WILEY

RESEARCH ARTICLE

Neural substrates of word category information as the basis of syntactic processing

Luyao Chen[1,2] | Junjie Wu[3] | Yongben Fu[3] | Huntae Kang[1] | Liping Feng[1] 🆔

[1]College of Chinese Language and Culture, Beijing Normal University, Beijing, China

[2]Department of Neuropsychology, Max Plank Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[3]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China

Correspondence
Liping Feng, College of Chinese Language and Culture, Beijing Normal University, Beijing 100875, P. R. China.
Email: flpjoy@163.com

Funding information
The National Social Science Fund of China, Grant/Award Number: 14BYY146

Abstract

The ability to use word category information (WCI) for syntactic structure building has been hypothesized to be the essence of human language faculty. The neural substrate of the ability of using the WCI for the complex syntactic hierarchical structure processing, however, is yet unknown. Therefore, we directly conducted an fMRI experiment by using a pseudo-Chinese artificial language with syntactic structures containing a center-embedded relative clause. Thirty non-Chinese native (Korean) speakers were randomly divided into two groups: one acquired WCI and WCI-based syntactic rules (the WCI group) before the scanning session, and the other did not (the non-WCI group). Both groups were required to judge the grammaticality of the testing sentences, with critical long-distance dependencies between two elements (the main verb and the relativizer). Behaviorally, the WCI group's accuracy was significantly higher and its reaction time was shorter. The scanning results showed that the left superior temporal gyrus (STG) and Broca's area were more strongly activated for the WCI group, and the dynamic causal modeling analyses revealed a distinct effective connectivity pattern for this group. Therefore, the present research, for the first time, reveals that the activation of and the functional connectivity between Broca's area and the left STG play a critical role in the ability of the rule-based use of the WCI which is crucial for complex hierarchical structure building, and might be substantially corresponding to the "labeling competence" within the linguistic framework.

KEYWORDS

artificial grammar, complex hierarchical syntactic structure, dynamic causal modeling, fMRI, labeling competence, rule-based, word category information

1 | INTRODUCTION

The essence of the faculty of human language has been studied for decades. Hauser, Chomsky, and Fitch (2002) proposed that recursion might play the core role in human language generation, whereas it is still unclear which specific recursive operation is critical for the human language faculty. Recently, some researchers have been focusing on "labeling", which can identify the syntactic word categories of the syntactic components and is essential for the syntactic hierarchy construction (Goucha, Zaccarella, & Friederici, 2017; Murphy, 2015). For

instance, "the apple" is labeled as a determiner phrase (DP) because of the Head, "the" (D), determining the hierarchy of this phrase, and then "the apple" can be combined with "eat" to be labeled as a verb phrase (VP). According to Chomsky (2013, 2015), labeling only functions after "Merge" at the interface for interpretation, where Merge refers to the binary combination of two syntactic objects so as to form a bigger component. But without labeling, Merge would merely string the words up without constructing the syntactic hierarchies (Goucha et al., 2017). Thus, Murphy (2015) raised the "Labeling Hypothesis", mainly suggesting that the labeling competence might be unique to humans and human language. To note, a "label" is only a theoretical item, while the substance underlying labels is the word category information (WCI), that is, as for the labels of the words or phrases, D, N (noun), and DP, all labels are based on the WCI. Narita (2014) also demonstrated that the labeling of a new phrase is closely related to

the syntactic categories of the constituents. Therefore, theoretically, using the WCI should be the basis for syntactic processing, as Friederici (2017, p. 28) proposed that the WCI "guides the buildup of syntactic structures." Goucha et al. (2017) further pinpointed that labeling is sensitive to the syntactic properties of both the inputs (i.e., the word category) and the structure that constitutes the output (i.e., the label of the new phrase), and allows people to recognize the labels that the single elements take and the combinatorial rules applying among them in language use.

However, the labeling competence is proposed mainly within the framework of linguistics. The neural substrate of dynamically using the WCI to process the syntactic structures still needs to be explored. Moreover, the cognitive mechanism underlying using the WCI to learn the complex syntactic rules and to process the complex hierarchical syntactic structures is necessary to be specified, which could either be the rule-based learning/processing mechanism concerning the rules that are compositional, generative, and rigorously descriptive with the property of inflexibility (Opitz & Hofmann, 2015), helping the participants to characterize the stimulus by examining part of the features (Pothos, 2007), or the other non-rule-based mechanisms focusing on the statistics (esp. the transitional probability) (the statistical-learning mechanism) (Saffran, Aslin, & Newport, 1996), the similarity to the exemplars of the learning session (the similarity-based learning mechanism) (Opitz & Friederici, 2003, 2004; Shanks, 1995) and the other possible non-rule information. To clarify, as the syntactic rule learning is also accompanied with the processing of the testing sentences, we would not differentiate the cognitive learning and the cognitive processing mechanisms strictly in this study. Therefore, our present fMRI research mainly aims at purely studying the neural substrates of the ability of using the WCI for the complex hierarchical syntactic structure processing, and our experiment results could be critical for identifying the corresponding cognitive mechanism so as to map the linguistic construct, the labeling competence, onto the cognitive one(s) in substance.

It is relatively rare to find direct experimental research focusing on the WCI as the basis of syntactic processing is rare. Behaviorally, the results of Rohrmeier, Fu, and Dienes's (2012) experiment first showed that participants could implicitly learn the syntactic rules for the construction of the recursive complex hierarchical structures with the WCI represented unconsciously, highlighting that the WCI is necessary for both syntactic rule learning and syntactic structure processing, and that cognitively, the learning (including the processing of the target structures) facilitated by the WCI is rule based. Our recent behavioral study (Chen et al., 2018) found that the participants who had acquired the WCI before the 6 blocks' pseudo-Chinese artificial grammar learning session could successfully construct the key complex syntactic rules, also indicating that WCI is essential for complex syntactic rule construction. We will discuss the cognitive mechanism of such a WCI-based learning by mainly proposing that the final successful acquisition of the complex syntactic rules on the basis of the rule-based use of the WCI (see the Discussion section).

As to the question of the neural substrates of using the WCI for syntactic processing, including the brain areas and the functional connections among them, previous research does provide some insights. By utilizing the modified version of BROCANTO (Friederici, Steinhauer, & Pfeifer, 2002), a minimal artificial language in which the

artificial words have their own corresponding word categories, Opitz and Friederici (2007) found that compared with the local dependent structure processing, the long-distance dependent rule-based hierarchical syntactic structure processing would more strongly activate the Broca's area (esp. Brodmann's area [BA] 44) in the left hemisphere. Due to the fact that BROCANTO has its own word categories, this study indicated that Broca's area might support the rule-based learning in which the WCI is critical for the complex syntactic rule construction, and this interpretation also applies to Opitz and Friederici (2003, 2004). Zaccarella and Friederici (2015) further reported that when participants were processing the minimal "Merge-able" pairs (e.g., DIESE FLIRK, in which the determiner was a German word, and the latter one a pseudo-word), the anterior ventral cluster of BA 44 was significantly activated for the Merge-able pairs. They claimed that Merge had two stages, a concatenation stage of accumulating the words, which depended on the frontal operculum, and a labeling stage helping to build up the hierarchical syntactic structures, supported by the anterior ventral part of BA 44. Goucha and Friederici (2015) also evidenced that when a sentence only reserves function words and inflectional affixes which still reflect the word categories, the left BA 44 would be highly activated. Although the underlying cognitive mechanisms were not demonstrated, both Zaccarella and Friederici (2015) and Goucha and Friederici (2015) have revealed the candidate brain areas for individuals' successful retrieving of the combinatorial rules operating on labels (also see the comments of Goucha et al., 2017), and these areas might be the candidate neural substrates of the WCI as the basis of the syntactic processing. Furthermore, Goucha et al. (2017) proposed that the Broca's area, the left superior temporal gyrus (STG), and the dorsal pathway connecting them might be in support of labeling. Such a dorsal pathway connecting Broca's area and the left temporal cortex (esp, the STG) is reported to be critical for (complex) syntactic processing (for a recent review, see Friederici, Chomsky, Berwick, Moro, & Bolhuis, 2017) in ontogenetical (Brauer, Anwander, Perani, & Friederici, 2013; Skeide, Brauer, & Friederici, 2016), second language acquisition (Yamamoto & Sakai, 2017), cross species (Dick, Bernal, & Tremblay, 2014), and primary progressive aphasia studies (Wilson et al., 2011), while the ventral tract connecting these brain regions might be involved in the semantic processing (Friederici et al., 2017) and local syntactic processing (Makuuchi & Friederici, 2013) or low-level syntactic feature identification (Friederici, 2012; Skeide et al., 2016). These researches mentioned above imply that the WCI-based complex hierarchical syntactic structure processing might be supported by the Broca's area, the left STG and their dorsal connections.

