RESEARCH ARTICLE

WILEY

Hierarchical syntactic processing is beyond mere associating: Functional magnetic resonance imaging evidence from a novel artificial grammar

Luyao Chen[1,2] | Tomás Goucha[2] | Claudia Männel[2,3] | Angela D. Friederici[2] | Emiliano Zaccarella[2]

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

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

[3]Department of Audiology and Phoniatrics, Charité – Universitätsmedizin Berlin, Berlin, Germany

Correspondence
Luyao Chen, Emiliano Zaccarella, and Angela D. Friederici, Department of Neuropsychology, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany.
Email: luyaochen@cbs.mpg.de (L. C.), zaccarella@cbs.mpg.de (E. Z.), and friederici@cbs.mpg.de (A. D. F.)

Abstract

Grammar is central to any natural language. In the past decades, the artificial grammar of the A^nB^n type in which a pair of associated elements can be nested in the other pair was considered as a desirable model to mimic human language syntax without semantic interference. However, such a grammar relies on mere associating mechanisms, thus insufficient to reflect the hierarchical nature of human syntax. Here, we test how the brain imposes syntactic hierarchies according to the category relations on linearized sequences by designing a novel artificial "Hierarchical syntactic structure-building Grammar" (HG), and compare this to the A^nB^n grammar as a "Nested associating Grammar" (NG) based on multilevel associations. Thirty-six healthy German native speakers were randomly assigned to one of the two grammars. Both groups performed a grammaticality judgment task on auditorily presented word sequences generated by the corresponding grammar in the scanner after a successful explicit behavioral learning session. Compared to the NG group, we found that the HG group showed a (a) significantly higher involvement of Brodmann area (BA) 44 in Broca's area and the posterior superior temporal gyrus (pSTG); and (b) qualitatively distinct connectivity between the two regions. Thus, the present study demonstrates that the build-up process of syntactic hierarchies on the basis of category relations critically relies on a distinctive left-hemispheric syntactic network involving BA 44 and pSTG. This indicates that our novel artificial grammar can constitute a suitable experimental tool to investigate syntax-specific processes in the human brain.

KEYWORDS
artificial grammar, associative processing, fMRI, hierarchical syntactic processing, language network, syntactic category

1 | INTRODUCTION

A hallmark of the human language faculty is the capacity to combine linear sequences of words into hierarchical structures for language comprehension (Dehaene, Meyniel, Wacongne, Wang, & Pallier, 2015; Ding, Melloni, Hang, Xing, & Poeppel, 2016; Everaert, Huybregts, Chomsky, Berwick, & Bolhuis, 2015; Fitch, 2014; Friederici, 2017; Hagoort, 2005; Nelson et al., 2017; Sheng

et al., 2019). The hierarchical relationships among the constituents (i.e., words and phrases) are based on their syntactic categories (e.g., nouns and verbs) and their syntactic relations, which determine how these constituents can be recursively combined to build up linguistic hierarchical syntactic structures (Adger, 2013; Berwick et al., 2013; Chomsky, 1995; Dehaene et al., 2015; Epstein, Kitahara, & Seely, 2014; Hauser, Chomsky, & Fitch, 2002; Hornstein & Nunes, 2008; Hornstein & Pietroski, 2009; Hornstein, 2009; Hoshi, 2018, 2019; Lenneberg, 1967; Miyagawa, Berwick, & Okanoya, 2013; Narita, 2014). For example, the determiner "[D the]" and the noun "[N apple]" together form the determiner phrase "[DP the apple]," which can further combine with the verb "[V eat]" to form the verbal phrase "[VP eat the apple]" as a larger constituent (see Figure 1a). The capacity to create hierarchical structures out of syntactic category relations has been deemed fundamental for language processing as a unique human language trait (Chen et al., 2019; Fujita, 2014; Berwick et al., 2013; Friederici, 2019; Goucha, Zaccarella, & Friederici, 2017; Miyagawa et al., 2013). In the current study, we approach the fundamental question of how the human brain imposes hierarchical syntactic structures on linearized word sequences (i.e., hierarchical syntactic processing) on the basis of the syntactic category relations by using an artificial grammar learning/processing paradigm.

