, 19*, 1493–1503.

Sonnweber, R., Ravignani, A., & Fitch, W. T. (2015). Non-adjacent visual dependency learning in chimpanzees. *Animal Cognition, 18*, 733–745.

Sun, Z., Shi, Y., Guo, P., Yang, Y., & Zhu, Z. (2021). Independent syntactic representation identified in left front-temporal cortex during Chinese sentence comprehension. *Brain and Language, 214*, Article 104907.

Tang, S. W. (2019). *Formal Chinese syntax* (2nd ed.). Shanghai (CHN: Shanghai Educational Publishing House. 形式汉语句法学.

Tyler, L. K., Shafto, M. A., Randall, B., Wright, P., Marslen-Wilson, W. D., & Stamatakis, E. A. (2010). Preserving syntactic processing across the adult life span: The modulation of the frontotemporal language system in the context of age-related atrophy. *Cerebral Cortex, 20*, 352–364.

Uddén, J., Ingvar, M., Hagoort, P., & Petersson, K. M. (2017). Broca's region: A causal role in implicit processing of grammars with crossed non-adjacent dependencies. *Cognition, 164*, 188–198.

Uddin, L. Q., Yeo, B. T., & Spreng, R. N. (2019). Towards a universal taxonomy of macro-scale functional human brain networks. *Brain Topography, 32*, 926–942.

Vallar, G., & Papagno, C. (2002). Neuropsychological impairments of verbal short-term memory. In A. D. Baddeley, M. D. Kopelman, & B. A. Wilson (Eds.), *The handbook of memory disorders* (2nd ed., pp. 249–271). England (UK): John Wiley & Sons.

Wang, N. (2000). System theory and the founding of the science of Chinese Characters formation. *Journal of Jinan University (philosophy & social science edition), 22*, 15–21.

Wang, P., Knösche, T. R., Chen, L., Brauer, J., Friederici, A. D., & Maess, B. (2021). Functional brain plasticity during L1 training on complex sentences: Changes in gamma-band oscillatory activity. *Human Brain Mapping, 42*, 3858–3870.

Wang, S., Mo, D., Xiang, M., Xu, R., & Chen, H. C. (2013). The time course of semantic and syntactic processing in reading Chinese: Evidence from ERPs. *Language & Cognitive Processes, 28*, 577–596.

Wang, F., Ouyang, G., Zhou, C., & Wang, S. (2015). Re-examination of Chinese semantic processing and syntactic processing: Evidence from conventional ERPs and reconstructed ERPs by residue iteration decomposition (RIDE). *PLoS One, 10*, 1–16.

Wang, J., & Zhang, J. (2016). The effects of category consistency and neighborhood size of the semantic radical on the semantic processing of Chinese character. *Acta Psychology Sinica, 48*, 1390–1400.

Wang, S., Zhu, Z., Zhang, J. X., Wang, Z., Xiao, Z., Xiang, H., & Chen, H. C. (2008). Broca's area plays a role in syntactic processing during Chinese reading comprehension. *Neuropsychologia, 46*, 1371–1378.

Weber, K., & Indefrey, P. (2009). Syntactic priming in German-English bilinguals during sentence comprehension. *NeuroImage, 46*, 1164–1172.

Wu, J., Yang, J., Chen, M., Li, S., Zhang, Z., Kang, C., Ding, G., & Guo, T. (2019). Brain network reconfiguration for language and domain-general cognitive control in bilinguals. *NeuroImage, 199*, 454–465.

Wu, C. Y., Zaccarella, E., & Friederici, A. D. (2018). Universal neural basis of structure building evidenced by network modulations emerging from Broca's area: The case of Chinese. *Human Brain Mapping, 40*, 1705–1717.

Xun, E. D., Rao, G. Q., Xiao, X. Y., & Zang, J. J. (2016). The construction of the BCC Corpus in the age of big data. *Corpus Linguistics, 1*, 93–109.

Yan, C. G., Cameron Craddock, R., He, Y., & Milham, M. P. (2013). Addressing head motion dependencies for small-world topologies in functional connectomics. *Frontiers in Human Neuroscience, 7*, 1–19.

Yang, J., Gates, K. M., Molenaar, P., & Li, P. (2015). Neural changes underlying successful second language word learning: An fMRI study. *Journal of Neurolinguistics, 33*, 29–49.

Yan, C. G., Wang, X Di, Zuo, X. N., & Zang, Y. F. (2016). DPABI: Data processing & analysis for (Resting-State) brain imaging. *Neuroinformatics, 14*, 339–351.

Yeh, S. L., Chou, W. L., & Ho, P. (2017). Lexical processing of Chinese sub-character components: Semantic activation of phonetic radicals as revealed by the Stroop effect. *Scientific Reports, 7*, 1–12.

Yu, J., & Zhang, Y. (2008). When Chinese semantics meets failed syntax. *NeuroReport, 19*, 745–749.

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A sub-region based functional investigation in the left pars opercularis. *Frontiers in Psychology, 6*, 1–9.

Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2017). Building by syntax: The neural basis of minimal linguistic structures. *Cerebral Cortex, 27*, 411–421.

Zaccarella, E., Schell, M., & Friederici, A. D. (2017). Reviewing the functional basis of the syntactic merge mechanism for language: A coordinate-based activation likelihood estimation meta-analysis. *Neuroscience & Biobehavioral Reviews, 80*, 646–656.

Zhang, H., Su, I., Chen, F., Ng, M. L., Wang, L., & Yan, N. (2020). The time course of orthographic and semantic activation in Chinese character recognition: Evidence from an ERP study. *Lang Cogn Neurosci, 35*, 292–309.

Zhang, Y., Yu, J., & Boland, J. E. (2010). Semantics does not need a processing license from syntax in reading Chinese. *Journal of Experimental Psychology Learning Memory and Cognition, 36*, 765–781.

Zhou, X., Tang, Y., Weng, X., Ma, L., & Li, D. (2001). Brain activation in reading regular and irregular Chinese characters. *NeuroImage, 13*, 634.

Zhu, D. X. (1982). *Lectures on grammar (语法讲义)*. Beijing (CHN): The Commercial Press.

Zhu, D. X. (1985). *The questions and answers on grammar (语法答问)*. Beijing (CHN): The Commercial Press.

Zhu, Y., Xu, M., Lu, J., Hu, J., Kwok, V. P. Y., Zhou, Y., Yuan, D., Wu, B., Zhang, J., Wu, J., & Tan, L. H. (2022). Distinct spatiotemporal patterns of syntactic and semantic processing in human inferior frontal gyrus. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-022-01334-6