Nevertheless, these experimental researches did not study the neural substrates of the WCI as the basis of syntactic processing in a comparatively direct way. Moreover, the Indo-European languages have rich morphological clues for facilitating speakers' or readers' use of WCI in the process of language cognition, while during syntactic structure processing in other languages, such as Chinese, the related neural substrates still need to be investigated. Therefore, we, for the first time, designed a micro, visually presented, pseudo-Chinese artificial language, inspired by Opitz and Friederici (2007) and Rohrmeier et al. (2012), based on the Chinese phonetic alphabet in which the letters are pseudo-Chinese characters (every single letter would be called a word in this artificial language) that were unfamiliar to the participants and could not be pronounced by them, and in which the

word order follows the Chinese syntax. This artificial language is composed of a complex hierarchical syntactic structure with a relative clause center embedded, that is, the main schematic syntactic expression is (N $V_{1/2}$ [[V_3 N] $R_{1/2}$] N), and the relative clause, ([V_3 N] $R_{1/2}$), is center embedded in the main clause (N $V_{1/2}$ N). During complex hierarchical syntactic structure processing, participants were required to form the critical long-distance dependencies between the subcategories of the verb ($V_{1/2}$) in the main clause and the subcategories of the relativizer ($R_{1/2}$) conducting the clause, so as to label a new complex VP containing the main verb, $V_{1/2}$, and the clause labeled by $R_{1/2}$, in which (V_3 N) is also center embedded between $V_{1/2}$ and $R_{1/2}$. In addition, although BROCANTO involves phonological cues for the classification of words (e.g., the word containing the vowel, /e/, such as "prez", should be a verb), such cues would be avoided in this experiment.

The present fMRI research is based on the results of our previous behavioral study (Chen et al., 2018) which focused on the WCI for the syntactic rule learning (see Supporting Information for the brief introduction). In the current experiment, one group of participants had already acquired both the WCI and the WCI-based syntactic rules of the artificial language (the WCI group) before the fMRI scanning session, whereas the other group had not (the non-WCI group) (see the Participants section for details). Both groups participated in the same task by judging the grammaticality of the testing materials, during which the WCI either functioned or the WCI group's performance (including the brain activation pattern) would be similar to the non-WCI group's. Apart from exploring the brain activation pattern, the dynamic causal modeling (DCM; Friston, Harrison, & Penny, 2003), commended to be "the most appropriate" approach for explaining the brain responses resulting from the experimental interventions (Stephan et al., 2010), was also performed to study the effective connections among the brain areas. DCM compares different hypothesized model structures through the Bayesian model selection (BMS) and finally picks up the optimal one considering the accuracy (model fit), complexity, and generalizability synthetically (Makuuchi & Friederici, 2013; Stephan et al., 2010). We expected that the brain activation pattern of the WCI group might be different from the non-WCI group in involving the core language areas such as Broca's area and the left temporal lobe (Fedorenko & Thompson-Schill, 2014), and the effective connectivity among them would support the ability of the rule-based use of the WCI to process the complex hierarchical syntactic structures.

2 | METHODS

2.1 | Participants

As the artificial language in the present study is based on Chinese, it is necessary to recruit non-Chinese native speakers to avoid mother tongue transfer (cf. Ling, 2016). Thirty native Korean speakers had participated in our previous behavioral study (Chen et al., 2018) (see Supporting Information for the introduction of the artificial grammar learning procedure and results), and all of them participated in this experiment, belonging to the WCI group (N = 15, 4 males; age:

M = 23.3, SD = 3.5) and the non-WCI group (N = 15, 4 males; age: M = 23.5, SD = 2.9). To note, as our previous behavioral study showed, at the end of the 6-block syntactic rule learning, the accuracy (ACC) of the WCI group in the last block (Block 6) (M = 82.4%, SD = 11.1%) was significantly higher than both the random level (t(14) = 11.30, p < .001) and that of the non-WCI group (M = 50.6%, SD = 9.2%) (t(28) = 8.57, p < .001), while the ACC of the non-WCI group in Block 6 was similar to the random level (t(14) = .24, p = .818), and the non-WCI group's word classification performance (M = 37.8%, SD = 16.4%) was significantly below 50% (t(14) = −2.88, p < .05). These results indicated that the WCI group could successfully acquire the key rules based on WCI, but the non-WCI group could not. Therefore, we hereby introduced the learning results to confirm that before the scanning session of our present fMRI research, the grouping of the WCI and the non-WCI groups is reliable.

All participants are right handed with normal or corrected-to-normal vision, reporting no physical or neuropsychological disease, and signed the research consent form. After the experiment, they were paid for participating. This experiment is approved by the Institutional Review Board of Beijing Normal University Imaging Center for Brain Research, National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, China. After the scanning session, the data from two participants of the WCI group and one participant of the non-WCI group were discarded due to excessive motion artifacts (>3 mm in translation and >3° in rotation), resulting in 13 participants in the WCI group (4 males; age: M = 23.5, SD = 3.6) and 14 participants in the non-WCI group (4 males; age: M = 23.6 SD = 2.9).

A prior operation span task (O-span, adapted from Unsworth, Heitz, Schrock, & Engle, 2005) was conducted before the experiment to balance the working memory span of the two groups. Participants were required to judge the correctness of equations such as "(7−5) × 3 = 6?," and to memorize English letters (each appeared after one equation) in order. The number of the letters increased as the equations progressed, and the ACC of the memorization of the letter sequences was recorded. Thus, the O-span performance of the remaining participants showed no significant difference between the two groups (the WCI group's ACC: M = 60.85%, SD = 6.39%; the non-WCI group's ACC: M = 61.64%, SD = 7.40%; t(25) = −.30, p = .768).

2.2 | Materials

The key syntactic rules are to merge $V_{1/2}$ and $R_{1/2}$ in a certain way. To balance the potential coindex effect (i.e., V_1 – R_1, and V_2 – R_2) facilitating the processing, half of the participants of the WCI group should use the critical converse rules: V_1 – R_2, and V_2 – R_1. We also added a verb modifier (Mod) which could adjacently occur before any verbs so as to vary the lengths of the long-distance dependencies to constrain the reading strategy, and to enrich the experimental material. Every word category and their corresponding words are listed as follows: (a) V_1: ㄊ, ㄷ, ㄋ; (b) V_2: ㄙ, ㄇ, ㄎ; (c) V_3: ㄩ, ㄐ, ㄑ; (d) N: ㄨ, ㄅ, ㄊ, ㄢ; (e) R_1: ㄌ; (f) R_2: ㄍ; and (g) Mod: 巛.

Given the blocked design adopted by our fMRI experiment, 64 testing sentences, each ending with a Chinese period "。", were generated with half being grammatically correct, and they were

randomly divided into eight task blocks, each having four correct sentences and four ungrammatical sentences. If the rules, $V_1 - R_1$ and $V_2 - R_2$, are grammatical, then $V_1 - R_2$ and $V_2 - R_1$ are ungrammatical, and vice versa. For instance, if $V_1 - R_1$ and $V_2 - R_2$ are grammatical, a correct example sentence can be "ㄨ ㄊ ㄍ ㄩ ㄤ ㄅ ㄣ。" whose correctness will be violated if the relativizer becomes "ㄌ." The lengths (6–7 words per sentence), word positions (esp. the Mod), and the word frequencies of the testing sentences (e.g., every token of N appears equivalently in every position of the sentences from each condition) were carefully controlled so as to avoid performance bias. Four task blocks constituted a run, and in each run, there was a baseline block between every two task blocks. The materials in the baseline blocks were strings of the symbol "#" also with a Chinese period "。" at the end, and their lengths matched the lengths of the testing sentences. Before the actual scanning, participants attended the practice, in which the practice materials were composed in a similar way to the test materials, with two task blocks and one baseline block between them. The practice sentences were different from the test ones.

2.3 | Procedure

Participants had practiced before the scanning session. The procedure of the scanning session had two runs, each having four task blocks, and between every two task blocks there was a baseline block (three in total for each run) (see Figure 1).

In the task blocks, every testing sentence was presented word by word to avoid participant reading strategies such as only paying attention to a certain word in a certain position of the sentence without reading the whole sentence, although these factors had been controlled for. The existence of Mod would change the lengths of the sentences, but the whole presentation time for a testing sentence (or a trial) was fixed. Taking a test trial for instance, before the first word of the sentence occurred, there was a 1,000 ms blank and a 300 ms fixation followed by a 200 ms blank. Every word of the sentence before the last word would be presented for 500 ms, but the duration of the blank after every word varied, being 400 ms for the 6-word sentences and 300 ms for the 7-word sentences. The last word with the period was for detecting the response, lasting for

3,000 ms for the 6-word sentences and 2,700 ms for the 7-word sentences, and the presentation of this screen would not terminate if the reaction time (RT) was shorter than the duration. Therefore, the duration of every trial was set to be 9 s. The presentation of the "#" strings was the same.

During the scanning, participants were required to judge the grammatical correctness of the test sentences by pressing the buttons once the last word of the testing sentence with the Chinese period appeared. The corresponding relations between the fingers and the buttons were also balanced.

2.4 | Behavioral data acquisition and analyses

The accuracy (ACC) and the RT of the participants' grammaticality judgment were recorded. A two-sample t test was performed to compare the ACCs of the two groups, and one-sample t tests were also conducted to compare the ACC of each group with the random level (50%), in which the p value was Bonferroni corrected ($p < .017$). As for RT, a two-sample t test was also adopted to assess the RT difference between the two groups. Besides, the comparison between the grammatically correct and the incorrect condition were also statistically analyzed for each group (see Supporting Information).

2.5 | fMRI data acquisition

The fMRI data were acquired by a 3T SiemensMAGNETON Trio system at the MRI Center of Beijing Normal University, using a single shot T2*-weighted gradient echo planar imaging (EPI) sequence at 252 time points for each run (two runs in total), including repetition time (TR) = 2,000 ms; echo time (TE) = 30 ms; flip angle (FA) = 90°; field of view (FOV) = 192 × 192 mm²; matrix size = 64 × 64; resolution within slices = 3 × 3 mm²; slice thickness/gap = 3 mm/1 mm; slice number = 33. Before the fMRI scanning, the high-resolution anatomical images were acquired using a T1-weighted multiplanar reconstruction sequence (TR = 2,530 ms; TE = 3.39 ms; FA = 7°; FOV = 256 × 256 mm²; matrix size = 256 × 256; resolution within slices = 1.33 × 1 mm²; slice thickness/gap = 1.33 mm/1 mm; slice number = 144).