Language processing in the human brain is known to be supported by a left-dominant fronto-temporal network (Berwick et al., 2013; Fedorenko, Hsieh, Nieto-Castañon, Whitfield-Gabrieli, &

Kanwisher, 2010; Friederici, 2017; Hagoort & Indefrey, 2014; Vigneau et al., 2006). A recent meta-analytic approach involving over 200 participants has shown that the processing of natural sentences compared to nonword lists involves the left inferior frontal gyrus (IFG), the inferior frontal pars orbitalis (IFGorb), the anterior temporal lobe (aTL), the posterior temporal lobe (pTL), and the angular gyrus (AG; see https://evlab.mit.edu/funcloc/download-parcels; Fedorenko et al., 2010).

Within this language network, the IFG, especially Brodmann area (BA) 44, and the pTL, especially the posterior superior temporal gyrus (pSTG), have been consistently shown to specifically support syntactic processing, regardless of stimulus complexity (e.g., Bornkessel, Zysset, Friederici, von Cramon, & Schlesewsky, 2005; Brennan & Pylkkänen, 2012; Chang, Dehaene, Wu., Kuo, & Pallier, 2020; den Ouden et al., 2012; Friederici, 2019; Friederici, Makuuchi & Bahlmann, 2009; Matchin, Brodbeck, Hammerly, & Lau, 2019; Matchin, Hammerly, & Lau, 2017; Jobard, Vigneau, Mazoyer, & Tzourio-Mazoyer, 2007; Just, Carpenter, Keller, Eddy, & Thulborn, 1996; Kinno, Kawamura, Shioda, & Sakai, 2008; Makuuchi, Bahlmann, Anwander, & Friederici, 2009; Pallier, Devauchelle, & Dehaene, 2011; Santi & Grodzinsky, 2012; Snijders et al., 2009; Zaccarella & Friederici, 2015; Zaccarella, Meyer, Makuuchi, & Friederici, 2017). Syntactic processing seems to increase the functional coupling between the two regions during the processing of both basic phrases and more complex sentences (den Ouden et al., 2012; Wu, Zaccarella, & Friederici, 2019). Studies using morphosyntactic
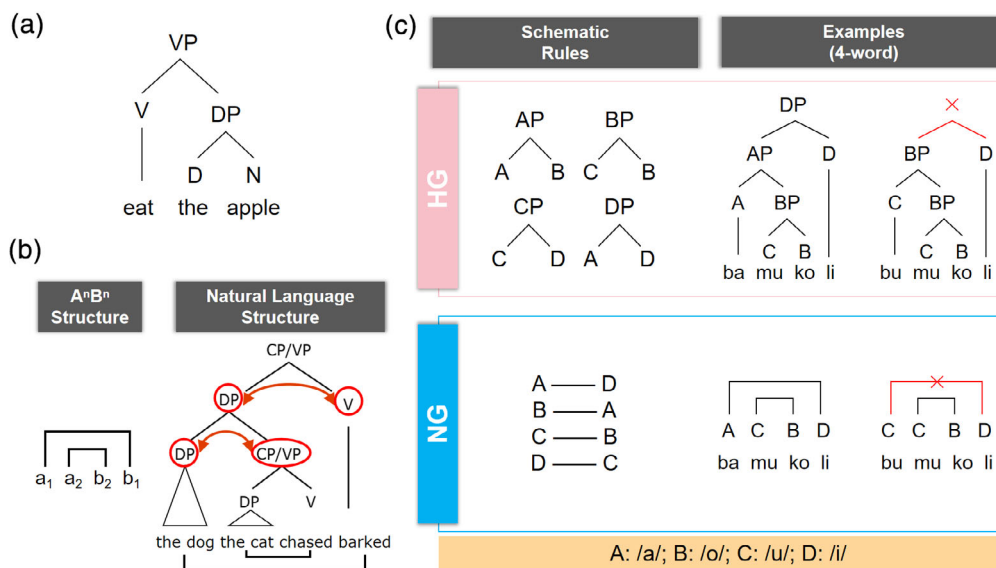