(a)

Task

	+		ㄊ		ㄇ		ㄅ		ㄢ		ㄌ		ㄨ。		
1000	300	200	500	400	500	400	500	400	500	400	500	400	3000		
	+		ㄨ		ㄊ		ㄍ		ㄩ		ㄊ		ㄅ		ㄣ。
1000	300	200	500	300	500	300	500	300	500	300	500	300	500	300	2700

Baseline

	+		#		#		#		#		#		#。		
1000	300	200	500	400	500	400	500	400	500	400	500	400	3000		
	+		#		#		#		#		#		#		#。
1000	300	200	500	300	500	300	500	300	500	300	500	300	500	300	2700

t (ms)

(b)

| Task | Baseline | Task | Baseline | Task | Baseline | Task |

8 trials   8 trials

FIGURE 1 The scanning procedure. (a) The procedure of presenting the task and the baseline trials. (b) The block organization of one run

2.6 | fMRI data analyses

Image preprocessing was performed using the DPARSF (Yan & Zang, 2010) and SPM8 (Wellcome Department of Cognitive Neurology, London, UK) implemented in MATLAB R2012a. The following steps were done in preprocessing: (a) slice timing correction was performed to minimize differences in acquisition between slices; (b) the images were realigned to the mid volume in the time series to correct for head motion; (c) then, the anatomical images were coregistered to the functional images for each participant; (d) the images were normalized to EPI template based on the Montreal Neurological Institue (MNI) stereotactic space to minimize cerebral differences between participants, and resampled into $3 \times 3 \times 3$ m^3 voxels; and (e) the images were smoothed with cubic Gaussian filter (6 mm full width at half-maximum).

Statistical analyses were performed using SPM8 in MATLAB R2012a. For each participant (first level), a general linear model was used to estimate effects of experiment condition at the voxel-based level, with a reference boxcar function of stimuli, which was convolved with a canonical hemodynamic response function. The data were high-pass filtered at 128 Hz. Regions of interest (ROIs) analyses were performed to directly investigate the candidate brain areas for the WCI as the basis of syntactic processing. Given that the Broca's area (BA 45 and BA 44) and the left STG are involved in the (complex) syntactic structure processing (Friederici et al., 2017; Skeide & Friederici, 2016), and the left fusiform gyrus (FG) supports the visual prelexical representation as a visual word form area (Dehaene, Le, Poline, Le, & Cohen, 2002; Jobard, Crivello, & Tzourio-Mazoyer, 2003), we defined the left BA 45, BA 44, and STG as the ROIs under the contrast between the WCI group and the non-WCI group, hypothesizing that these brain areas would support the WCI-based complex hierarchical syntactic structure processing, whereas the left FG was hypothesized to be a common area for both groups to process the word form. Therefore, the left FG was defined under the conjunction of the brain activation of the two groups as a control ROI. In case of the selection bias, the anatomical ROIs of the left BA 45 (Amunts et al., 1999), BA 44 (Amunts et al., 1999), STG (composed of the TE 1.0-1.2 from Morosan et al., 2001, and the TE 3 from Morosan, Rademacher, Palomero-Gallagher, & Zilles, 2005), and the left FG (Caspers et al., 2013) were defined on the basis of the maximum probabilistic cytoarchitectonic maps from the SPM Anatomy Toolbox 2.2b (Eickhoff et al., 2005; also see the comment of Stephan et al., 2010), and they were utilized to mask the results of the whole-brain analyses for extracting the peak MNI coordinates centered on which spherical ROIs (radius = 4 mm) were defined (see Figure 3 for the positions of the ROIs in the left hemisphere): the left BA 45 (−54, 27, 3), the left BA 44 (−51, 3, 12), the left STG (−54, −3, 3), and the left FG (−45, −66, −12). To investigate the neural activation evoked by the task, we extracted the β values of the ROIs as the mean activation on the individual level. Then, two-sample t test was performed on each ROI to investigate the activation differences between the WCI group and the non-WCI group, and the threshold was Bonferroni corrected to be $p < .0125$. The methods and the results of the whole-brain analyses were less interested in the current study, and readers may find this part in the Supporting Information.

2.7 | Dynamic causal modeling

Based on the group peak MNI coordinates of the anatomic ROIs, we redefined the corresponding spherical volumes of interest (VOIs) for the DCM. Considering the individual brain activation variances, the radius of the VOIs defined at the group level was extended to 6 mm, and these anatomic VOIs were utilized as the masks (Liu et al., 2010) to detect each participant's local maxima close to the group maxima, thresholded with $p < .05$, uncorrected (Heim et al., 2009; Heim, Eickhoff, & Amunts, 2009; Makuuchi & Friederici, 2013). The peak MNI coordinates at the individual level were the centers of the spherical VOIs of a 4 mm radius for each participant, and the mean peak MNI coordinates with the SDs of the VOIs were listed as follows: (a) the WCI group: the left BA 45 (M: −52.85, 24.00, 1.38; SD: 2.30, 3.87, 2.90), the left BA 44 (M: −48.92, 1.15, 11.31; SD: 3.33, 3.58, 2.18), the left STG (M: −49.38, −2.77, 3.92; SD: 3.38, 3.77, 2.56), and the left FG (M: −44.77, −66.23, −12.23; SD: 2.59, 2.86, 2.59); (b) the non-WCI group: the left BA 45 (M: −54.00, 25.93, 0.64; SD: 2.88, 4.01, 2.41), the left BA 44 (M: −49.29, 2.36, 13.50; SD: 3.05, 3.93, 1.95), the left STG (M: −48.21, −3.43, 2.79; SD: 2.75, 3.30, 3.81), and the left FG (M: −45.86, −66.86, −11.36; SD: 2.18, 2.74, 2.41). The time series data were extracted from each VOI as the eigenvariate, that is, the first principal component, of all voxel time series within it at the single subject level for the model definition later on.

The differences in the model structure between the WCI and the non-WCI group were of primary interest. Hence, we constructed the models for each group respectively, and the experimental perturbation came from Task (relative to Baseline). For each group, in the process of model definition and estimation, the intrinsic connections, the modulatory effects (Task), and the input were taken into account. We defined the left FG as the input brain area for delivering the word form information to higher-order language processing areas. The left FG and the left STG were reported to belong to the "graphophonological conversion access" (Jobard et al., 2003) or the "ventral visual written words stream" (Marinkovic et al., 2003) for word reading, and the functional connection between these areas was positively correlated with Chinese reading performance (He, 2015). Therefore, the left FG was assumed to be intrinsically connected with the left STG in our models. The interconnections between the Broca's area (BA 45 and BA 44) and the left STG were also intrinsic (see Friederici et al., 2017 and the related references therein). The functional connections among the VOIs were modeled to be bidirectional (cf. Makuuchi & Friederici, 2013) and would be modulated by the experimental factor, Task (relative to the baseline), together with their self-connections.

The same four models were constructed for each group (Figure 2). Because of the intrinsic connection between the left FG and the left STG, conveying the visual word(s) information, the left STG should be functionally connected to one or both subregions of Broca's area so as to further process the language information if any modulatory effect existed. Therefore, the model merely composed of the bidirectional connection between the left BA 44 and BA 45 (BA 44 ↔ BA 45) and STG ↔ FG was not constructed.

All the models would be compared within each group by BMS which is sufficient for inference on the model space (Stephan et al.,
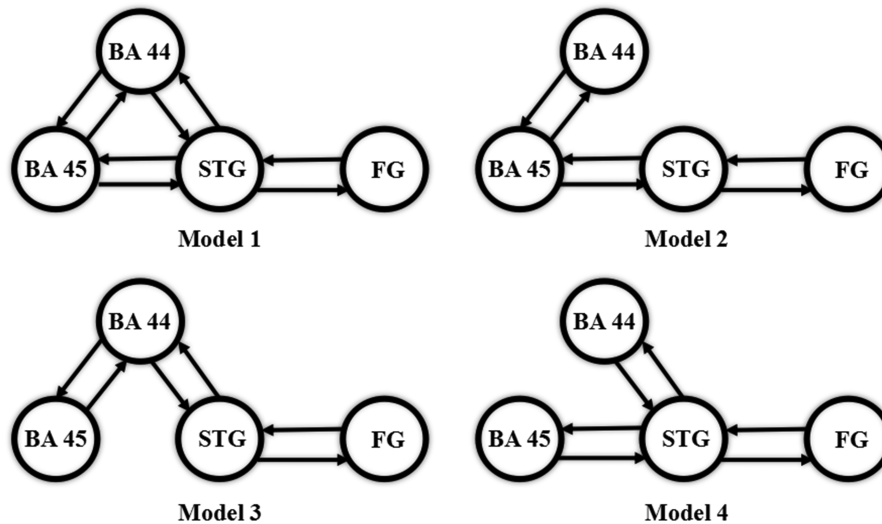
FIGURE 2 The same four dynamic causal models for each group (only presenting the bidirectional connections among the VOIs, on which the modulatory effect, Task, might exert). Model 1 hypothesizes the full connections among the left BA 44, BA 45, and STG; Models 2 and 3 hypothesize the unipath interconnecting the left STG and Broca's area (either BA 45 or BA 44); Model 4 stands for a dual path model in which the left STG is functionally connected with the left BA 44 and BA 45, respectively, while there is no collaboration between BA 44 and BA 45. To note, the left FG is the brain area for receiving the input

2010), with inference to the fixed-effects (FFX) assuming that the optimal model was the same for all the participants from their own groups. The posterior model probability would be calculated for each model to decide the optimal one for each group (Liu et al., 2010). The mean modulatory parameters of the interarea connections in the optimal model were compared to 0 by one-sample t test with the threshold under the Bonferroni correction. If the two groups share the same optimal model, a further two-sample t test comparing the corresponding parameters would be performed with the threshold Bonferroni corrected.

3 | RESULTS

3.1 | Behavioral results

The intergroup comparisons showed that the WCI group performed much better than the non-WCI group (the ACC of the WCI group: $M = 86.5\%$, $SD = 16.8\%$; the ACC of the non-WCI group: $M = 53.7\%$, $SD = 5.6\%$; $t(25) = 6.72$, $p < .001$), and the random level ($t(12) = 18.61$, $p < .001$), while the non-WCI group's performance was not significantly different from 50% ($t(13) = 2.47$, $p = .028$ (>.017),
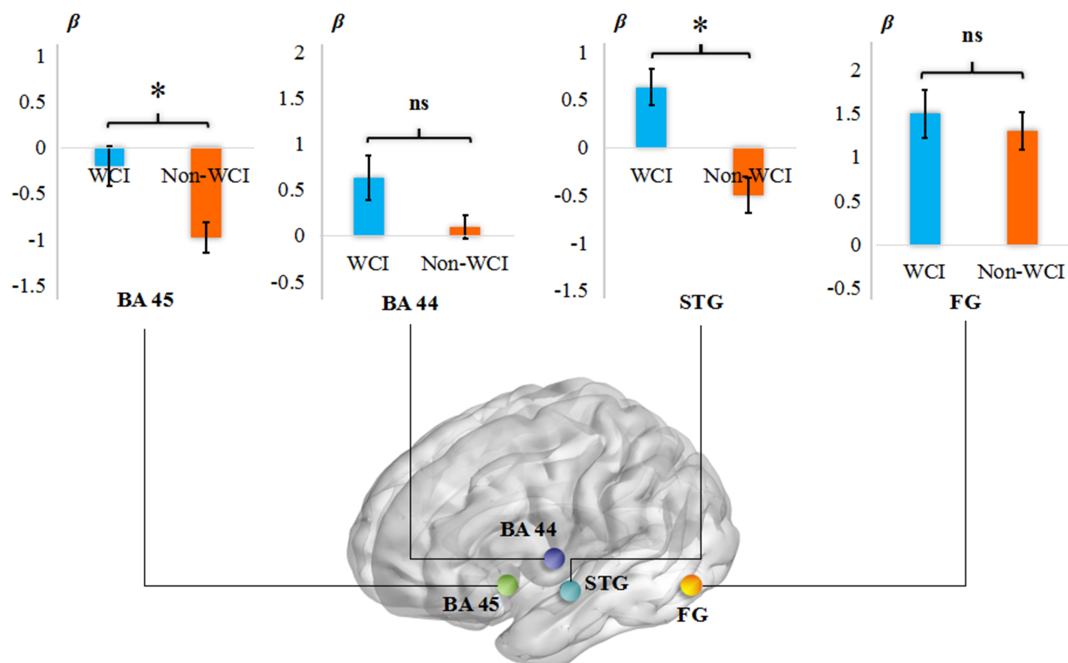
FIGURE 3 The comparison of the β values of the WCI group (WCI) and the non-WCI group (non-WCI) for the ROIs. Error bars show the SE of the mean. $*p < .0125$; ns = not significant, Bonferroni corrected. Note, the activation difference in the left BA 45 between WCI and non-WCI is a difference on the negative side: less activation for non-WCI [Color figure can be viewed at wileyonlinelibrary.com]

Bonferroni corrected). The WCI group's RT was also significantly shorter than the non-WCI group (the RT of the WCI group: M = 751.58 ms, SD = 131.08 ms; the RT of the non-WCI group: M = 1,062.64 ms, SD = 295.10 ms; $t(25)$ = −3.49, p < .005).

The results of the comparison between the grammatically correct and the incorrect conditions within each group can be found in the Supporting Informarion.

3.2 | The results of the ROI analyses

The ROI analyses' results revealed a larger activation in the WCI group compared to the non-WCI group (see Figure 3), significant for BA 45 and STG ($ts(25)$ ≥ 2.91, ps < .01), whereas the difference of BA 44 failed to reach the statistically significant level ($t(25)$ = 1.94, p = .069). Although BA 44 reflected an increase in activation, the difference in BA 45 was characterized by a decrease in activation from the non-WCI group compared to the WCI group. Moreover, there was no significant difference between the two groups in the left FG assumed as a common area for visual word form processing (t (25) = 0.56, p = .581).

3.3 | DCM results

The results of BMS showed that Model 1 was the optimal model for the WCI group (the posterior model probability was more than 99.99% compared with the other models whose posterior model probabilities were all less than 0.01%), whereas Model 3 was the best for the non-WCI group (the posterior model probability was also more than 99.99%, and the posterior model probabilities of the left three models were less than 0.01% as well). All the connection strengths (i.e., the modulatory parameters) of Model 1 except BA 45 ↔ BA 44 were significantly different from the baseline (0) ($ts(12)$ ≥ 6.07 or $ts(12)$ ≤ −4.17, ps < .006, Bonferroni corrected; for BA 45 ↔ BA 44: −1.33 ≤ $ts(12)$ ≤ 0.34, p > .006, Bonferroni corrected). As to Model 3, the t-test results showed that only the connection strength of BA 45 → BA 44 failed to reach the significant level ($ts(12)$ = −2.06, ps = .061, Bonferroni corrected; for the other connection strengths: ts (12) ≥ 5.19 or $ts(12)$ ≤ −4.56, p < .008, Bonferroni corrected). Figure 4 showed the schematic plots of the optimal models for the corresponding groups.

4 | DISCUSSION

The current study aimed at exploring the neural substrates of the WCI as the basis for the complex hierarchical syntactic structure processing, during which the WCI functions to label the new components by forming the critical syntactic relations (i.e., the long-distance dependencies) and constructing the syntactic hierarchies. Behaviorally, the WCI group outperformed the non-WCI group and the random level and could successfully process the complex syntactic hierarchical structures in a rule-based way. In terms of the neural substrates, the WCI group had stronger activations of the Broca's area (especially the left BA 45) and the left STG respectively. Moreover, its effective connection pattern of these regions was clearly different from that of the non-WCI group. These results, for the first time, elucidated that the WCI plays a critical role in complex syntactic hierarchical structure processing, even though the experimental language lacks rich morphological information, such as Chinese, and that the rule-based use of the WCI for the complex hierarchical syntactic structure processing involves a left temporo-frontal network consisting of the STG and Broca's area in the inferior frontal gyrus (IFG). We will discuss the results in detail below.

4.1 | Mapping the labeling competence onto the ability of the rule-based use of the WCI

Given the differences between the behavioral performances of the two groups, it is apparent that, without the WCI, the complex syntactic hierarchical structure building would be highly constrained. Under the framework of the "strong minimalist thesis" which assumes the Universal Grammar (UG) to be the simplest computational principles, Chomsky (2017) emphasized that Merge should be the candidate for the simplest computational operation, which constructs a *new* object Z from X and Y. The problem is, how do we know the set {X, Y} (= Z) is *new* and different from a random string like [X, Y]? Moreover, why are X and Y selected to form a new syntactic object but not the others? Under the *linguistic* perspective of labeling, the answers appear to be clear. Thanks to WCI, the syntactic objects could be selectively picked out to enter the further computation process during which they would be merged and labeled (also see Zaccarella & Friederici, 2015) and then enter the next processing phase recursively; simultaneously, the syntactic hierarchical structure could be built in a cascadic way.

This is demonstrated in the present study. The WCI group could form both the local (such as Mod – V and N – V) and the long-distance dependencies (in particular the critical relations between $V_{1/2}$ and $R_{1/2}$) among the syntactic objects on the basis of the WCI, while the processing of the non-WCI group was "irregular" and had no difference from the random level, indicating that this group failed in abstracting out the accurate WCI and the key syntactic rules. This is in line with Gómez and Gerken's (2000) comment that the "category-based abstraction" is critical for rule extraction. The random (i.e., unsystematic) processing of the non-WCI group further indicated that this group might mainly focus on dealing with the local relations among the single words or chunks at a comparatively concrete level, even though the materials had been controlled and did not show a performance bias as indicated by the ACC of the non-WCI which was not significantly different from the random level. Moreover, both groups' performances would be significantly different from 50% due to the superficial clues/strategies. Therefore, the abilities to identify the WCI/labels of syntactic objects, to compute the relations among the word categories, and to construct the labeled hierarchical syntactic structures constitute might be essential for the human language faculty, and thereby the capacity to build syntactic hierarchical structure.