FIGURE 1 (a) Tree diagram of the hierarchical syntactic structure of "eat the apple," and its underlying structure-building rules. (b) Superficial resemblance and hierarchical differences between the $A^n B^n$ structures and the natural language structures. Here, we do not differentiate the complementizer phrase (CP) with the verbal phrase (VP) for a parsimonious expression of the syntactic nodes. Red circles were used to highlight the critical syntactic nodes, and the double-headed arrows indicated the syntactic dependencies between the syntactic nodes. Triangles under the determiner phrase (DPs) in the natural language structure were used to simplify the expression of the internal structures in this example. The lines under the natural language structure superficially correspond to the dependencies of the $A^n B^n$ structure. (c) The schematic rules with the vowels for each category (in the bottom yellow box) were presented with grammatical and ungrammatical examples. Upper panel: structure-building rules and examples for the Hierarchical syntactic structure-building Grammar (HG). Lower panel: Associative rules and examples for the Nested associating Grammar (NG)

cues to syntactic categories, like inflectional suffixes and case marking, further confirmed the involvement of Broca's area during hierarchical syntactic processing (Ohta, Fukui, & Sakai, 2013), in which BA 44 is specifically sensitive to the syntactic categories (Goucha & Friederici, 2015). Taken together, the reported studies suggest that BA 44 and the pSTG might be specifically involved in building syntactic hierarchies during natural language processing, according to the syntactic relationships between word categories.

In addition to studies focusing on natural language, a second approach to test syntactic processing in the brain is to use artificial-grammar paradigms. Compared to natural language, artificial grammars have two additional advantages: (a) they remove semantics as a confounding factor, and (b) they ensure cross-species comparability for the understanding of language evolution (e.g., Gómez & Gerken, 2000; Misyak, Christiansen, & Tomblin, 2010; Morgan-Short, 2020; Petersson, Folia, & Hagoort, 2012; Petkov & ten Cate, 2019; Tagarelli, Shattuck, Turkeltaub, & Ullman, 2019; Uddén & Männel, 2018; Wilson et al., 2018).

Regarding (a), regions within the language network that have been frequently reported to be related to semantic processing include the IFGorb, aTL, and AG (Dapretto & Bookheimer, 1999; Davey et al., 2015; Hagoort & Indefrey, 2014; Hartwigsen et al., 2016; Jung & Lambon Ralph, 2016; Lau, Phillips, & Poeppel, 2008; Matchin et al., 2017; Pylkkännen, 2019; Vigneau et al., 2006; Whitney, Kirk, O'Sullivan, Lambon Ralph, & Jefferies, 2012). Although the middle frontal gyrus (MFG) (coarsely corresponding the dorsal BA 6) was also included in Fedorenko's language atlas, Fedorenko, Behr, and Kanwisher (2011) found that this region responded to additional non-language cognitive tasks, and thus they suggested that it might not be specific to language processing. Moreover, BA 6 has not systematically been found across language tasks (e.g., Chang et al., 2020; Matchin et al., 2017; Pallier et al., 2011). Rather, MFG, a broader prefrontal region also covering the dorsal BA 6, was reported together with the inferior parietal lobule (ParInf) as core regions for cognitive control (Uddin, Yeo, & Spreng, 2019). We thus differentiate a network comprising BA 44 and pSTG as supporting hierarchical syntactic processing, and a network formed by the IFGorb, aTL, and AG as supporting semantic processing (see Section 2).