It is noteworthy that, because using the WCI for the complex hierarchical syntactic structure processing is closely related to its role played in the complex syntactic rule acquisition, we would like to discuss the underlying cognitive mechanism of using the WCI, fostering both the acquisition/learning of the syntactic rules and the processing
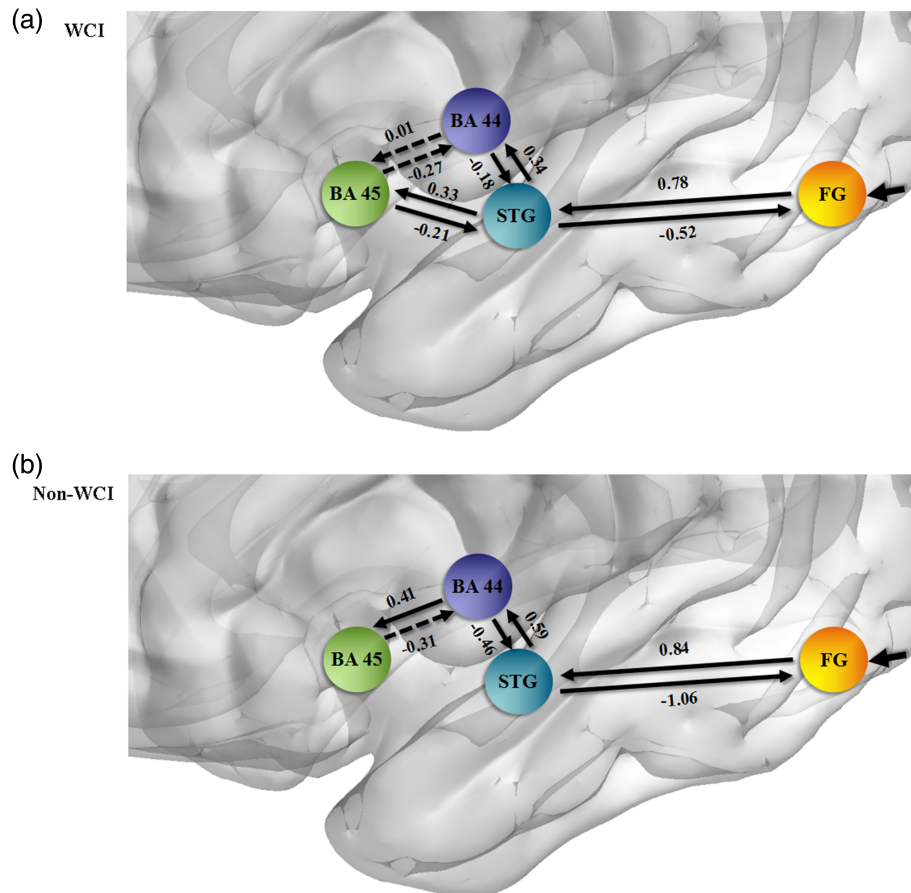
FIGURE 4 The schematic plots of the optimal dynamic causal models for (a) the WCI group and (b) the non-WCI group, respectively. The mean modulatory parameter of every connection was given. The dashed arrows denote the connections with the modulatory parameters that were not significantly different from the baseline. The solid arrow pointing to the left FG stands for the driving input with its direction [Color figure can be viewed at wileyonlinelibrary.com]

of the complex hierarchical syntactic structures, which is also enlightening for us to map the linguistic construct (i.e., the labeling competence) onto the cognitive one(s). In light of the learning results (see Introduction, Participants, and the Supporting Information) and the behavioral results of the present study, compared with the non-WCI group's poor performance, the WCI group successfully acquired the complex syntactic rules and could use them to process the new sentences in the current fMRI scanning session, indicating that the WCI group could use the WCI to construct the target syntactic rules and to process the complex hierarchical syntactic structures in a rule-based fashion, but not on the other non-rule-based learning/processing mechanisms. The statistical learning mainly relies on the transition probabilities between adjacent elements, that is, "in local dependencies and usually at fixed distances" (Goucha et al., 2017), which is assumed not to be sufficient to account for the syntactic rule acquisition, especially when the rules are comparatively complex (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015; Ding, Melloni, Hang, Xing, & Poeppel, 2016; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006; Goucha et al., 2017; Opitz & Friederici, 2004, 2007; Peña, Bonatti, Nespor, & Mehler, 2002; Skeide & Friederici, 2016). As the experimental materials used in the present study and our previous behavioral research (Chen et al., 2018) consist of the complex hierarchical syntactic structures with the long-distance dependencies and the relative clause center embedded, the WCI group might follow a

rule-based way to use the WCI for both the complex syntactic rule acquisition and the complex hierarchical syntactic structure processing. Moreover, the transition probabilities between the adjacent words are the same for both types of the trials (i.e., the correct and the incorrect ones), so the WCI group may unlikely use these statistics to identify the grammaticality of the trials. As to the similarity-based learning, Domangue, Mathews, Sun, Roussel, and Guidry (2004) claimed that the accuracy of the similarity-based learning is very limited. Opitz and Friederici (2004) found that the similarity-based learning would activate the left anterior hippocampus, which is different from the area (i.e., the left ventral premotor cortex) activated by the rule-based learning, which was also demonstrated by Hauser, Hofmann, and Opitz (2012). On one hand, the chunk strength has been carefully controlled during the learning (see Supporting Information), and thus, both groups could hardly extract the rules by merely evaluating the superficial similarity to the exemplars in the training material. On the other hand, the brain activation pattern of our research is compatible with Opitz and Friederici (2004, 2007) and Hauser et al. (2012) (see the discussions for details below; also see the critical review of Dehaene et al., 2015). Therefore, we assume that the WCI group's use of the WCI to process the complex hierarchical syntactic structure is in a rule-based way. A recent research of Opitz and Hofmann (2015) further evidenced that the non-rule-based learning mechanism (the similarity-based learning in their study) could function

at the beginning of the learning session, and the rule-based learning might be decisive for the learning result, in line with Opitz and Friederici (2004). Similarly, it is reasonable that the non-rule-based learning might narrow down the solution space by ruling out some transitions among the word categories at the very beginning of the learning so as to promote the efficiency of using the WCI to establish the syntactic rules, but it seems unlikely to determine the final results, especially when the syntactic rules are, to some extent, complicated. Moreover, as Opitz and Hofmann (2015) proposed, the rule-based learning will happen especially when the learning is explicit. Due to the fact that we explicitly told all the participants to extract the possible abstract rules and the WCI-group was explicitly equipped with the WCI knowledge in the learning session, the rule-based use of the WCI for syntactic rule acquisition should be more convincing.

After demonstrating that the WCI group's use of the WCI for the complex hierarchical syntactic structure processing (and the complex syntactic rule acquisition) is relying on the rule-based learning/processing mechanism, we could finally map the linguistic construct, the labeling competence, substantially onto the cognitive construct, that is, the ability of the rule-based use of the WCI. Although the mapping between the two constructs seems reasonable, the specialty of the construct, the labeling competence, is undetermined yet. Under the cognitive perspective, the "labeling competence" seems to be a mere linguistic construct, although this study has revealed the underlying cognitive mechanism for the rule-based use of the WCI to process the complex hierarchical syntactic structures. From the linguistic perspective, it is interesting to know whether the ability to use the WCI for the syntactic processing in a rule-based way is language-specific or cognitive-general. If such an ability is language specific, then the labeling competence should also be a special construct in the field of cognition, distinct from the other general cognitive mechanisms. We will leave this issue open to the future researches.

4.2 | Brain areas supporting the rule-based use of the WCI for the syntactic processing

4.2.1 | The left STG

In the present experiment, the left STG was significantly activated for the WCI group in the ROI analyses (and was further confirmed by the whole-brain analyses [see Supporting Information]), indicating its importance in supporting the rule-based use of the WCI for complex hierarchical syntactic structure processing. The syntactic processing role played by the left STG has been reported by several researches (Brennan, Stabler, Wagenen, Luh, & Hale, 2016; Caplan et al., 2002; Caplan, Chen, & Waters, 2008; Ding et al., 2016; Friederici, 2011; Friederici et al., 2003; Friederici, Kotz, Scott, & Obleser, 2010; Hagoort & Indefrey, 2014; Matchin, Hammerly, & Lau, 2017; Matchin, Sprouse, & Hickok, 2014; Nelson et al., 2017; Newman, Just, Keller, Roth, & Carpenter, 2003; Pallier, Devauchelle, & Dehaene, 2011; Röder, Stock, Neville, Bien, & Rösler, 2002; Stowe et al., 1998). Generally, the left STG could respond to the expressions as long as they were hierarchically structured. Critically, Pallier et al. (2011) showed that the increase of the scale of the phrase structure was positively correlated to the activation level of pSTS/STG, similar to the recent research of Nelson et al. (2017) which reported that the activation of

STG increased with the number of the syntactic nodes to be merged. Ding et al. (2016) found STG to be present in the patterns of the high-frequency γ energy oscillation from phrase level up to the sentence level. Skeide et al. (2016) further found that the level of the activation of STG could predict the development of the individual complex syntactic competence indexed by the performance of the processing of the complex hierarchical syntactic structure with the relative clause center embedded.