As for (b), artificial grammars mimicking human language syntax have been extensively used in the last decades to test the language faculty in cross-species comparison between humans and nonhuman primates (Fitch & Hauser, 2004; see a recent review of Petkov & ten Cate, 2019). Previous studies have shown that humans are sensitive to violations in "a_1 a_2 b_2 b_1" sequences produced by the "A^nB^n" grammar (co-indexed elements signal an association), in which "a_2–b_2" is *nested* within the center-embedding pair of "a_1–b_1," forming multilevel associations (Bahlmann, Schubotz, & Friederici, 2008; Bahlmann, Schubotz, Mueller, Koester, & Friederici, 2009; Fitch & Hauser, 2004; Friederici, Bahlmann, Heim, Schubotz, & Anwander, 2006), and that even infants as young as 5 months of age are able to encode relations across nested dependencies, as indicated by brain responses to violations of multilevel associations (Winkler, Mueller, Friederici, & Männel, 2018). However, a cross-species comparison revealed that

nonhuman primates were only able to learn and process sequences based on local transition probabilities of the types such as $(AB)^n$, while they failed to process the more complex A^nB^n structures (Fitch & Hauser, 2004). This has been taken as supporting evidence that the human language faculty goes beyond nonhuman animals' sequence processing capacities. However, it can be argued that the limited time for implicit exposure, the intensity of feeding treats during the re-familiarization phase, the reliability of "head-turning" observation technique (see Fitch & Hauser, 2004), and even the animals' "willingness" or "awareness" of using basic nonrule strategies might have influenced the results (e.g., Rey, Perruchet, & Fagot, 2012; Uddén et al., 2009). A recent, better controlled study has shown that nonhuman primates (rhesus monkeys) are capable of forming multilevel representations of nested A^nB^n structures during a nonlinguistic sequence generation task (Ferrigno, Cheyette, Piantadosi, & Cantlon, 2020). The crucial point, however, is that such nested multilevel associations of the type $a_2a_1b_1b_2$ only superficially resemble hierarchical relations found in human language, since the innermost pair ("a_1b_1") does not depend on the outermost pair ("a_2b_2"). In language, the innermost clause conversely always depends on the outermost clause. As shown in Figure 1b, in the artificial multilevel associating sequence "$a_2a_1b_1b_2$," there is no dependency between "a_1b_1" and "a_2b_2." In contrast, in the natural sentence, "that the cat chased" comprises an additional syntactic node (i.e., complementizer phrase [CP] or a verbal phrase [VP]), which modifies "the dog" (a determiner phrase [DP]) in the main clause. "The dog that the cat chased" is a complex DP, which can be further combined with the main verb "barked" to generate a higher syntactic node (CP/VP). Therefore, both types of sequences have distinct inner structures. The structural difference between the a^nb^n grammar and the natural language hierarchical grammar might represent a possible explanation why nested multilevel associations could be successfully processed by certain nonhuman animals (Abe & Watanabe, 2011; Ferrigno et al., 2020; Jiang et al., 2018; Rey et al., 2012; Stobbe, Westphal-Fitch, Aust, & Fitch, 2012). In this sense, the nested multilevel associations are also claimed to be solved by animals' nonrule strategies such as counting and repetition/symmetry detection (Beckers, Bolhuis, Okanoya, & Berwick, 2012; Berwick et al., 2013; de Vries, Monaghan, Knecht, & Zwitserlood, 2008; Fitch & Friederici, 2012; Friederici, 2018; Petkov & ten Cate, 2019; Stobbe et al., 2012).

Overall, it seems that previous artificial grammars containing nested multilevel associations, such as the A^nB^n grammar, do not fully capture human syntax, and can be solved using associative strategies that render the specificity of the human language faculty underestimated. This motivated the present study to take the lead in designing a novel artificial grammar which guarantees hierarchical syntactic processing on the basis of syntactic category relations in the human brain, and this grammar thus serves as a more ecological model approaching natural human syntactic processing when compared with previous artificial grammars.

In light of the considerations above, we aim to go beyond previous research to verify the involvement of syntax-related cortical areas for hierarchical syntactic processing on the basis of category relations

and compare these directly to simple associative processing at work for nested multilevel associations. To this end, we conducted a functional magnetic resonance imaging (fMRI) experiment to test whether brain regions within the left-hemispheric syntactic network (BA 44 and pSTG), would differently respond to hierarchical syntactic processing and multilevel associative processing. In order to exclude semantic interference, two artificial grammars were designed corresponding to the two processing mechanisms. First, we created a "Hierarchical syntactic structure-building Grammar" (HG) which relies on four structure-building rules applying to sets of functionally distinct categories (Figure 1c upper panel). Participants were explicitly instructed to identify the category relations between two constituents, such as "[X x]" and "[Y y]," and to combine them as a *new* phrase, "[XP x y]" (capitalized letters denote the syntactic categories). This is comparable to the natural language example mentioned earlier, in which the determiner "[D the]" and the noun "[N apple]" can be combined together to form the DP "[DP the apple]." To clarify, we conventionally notated this phrase as "XP" here for convenience of expression, but it symbolizes the abstract category "X" in HG. Both "[XP x y]" and "[X x]" belong to the same category "X," sharing the same syntactic features and functions. Second, we developed a "Nested associating Grammar" (NG) according to the previous $A^n B^n$ grammar, in which multiple associations are *nested* to form multilevel associations. This grammar is composed of four association rules (for each rule, two specific categories are associated, such as "X–Y" and "V–W"), and each pair of associations can be nested in between another one, such as "V X Y W" (Figure 1c lower panel). Two groups of participants learned these grammars separately, and the successful learners underwent an fMRI scanning session, during which they had to judge the grammaticality of the sequences according to the grammars they had learned. Thus, our fMRI study focused on the processing phase, during which participants applied the rules they had already learnt to solve the grammaticality judgment task in the scanner. As to the results, we expected to find a differential involvement of the left BA 44 and pSTG during the application of the structure-building rules for hierarchical syntactic processing, relative to the associative processing of the nested multilevel associations. As a control test, we expected to find no activation within the semantic network (IFGorb, aTL, and AG) for either grammar.

2 | METHODS

2.1 | Participants

Thirty-six participants were randomly allocated to two groups (HG: N = 18, including 11 females and 7 males, age: M = 27.89, SD = 4.07; NG: N = 18, including 10 females and 8 males, age: M = 27.44, SD = 3.03). All participants were right-handed German native speakers with normal or corrected-to-normal vision and audition, reporting no history of psychiatric or neurological diseases. Moreover, they were nonmusicians and not early bilinguals. All gave written informed consent before the experiment, which had been approved by the local ethics committee (078/19-ek), and received around 20 Euros for

participation. Note that these participants were recruited from a pool of participants who successfully acquired the rules in the learning session, and the data of participants with excessive motion artifacts (one in each group) were discarded after the scanning session (see Section 2.3 for details). To reliably confirm the homogeneity between groups, a battery of cognitive tests was conducted, and the results showed no between-group differences (see Supporting Information 1).

2.2 | Materials

For each grammar, we created nonsense monosyllables (of a consonant-vowel structure) to compose the artificial language vocabulary. Here, we assigned artificial words sharing the same vowel to the same category (Bahlmann et al., 2008; Friederici et al., 2006; Opitz & Friederici, 2003, 2004, 2007), with four German vowel categories, /a/, /o/, /u/, and /i/, and 12 tokens each (e.g., /ba/and/bo/). All word tokens within one category were distributed across two sessions: a behavioral learning session and an fMRI scanning session (see Section 2.3), counterbalancing their phonological similarity to real German words and the real words' frequency (see Supporting Information 2.1 for details). The counterbalance of word tokens between the two sessions was to avoid similarity transfer effect, that is, participants process the stimuli in the scanning session simply based on material similarities to the stimuli in the learning session (see also Opitz & Hofmann, 2015). Therefore, no superficial exemplar overlaps between the two sessions could be detected. All monosyllabic word tokens were recorded in 32-bit stereo at a sampling rate of 44,100 Hz by a female professional German native broadcaster. The audio files were normalized to 70 dB and 200 Hz by using Praat (http://www.praat.org).

The schematic rules for each grammar are illustrated in Figure 1c. We chose four-word sequences as the stimuli for the scanning session because these sequences in the current study maximally contained one center-embedded dependency for both grammars.