However, the comparatively specific role played by the left STG in the (complex) hierarchical syntactic structure processing was still disputed. It was either deemed as a memory component (Boeckx, Martinez-Alvarez, & Leivada, 2014; Hagoort, 2005, 2013; Kuhnke, Meyer, Friederici, & Hartwigsen, 2017; Meyer, Obleser, Anwander, & Friederici, 2012; Xiang, Fonteijn, Norris, & Hagoort, 2010), related to the storage of the syntactic node information (Hagoort, 2005), and the storage and the check of the syntactic frame information (Herrmann, Maess, Hasting, & Friederici, 2009), or an area for syntactic integration (Friederici, 2011; Friederici et al., 2003, Friederici et al., 2010; Matchin et al., 2014; Newman et al., 2003). Combining with the results of DCM, we are inclined to consider the left STG as a multi-functional component which could retrieve the WCI (or label) of the syntactic objects (i.e., the node information) and the labeled hierarchical syntactic templates for later processing, and then integrate and store the various new syntactic information generated from Broca's area in a rule-based way. Therefore, the more complex hierarchical syntactic structure labeled in Broca's area would cause the increase of the complexity of its integration and storage in the left STG, and conversely, the more complex constituent retrieved from the left STG would also augment the complexity of processing in Broca's area. Moreover, the role of the left STG also implied that this region was crucial for relatively abstract information processing, that is, the operations at the WCI level, but not only for the concrete element computation.

4.2.2 | Broca's area

The ROI analyses revealed the significant activation differences of Broca's area between the two groups. More specifically, the left BA 45 was more strongly activated for the WCI group, while the left BA 44 failed to show a similar intergroup difference. Broca's area has been numerously reported to be involved in the hierarchical syntactic structure processing for both the simple (Ding et al., 2016; Fedorenko, Duncan, & Kanwisher, 2012; Folia, Forkstam, Ingvar, Hagoort, & Petersson, 2011; Goucha & Friederici, 2015; Iijima, Fukui, & Sakai, 2009; Nelson et al., 2017; Okada et al., 2013; Pallier et al., 2011; Wang et al., 2008; Zaccarella & Friederici, 2015) and the complex hierarchical structures (Bahlmann, Schubotz, & Friederici, 2008; Friederici, 2004; Friederici et al., 2006; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Keller, Carpenter, & Just, 2001; Makuuchi, Bahlmann, Anwander, & Friederici, 2009; Skeide et al., 2016; Stowe et al., 1998). Broca's area appears to participate in structured sequence processing regardless of the complexity, in which "structured" denoted the syntactic dependencies among the elements in the sequence (Uddén & Bahlmann, 2012). To note, some researches also suggested the role of the syntactically specific or language-specific working memory played

by Broca's area (Boeckx et al., 2014; Caplan et al., 2002; Cooke et al., 2002; Fiebach, Schlesewsky, Lohmann, von Cramon, & Friederici, 2005; Santi & Grodzinsky, 2012; Stowe et al., 1998), but according to Makuuchi et al. (2009), such a working memory competence was supported by the left inferior frontal sulcus rather than Broca's area in IFG.

Broca's area can be receptorarchitectonically partitioned into several subregions, among which the left posterior BA 45 and the left ventral BA 44 are proposed to be related to syntactic processing (Amunts et al., 2010). The syntactic function of the left BA 44 has been demonstrated by several studies. For instance, Makuuchi et al. (2009) varied the length of dependency and the number of hierarchies to detect the brain areas for working memory and syntactic processing respectively, proposing that the left BA 44 supported syntactic processing. Goucha and Friederici (2015) found that the left BA 44 was activated only under the jabberwocky condition with the inflectional affixes reserved. More elaborately, Zaccarella and Friederici (2015) pinpointed that the anterior-ventral part of the left BA 44 was responsible for Merge with its anterior-ventral part transforming the string of words into a "hierarchically labeled syntactic structure." Interestingly, activation difference of the left BA 44 between the WCI group and the non-WCI group failed to reach the level of significance. One possibility is noteworthy: the activation of the left BA 44 for the WCI group was subtracted away (Friederici et al., 2017) because of the ungrammatical Merge conducted on the concrete words or chunks (but not the labels at the abstract level) by the non-WCI group. This possibility also implies that the intention to merge two constituents into a new one is intrinsic for human beings despite the grammaticality, although, apparently, the degree of the abstractness and the grammaticality would definitely constrain the rule-based use of the WCI for hierarchical syntactic structure construction in a recursive fashion.

As the left posterior BA 45 could be receptorarchitectonically distinguished from the anterior part, and the current experiment is independent of the semantic information, the stronger activation of the left BA 45 for the WCI group should be related to syntactic processing, especially the rule-based use of the WCI. Recently, Santi and Grodzinsky (2012) proposed that BA 45 could be significantly activated as long as the elements of a certain structure were predictable. Matchin et al.'s (2014) research showed that BA 45 was activated both in the wh-movement and the back anaphora structures, claiming that BA 45 was responsible for the syntactic prediction regardless of the types of the syntactic structures. The critical long-distance dependencies between $V_{1/2}$ and $R_{1/2}$ and the other syntactic dependencies in our experiment were similar to the back anaphora structure(s) in that the former word categories would predict the latter ones. Hence, we assumed that the left BA 45 might support the syntactic prediction on the basis of the syntactic rules during the syntactic processing. Moreover, we do not deny the semantic processing role possibly played by the left BA 45; it is unreasonable to assume a certain brain area has only one function, especially when it could be separated into several subregions, and when it may be involved in different brain networks (Friederici & Singer, 2015).

4.3 | Effective connectivity supporting the rule-based use of the WCI for the syntactic processing

The DCM results in the present study showed that the optimal model for the WCI group was Model 1, a full effective connection model with the left STG connecting to Broca's area in a dual-path fashion, and the non-WCI group Model 3 with the left STG connecting to Broca's area (the left BA 44) via a unipath, lacking the left STG ↔ the left BA 45.

Previous research has demonstrated that the dorsal tract (anatomically, the arcuate fasciculus) connecting the left STG (especially the posterior part) and the Broca's area (especially the left BA 44) might be critical for syntactic processing (Brauer et al., 2013; Dick et al., 2014; Friederici, 2011; Friederici, 2012; Friederici, 2017; Friederici & Gierhan, 2013; Friederici & Singer, 2015; Meyer et al., 2012; Skeide & Friederici, 2016; Wilson et al., 2011) and was deemed to be human specific (Friederici, 2016; Goucha et al., 2017). Some research also proposed the ventral pathway connecting the left temporal lobe and the frontal lobe was subdivided in a fiber tract to the left frontal operculum functionally relevant for local combinatorics and a fiber tract going to the left BA 45 relevant for lexical-semantic processing (Friederici et al., 2017). However, due to the fact that the anatomical connections could not determine the effective connectivity because the synaptic connections could be functionally expressed in a dynamic and context-sensitive way (see Stephan et al., 2010 and the related references therein), the current DCM results could not decide whether the effective connections belong to which anatomical pathway(s). But the clear pattern is that Model 1 for the WCI group hypothesized a tighter relationship between the left STG and Broca's area because of the dual-path effective connectivity, and this pattern might be critical for the rule-based use of the WCI for the complex hierarchical syntactic structure processing. Flöel, de Vries, Scholz, Breitenstein, and Johansen-Berg (2009) found that the white matter integrity within Broca's area and between the STG and Broca's area could both predict the grammar learning performance. The closer and full connection pattern of Model 1 does support the WCI group's syntactic performance.

The most critical difference between these two models was located in the existence of the left STG ↔ the left BA 45 for Model 1. den Ouden et al.'s (2012) DCM analysis also indicated that the effective connectivity between the left BA 45 and the left posterior STG was critically in service of complex syntactic processing. Combined with the functions of the left BA 45 and the left STG discussed above, the left STG ↔ the left BA 45 of Model 1 might be in the service of label/WCI recognition, retrieval, and syntactic prediction with checking on the basis of WCI, following the syntactic rules. Without the left STG ↔ the left BA 45, the non-WCI group's ability of the rule-based using the WCI for syntactic processing was highly constrained, failing to process the complex hierarchical syntactic structures at the WCI level.

Interestingly, Model 1's modulatory parameters of the left BA 45 ↔ the left BA 44 failed to be different from the baseline significantly. One possibility is, as suggested by Makuuchi and Friederici (2013), that the effective connectivity of the left BA 45 ↔ the left BA 44 was an indirect one, interfered by the dual pathways between the

left STG and Broca's area. A more convincing possibility is that the left BA 45 and BA 44 could have been automatically coactivated for the WCI group irrespective of the exertion of the modulatory effect, that is, this coactivation pattern of the WCI group had already been formed even before the participants started the task because of the unconstrained ability of the rule-based use of the WCI (cf. Heim, Eickhoff, & Amunts, 2009). As for the non-WCI group, the modulatory effect of the left BA 44 → the left BA 45 was significantly beyond the baseline, indicating this connection to be more controlled, possibly due to the lexical decision of BA 45 (Heim, Eickhoff, & Amunts, 2009; Heim, Eickhoff, Ischebeck, et al., 2009) for the ungrammatical combinations or chunks, whereas the feedback from the left BA 45 to the left BA 44 was automatic, independent of the modulators. We would leave this phenomenon open to future research.

5 | FURTHER CAVEATS FOR DISCUSSION

The differences between the WCI and the non-WCI groups also well reflected the difference of proficiency between the two groups, as evidenced by the behavioral results of the scanning session that the WCI group outperformed the non-WCI group. Indeed, proficiency was promoted by the ability of the rule-based use of the WCI, that is, owing to the WCI and the rule-based use of the WCI, the WCI group could process the complex hierarchical syntactic structures in a more proficient and efficient fashion, whereas such an ability of the non-WCI group was highly constrained by the lack of the WCI, so the participants of this group could hardly rely on the rule-based use of the WCI and became much less proficient. Therefore, in the present study, proficiency per se is, by no means, contrary to but rather is the reflection of the unconstrained ability of the rule-based use of the WCI, and both the brain activation and the effective connectivity pattern of the WCI group revealed the neural substrates of the ability of the rule-based use of the WCI, which support this group to process the complex hierarchical syntactic structures more proficiently.