We additionally constructed ungrammatical structures to evaluate participants' grammar performance via grammaticality judgments, thus ensuring participants' engagement in stimuli processing. For HG, violations implied that the constituents in a sequence could not be combined. For example, in Figure 1c, "C C B D" is ungrammatical because "C C B" can be combined as a phrase belonging to Category B by applying the structure-building rule, but "B D" cannot be combined due to lacking such an underlying rule. For NG, violations referred to the occurrence of an unexpected element, unpredicted by the underlying dependency. For example, in Figure 1c, "C" could not go with "D," resulting in a nonadjacent dependency violation. Each grammar contained the same number of grammatical and ungrammatical sequences. Furthermore, for both grammars, positions and frequencies of each category and the distribution of word-categorical bi- and trigrams (i.e., combinations of two- and/or three-word categories) were carefully controlled (see Supporting Information 2.2). This was to ensure that participants could not develop alternative strategies to distinguish grammatical from ungrammatical stimuli.

For each participant, a new stimulus list was created through randomly substituting the structures with the word tokens of the different categories, and auditorily presented (Presentation 17.0; https://www.neurobs.com).

2.3 | Procedures

An explicit learning session was adopted to help participants grasp the rules for solving the grammaticality judgment task in the scanner for each grammar. Therefore, the current study focused on how participants applied these rules, once they successfully passed the learning session, to process the structures in the scanning session (see also Bahlmann et al., 2008, 2009; Jeon & Friederici, 2013; Martins, Bianco, Sammler, & Villringer, 2019; Ohta et al., 2013).

2.3.1 | Learning session

Participants underwent an explicit behavioral learning session to learn and consolidate their rule knowledge around 1 hr before subsequent stimuli processing in the scanner (see Figure 2).

First, participants were explicitly instructed about the grammar rules with examples on the screen and given 2 min to memorize all the rules. Then they entered the "single rule learning" phase, in which each rule would be specifically trained with six 2-word trials. After learning the four rules separately, participants moved on to the "mixed rule learning" phase, which contained all the rules, with 16 2-word trials. The single and mixed rule learning phases would repeat four times. To help the HG group grasp the structure-building rules, participants were asked to judge whether the two-word combinations could be combined as a given category, that is, a "category judgment" task. For instance, participants would hear a string of "ba ko a" and were instructed to judge whether "ba ko" could form an "/a/ phrase." This task was set to strengthen participants' knowledge of transforming the word combinations into abstract upper-level categories (i.e., represented as the categorical vowels). Both groups went through a four-word training phase, which was composed of 24 blocks with 10 trials per block. All the phases and each block within a phase contained 50% ungrammatical trials.