Moreover, different from the WCI group who had acquired the corresponding syntactic rules so as to process the syntactic structures by using the WCI in a rule-based way, the non-WCI group might incline to rely on "guessing," which could, in theory, partly affect this group's brain activation and effective connectivity patterns. In accordance with the scanning result that there was no significant difference of the activation of the left BA 44 crucial for Merge (Zaccarella & Friederici, 2015) between the two groups, the participants of the non-WCI group might mainly resort to the local combinations of certain adjacent concrete words (but not the WCI) to falsely process the target structures. Thus, in the view of "guessing," the individuals of the non-WCI group might utilize the non-rule-based information (such as the superficial similarity) alongside with guessing to target various chunks/combinations for syntactic processing, whose ability of the rule-based use of the WCI was highly constrained, and therefore the resulted brain activation and connectivity patterns were distinct from those of the WCI group.

Furthermore, given the significantly longer RT of the non-WCI group, the "time-on-task" effect that the variability of the trial-by-trial RT might affect the hemodynamic response (Barber, Pekar, &

Mostofsky, 2016; Domagalik, Beldzik, Oginska, Marek, & Fafrowicz, 2014; Grinband et al., 2011; Grinband, Wager, Lindquist, Ferrera, & Hirsch, 2008; Yarkoni, Barch, Gray, Conturo, & Braver, 2009) also seemed to result in the group differences of the activation and the connectivity patterns. However, due to the current blocked design (each block lasted for 72 s and the duration of each trial was 9 s), the regressors were constructed as boxcar functions with the block duration equal to the duration of the stimuli (cf. Grinband et al., 2008) (to note, the trial duration was also fixed to 9 s, and the final response screen would not disappear if the RT was shorter than its duration), which was different from the "impulse functions" adopted by most rapid event-related designs that underestimated the nontrivial impact of the RT variation of the stimuli with much shorter durations (Grinband et al., 2008). It is also noteworthy that the RT of the current experiment was collected at the final response screen, before which the critical long-distance dependencies should have already been processed. We would like to leave the question whether such blocked designs should regress out the RT as a regressor of no interest open to future methodological researches. Moreover, the slower RT would induce greater amplitude of the hemodynamic response (Barber et al., 2016; Domagalik et al., 2014; Grinband et al., 2011; Yarkoni et al., 2009), but the WCI group with shorter RT showed more activation of Broca's area and the left STG, indicating that the group differences should not be ascribed to the task-on-time effect. The studies further revealed that the task-on-time effect might engage the attentional network (Domagalik et al., 2014) or a frontal–parietal network reflecting the task-on-time effect of sustained attention (Yarkoni et al., 2009), distinct from the frontal-temporal language network we have been discussing. Therefore, although behaviorally the two groups differed in RT, the "time-on-task" effect seemed unlikely to become the main factor to affect the brain activation/connectivity differences between the two groups.

6 | CONCLUSION

By comparing both the behavioral performances and the neural activities between the WCI group and the non-WCI group, our fMRI research, for the first time, revealed that the left STG, Broca's area, and the effective connectivity showing a dual-pathway pattern between them was crucial for the (ability of the) rule-based use of the WCI for the complex hierarchical syntactic structure processing. Furthermore, the labeling competence within the linguistic framework could be mapped onto the cognitive construct, that is, the ability of the rule-based use of the WCI. However, the specific suboperations of the whole process and the corresponding neural substrates await refinement. And the specific anatomical and functional connections between the left temporal lobe and Broca's area still need to be elaborated. Moreover, the specialty of the ability of the rule-based use of the WCI remains yet unclear. Nevertheless, as such an ability is also crucial in the highly analytic language processing (such as for Chinese, which lacks of abundant morphological marks) and the neural substrates are part of the core language network, we further hypothesize that the ability of the rule-based use of the WCI might be universal as

the essence of human language faculty under the perspective of neurolinguistics.

ORCID

Liping Feng https://orcid.org/0000-0002-5463-0476

REFERENCES

Amunts, K., Lenzen, M., Friederici, A. D., Schleicher, A., Morosan, P., Palomero-Gallagher, N., & Zilles, K. (2010). Broca's region: Novel organizational principles and multiple receptor mapping. *PLoS Biology*, *8*, 1000489.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B. M., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and intersubject variability. *The Journal of Comparative Neurology*, *412*, 319–341.

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

Barber, A. D., Pekar, J. J., & Mostofsky, S. H. (2016). Reaction time-related activity reflecting periodic, task-specific cognitive control. *Behavioural Brain Research*, *296*, 100–108.

Boeckx, C., Martinez-Alvarez, A., & Leivada, E. (2014). The functional neuroanatomy of serial order in language. *Journal of Neurolinguistics*, *32*, 1–15.

Brauer, J., Anwander, A., Perani, D., & Friederici, A. D. (2013). Dorsal and ventral pathways in language development. *Brain and Language*, *127*, 289–295.

Brennan, J. R., Stabler, E. P., Wagenen, S. E. V., Luh, W.,. M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic com-prehension. *Brain and Language*, *157–158*, 81–94.

Caplan, D., Chen, E., & Waters, G. (2008). Task-dependent and task-independent neurovascular responses to syntactic processing. *Cortex*, *44*, 257–275.

Caplan, D., Vijayan, S., Kuperberg, G., West, C., Waters, G., Greve, D., & Dale, A. M. (2002). Vascular responses to syntactic processing: Event-related fMRI study of relative clauses. *Human Brain Mapping*, *15*, 26–38.

Caspers, J., Zilles, K., Eickhoff, S. B., Schleicher, A., Mohlberg, H., & Amunts, K. (2013). Cytoarchitectonical analysis and probabilistic mapping of two extrastriate areas of the human posterior fusiform gyrus. *Brain Structure & Function*, *218*, 511–526.

Chen LY, Han X, Wu JJ, Fu YB, Kang KT, Li ZS, Feng LP (2018) Researching on the word category information-based labeling competence for second language complex hierarchical syntactic rule construction. (Submitted for publication.) Chinese Teaching in the World.

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* (pp. 3–16). Amsterdam: John Benjamins Publishing Company.

Chomsky, N. (2017). Language architecture and its import for evolution. *Neuroscience and Biobehavioral Reviews*, *81*, 295–300.

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

Dehaene, S., Meyniel, F., Wacongne, C., Wang, L., & Pallier, C. (2015). The neural representation of sequences: From transition probabilities to algebraic patterns and linguistic trees. *Neuron*, *88*, 2–19.

Dehaene, S., Le, C. G., Poline, J. B., Le, B. D., & Cohen, L. (2002). The visual word form area: A prelexical representation of visual words in the fusiform gyrus. *Neuroreport*, *13*, 321–325.

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

Ding, N., Melloni, L., Hang, Z., Xing, T., & Poeppel, D. (2016). Cortical tracking of hierarchical linguistic structures in connected speech. *Nature Neuroscience*, *19*, 158–164.

Dick, A. S., Bernal, B., & Tremblay, P. (2014). The language connectome: New pathways new concepts. *The Neuroscientist*, *20*, 453–467.

Domagalik, A., Beldzik, E., Oginska, H., Marek, T., & Fafrowicz, M. (2014). Inconvenient correlation—RT-bold relationship for homogeneous and fast reactions. *Neuroscience*, *278*, 211–221.

Domangue, T. J., Mathews, R. C., Sun, R., Roussel, L. G., & Guidry, C. E. (2004). Effects of model-based and memory-based processing on speed and accuracy of grammar string generation. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *30*, 1002–1011.

Eickhoff, S. B., Stephan, K. E., Mohlberg, H., Grefkes, C., Fink, G. R., Amunts, K., & Zilles, K. (2005). A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data. *NeuroImage*, *25*, 1325–1335.

Fedorenko, E., & Thompson-Schill, S. L. (2014). Reworking the language network. *Trends in Cognitive Sciences*, *18*, 120–126.

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*, 2059–2062.

Fiebach, C. J., Schlesewsky, M., Lohmann, G., von Cramon, D. Y., & Friederici, A. D. (2005). Revisiting the role of Broca's area in sentence processing: Syntactic integration versus syntactic working memory. *Human Brain Mapping*, *24*, 79–91.

Flöel, A., de Vries, M.,. H., Scholz, J., Breitenstein, C., & Johansen-Berg, H. (2009). White matter integrity in the vicinity of Broca's area predicts grammar learning success. *Neuroimage*, *47*, 1974.

Folia, V., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2011). Implicit artificial syntax processing: Genes preference and bounded recursion. *Biolinguistics*, *5*, 105–132.

Friederici, A. D. (2016). Evolution of the neural language network. *Psychonomic Bulletin & Review*, *21*, 41–47.

Friederici, A. D. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*, 262–268.

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

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

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., Chomsky, N., Berwick, R. C., Moro, A., & Bolhuis, J. J. (2017). Language, mind and brain. *International Journal of Psychophysiology*, 69, 163–164.

Friederici, A. D., & Gierhan, S. M. E. (2013). The language network. *Current Opinion in Neurobiology*, 23, 250–254.

Friederici, A. D. (2004). Processing local transitions versus long-distance syntactic hierarchies. *Trends in Cognitive Sciences*, 8, 245–247.