The learning session was designed to terminate when participants achieved our *a priori* defined performance criterion of two successive blocks reaching 90% accuracy during the four-word training phase. Setting a behavioral criterion aims to balance the behavioral performances under respective conditions before scanning (see also
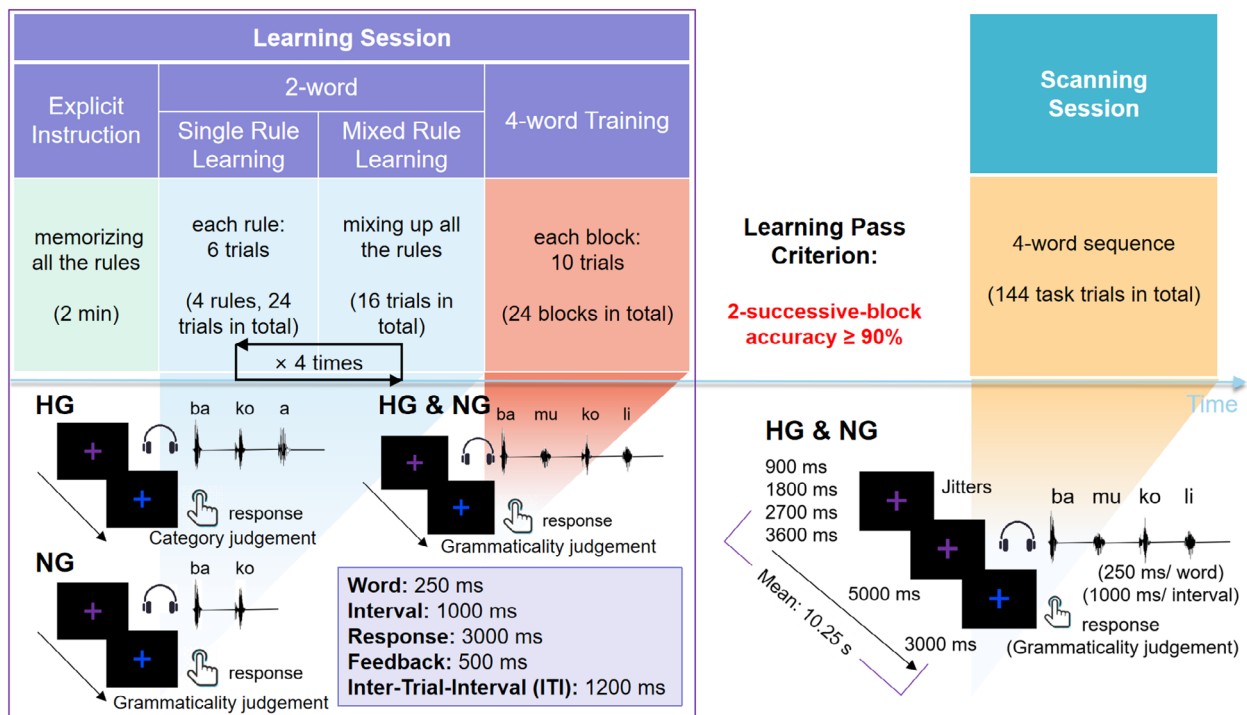

FIGURE 2 Procedure of learning and scanning sessions. For the learning session (within the left purple frame), the upper panel shows the learning procedures of the rules of each grammar. Note that the task for HG was "category judgment" during single and mixed rule learning phases, and participants underwent two- and eight-word training phases (8 and 24 blocks, respectively, with each block having 10 trials) to get used to the grammaticality judgment task before the four-word training (not shown in the figure). The lower panel shows the trial presentation according to the parameters listed at the bottom right. After successfully achieving the behavioral criterion of the four-word training phase, participants would undergo the scanning session (see schematic illustration of the fMRI task [on the right]). HG, Hierarchical syntactic structure-building Grammar; NG, Nested associating Grammar

Bahlmann et al., 2008, 2009). According to this criterion, 19 out of 33 participants passed the HG learning session, while 19 out of 20 participants succeeded in NG learning. The original sample size for each group was set to achieve the between-group behavioral balance as mentioned above. The higher drop-out rate for the HG is due to this grammar's inherent structural complexity. Crucially, the same outcome was also reported in Martins et al. (2019), who had a higher drop-out rate (50%) for learning the most complex grammar with a criterion that was much looser (accuracy of the last learning session [containing 20 trials] >80%) than that of our current study. In line with their study, we set the drop-out learners aside and only included those HG learners who could successfully apply the HG rules for the corresponding task after a relatively short learning time (~1 hr), similar to the successful NG learners. Therefore, the comparison between the two groups of successful learners should be reliable.

For each trial of the learning session, participants saw a purple fixation at the screen center, meanwhile hearing a sequence of the stimuli. Each word or category lasted 250 ms, followed by a 1,000 ms silence. There was a 1,000 ms interval between the last word (or category) and the response cue (a blue fixation). Therefore, only after presentation of the whole sequence could participants deliver a response, rendering the reaction time unreliable in the delayed response mode (see also Bahlmann et al., 2008 for the "delayed-response mode"). Participants had 3,000 ms to respond, and they immediately received visual feedback (duration: 500 ms). Response buttons to right-hand index and middle finger assignments were counterbalanced across participants. Between the trials in the learning session, there was an intertrial interval (ITI) of 1,200 ms. Moreover, accuracy was shown to the participants for each block within each training phase and for single/mixed rule learning phase (1,000 ms). The whole learning session took about 1 hr.