Friederici, A. D., Kotz, S. A., Scott, S. K., & Obleser, J. (2010). Disentangling syntax and intelligibility in auditory language comprehension. *Human Brain Mapping*, 31, 448–457.

Friederici, A. D., & Singer, W. (2015). Grounding language processing on basic neurophysiological principles. *Trends in Cognitive Sciences*, 19, 329–338.

Friederici, A. D., Steinhauer, K., & Pfeifer, E. (2002). Brain signatures of artificial language processing: Evidence challenging the critical period hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 529–534.

Friederici, A. D., Rüschemeyer, S. A., Hahne, A., Fiebach, C. J., & Christian, J. (2003). The role of left inferior frontal and superior temporal cortex in sentence comprehension: Localizing syntactic and semantic processes. *Cerebral Cortex*, 13, 170–177.

Friston, K.,, J., Harrison, L., & Penny, W. (2003). Dynamic causal modeling. *NeuroImage*, 19, 1273–1302.

Gómez, R. L., & Gerken, L. A. (2000). Infant artificial language learning and language acquisition. *Trends in Cognitive Sciences*, 4, 178–186.

Goucha, T. B., & 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 the homo loquens as a builder of labeled structures: Neurocognitive considerations. *Neuroscience and Biobehavioral Reviews*, 81, 213–224.

Grinband, J., Wager, T. D., Lindquist, M., Ferrera, V. P., & Hirsch, J. (2008). Detection of time-varying signals in event-related fMRI designs. *NeuroImage*, 43, 509–520.

Grinband, J., Savitsky, J., Wager, T. D., Teichert, T., Ferrera, V. P., & Hirsch, J. (2011). The dorsal medial frontal cortex is sensitive to time on task, not response conflict or error likelihood. *NeuroImage*, 57, 303–311.

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, 416.

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

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

Hauser, M. F. A., Hofmann, J., & Opitz, B. (2012). Rule and similarity in grammar: Their interplay and individual differences in the brain. *NeuroImage*, 60, 2019–2026.

He, Q. (2015). *The effects of Chinese word acquisition on reading networks at rest*. Shanghai, China: East China Normal University.

Heim, S., Eickhoff, S.,. B., & Amunts, K. (2009). Different roles of cytoarchitectonic BA 44 and BA 45 in phonological and semantic verbal fluency as revealed by dynamic causal modelling. *NeuroImage*, 48, 616–624.

Heim, S., Eickhoff, S. B., Ischebeck, A. K., Friederici, A. D., Stephan, K. E., & Amunts, K. (2009). Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. *Human Brain Mapping*, 30, 392–402.

Herrmann, B., Maess, B., Hasting, A. S., & Friederici, A. D. (2009). Localization of the syntactic mismatch negativity in the temporal cortex: An MEG study. *NeuroImage*, 48, 590–600.

Iijima, K., Fukui, N., & Sakai, K. L. (2009). The cortical dynamics in building syntactic structures of sentences: An MEG study in a minimal-pair paradigm. *NeuroImage*, 44, 1387–1396.

Jobard, G., Crivello, F., & Tzourio-Mazoyer, N. (2003). Evaluation of the dual route theory of reading: A metanalysis of 35 neuroimaging studies. *NeuroImage*, 20, 693–712.

Just, M. A., Carpenter, P. A., Keller, T. A., Eddy, W. F., & Thulborn, K. R. (1996). Brain activation modulated by sentence comprehension. *Science*, 274, 114–116.

Keller, T. A., Carpenter, P. A., & Just, M. A. (2001). The neural bases of sentence comprehension: A fMRI examination of syntactic and lexical processing. *Cerebral Cortex*, 11, 223–237.

Kuhnke, P., Meyer, L., Friederici, A. D., & Hartwigsen, G. (2017). Left posterior inferior frontal gyrus is causally involved in reordering during sentence processing. *NeuroImage*, 148, 254–263.

Ling, X. L. (2016). *Implicit learning of artificial phrase structure grammar*. Shanghai, China: East China Normal University.

Liu, L., Vira, A., Friedman, E., Minas, J., Bolger, D., Bitan, T., & Booth, J. (2010). Children with reading disability show brain differences in effective connectivity for visual but not auditory word comprehension. *PLoS One*, 5, e13492.

Makuuchi, M., & Friederici, A. D. (2013). Hierarchical functional connectivity between the core language system and the working memory system. *Cortex*, 49, 2416–2423.

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.

Marinkovic, K., Dhond, R. P., Dale, A. M., Glessner, M., Carr, V., & Halgren, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, 38, 487–497.

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., Sprouse, J., & Hickok, G. (2014). A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and Language*, 138, 1–11.

Meyer, L., Obleser, J., Anwander, A., & Friederici, A. D. (2012). Linking ordering in Broca's area to storage in left temporo-parietal regions: The case of sentence processing. *NeuroImage*, 62, 1987–1998.

Morosan, P., Rademacher, J., Schleicher, A., Amunts, K., Schormann, T., & Zilles, K. (2001). Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, 13, 684–701.

Morosan, P., Rademacher, J., Palomero-Gallagher, N., & Zilles, K. (2005). Anatomical organization of the human auditory cortex: Cytoarchitecture and transmitter receptors. In R. Konig, P. Heil, E. Budinger, & S. H. Mahwah (Eds.), *Auditory cortex: A synthesis of human and animal research* (pp. 27–50). Mahwah, NJ: Erlbaum.

Murphy, E. (2015). Labels cognomes and cyclic computation: An ethological perspective. *Frontiers in Psychology*, 6, 715–732.

Narita, H. (2014). *Endocentric structuring of projection-free syntax* (Vol. 218). Amsterdam, Netherlands: John Benjamins Publishing Company.

Nelson, M. J., El Karoui, I., Giber, K., Yang, X., Cohen, L., Koopman, H., ... Dehaene, S. (2017). Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences of the United States of America*, 114, E3669–E3678.

Newman, S. D., Just, M. A., Keller, T. A., Roth, J., & Carpenter, P. A. (2003). Differential effects of syntactic and semantic processing on the subregions of Broca's area. *Cognitive Brain Research*, 16, 297–307.

Okada, R., Okuda, T., Nakano, N., Nishimatsu, K., Fukushima, H., Onoda, M., ... Kato, A. (2013). Brain areas associated with sentence processing: A functional MRI study and a lesion study. *Journal of Neurolinguistics*, 26, 470–478.

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 neural dissociation of rule-based versus similarity-based learning. *The 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.

Opitz, B., & Hofmann, J. (2015). Concurrence of rule- and similarity-based mechanisms in artificial grammar learning. *Cognitive Psychology*, 77, 77–99.

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

Peña, M., Bonatti, L. L., Nespor, M., & Mehler, J. (2002). Signal-driven computations in speech processing. *Science*, *298*, 604–607.

Pothos, E. M. (2007). Theories of artificial grammar learning. *Psychological Bulletin*, *133*, 227–244.

Röder, B., Stock, O., Neville, H., Bien, S., & Rösler, F. (2002). Brain activation modulated by the comprehension of normal and pseudo-word sentences of different processing demands: A functional magnetic resonance imaging study. *NeuroImage*, *15*, 1003–1014.

Rohrmeier, M., Fu, Q., & Dienes, Z. (2012). Implicit learning of recursive context-free grammars. *PLoS One*, *7*, e45885.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*, 1926–1928.

Santi, A., & Grodzinsky, Y. (2012). Broca's area and sentence comprehension: A relationship parasitic on dependency displacement or predictability? *Neuropsychologia*, *50*, 821–832.

Shanks, D. R. (1995). *The psychology of associative learning*. Cambridge, UK: Cambridge University Press.

Skeide, M. A., Brauer, J., & Friederici, A. D. (2016). Brain functional and structural predictors of language performance. *Cerebral Cortex*, *26*, 2127–2139.

Skeide, M. A., & Friederici, A. D. (2016). The ontogeny of the cortical language network. *Nature Reviews. Neuroscience*, *17*, 323–332.

Stephan, K. E., Penny, W. D., Moran, R. J., den Ouden, H. E. M., Daunizeau, J., & Friston, K. J. (2010). Ten simple rules for dynamic causal modeling. *NeuroImage*, *49*, 3099–3109.

Stowe, L. A., Broere, C. A., Paans, A. M., Wijers, A. A., Mulder, G., Vaalburg, W., & Zwarts, F. (1998). Localizing components of a complex task: Sentence processing and working memory. *Neuroreport*, *9*, 2995–2999.

Uddén, J., & Bahlmann, J. (2012). A rostro-caudal gradient of structured sequence processing in the left inferior frontal gyrus. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 2023–2032.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498–505.

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.

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., … Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, *72*, 397–403.

Xiang, H. D., Fonteijn, H. M., Norris, D. G., & Hagoort, P. (2010). Topographical functional connectivity pattern in the perisylvian language networks. *Cerebral Cortex*, *20*, 549–560.

Yan, C., & Zang, Y. (2010). DPARSF: A MATLAB toolbox for "pipeline" data analysis of resting-state fMRI. *Frontiers in Systems Neuroscience*, *4*, 13–19.

Yarkoni, T., Barch, D. M., Gray, J. R., Conturo, T. E., & Braver, T. S. (2009). BOLD correlates of trial-by-trial reaction time variability in gray and white matter: A multi-study fmri analysis. *PLoS One*, *4*, e4257.

Yamamoto, K., & Sakai, K. L. (2017). Differential signatures of second language syntactic performance and age on the structural properties of the left dorsal pathway. *Frontiers in Psychology*, *8*, 829.

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*, 1818.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.