2.3.2 | Scanning session

Participants who reached the behavioral learning criterion were invited to the fMRI scanning session (see Figure 2). The stimuli contained 72 grammatical and 72 ungrammatical four-word sequences, and the timing of stimulus presentation was identical to the four-word training phase in the learning session. Additionally, 48 short-break trials of the same duration as the task stimuli were randomly inserted to estimate the hemodynamic response more efficiently. The ITI was jittered to 900; 1,800, 2,700; or 3,600 ms. The mean duration of each trial (including ITI) was around 10.25 s.

An in-house algorithm was applied for pseudo-randomizing the trials to achieve optimal efficiency of fMRI signal acquisition, separately for each participant. In between every 64 trials (3 sets in total, and each set was composed of 24 grammatical and 24 ungrammatical task trials, and 16 short break trials), there was a 20-s break to further help the hemodynamic response return to baseline. All break trials included a fixation at the center of the screen. In contrast to the behavioral learning session, no feedback was provided after the grammaticality judgment.

2.4 | Behavioral data analyses

For the analysis of the behavioral data acquired in the scanner, we applied a two-way (grammar group × grammaticality) repeated-measures analysis of variance. We used accuracy as the most consistent index of behavioral performance to assess participants' grammar processing (see also Bahlmann et al., 2008, 2009).

2.5 | Imaging data acquisition

Functional imaging data were acquired via a 3.0-Tesla Siemens PRISMA magnetic resonance scanner (Siemens AG, Erlangen, Germany) using a 32-radiofrequency-channel head coil. A T2*-weighted gradient echo planar imaging (EPI) sequence was adopted for acquiring the functional magnetic resonance images with the following parameters: repetition time (TR) = 2,000 ms; echo time = 23.6 ms; flip angle = 80°; field of view = 204×204 mm^2; matrix size = 102×102 mm^2; in-plane resolution = 2×2 mm^2; slice thickness = 2 mm; number of slice = 60; gap = 0.26 mm; alignment to AC–PC plane. Signals from different slices were acquired by the multiband scanning technique (MB = 3) to efficiently minimize slice-timing effects. High-resolution anatomical T1-weighted images for co-registration were selected from the institute's database for corresponding participants.

2.6 | Imaging data preprocessing

Functional imaging data preprocessing was performed using statistical parametric mapping (SPM 12; Welcome Department of Cognitive Neurology, London, UK) implemented in the environment of MATLAB R2017b, including: (a) slice time correction with individual slicing-time series for each participant to minimize acquisition differences among slices; (b) spatial realignment of images to the first set of volumes acquired from the three bands at the very beginning for each individual to correct for head motion; (c) co-registration from the anatomical images to the functional images for each participant; (d) normalization of the images to the EPI template based on Montreal Neurological Institute (MNI) stereotactic space to minimize cerebral differences between participants, and resampled into $2 \times 2 \times 2$ mm^3 voxels; and (e) smoothing the images with a 3D Gaussian kernel with full-width at half-maximum of 5 mm.

2.7 | Language network-level statistical analyses

Whole brain-level statistical analysis commonly contains two procedures—the first-level analysis for individual brain activation and the second-level analysis for between-group activation comparisons (Huettel, Song, & McCarthy, 2004). At the first level, a general linear model (GLM) for each participant within each group was set up by taking the grammatical and ungrammatical conditions as two regressors of interest, with each task trial onset and its stimuli duration

(~5 s) of the conditions modulated as a boxcar convolved with a canonical hemodynamic response function. Six head motion parameters were also recruited for the model as covariates of no interest. The data were further high-pass filtered at 128 Hz to eliminate low-frequency drifts.

At the second level of group comparison, we calculated main effects of the two factors, grammar group and grammaticality, and their interaction as *t* tests. Since we are particularly interested in how the subregions within the left language network support hierarchical syntactic processing through manipulating the category-relation-based structure-building rules when compared to the nested multi-level association processing, we restricted our second-level analysis to the syntactic network composed of the IFG and pTL from the functional left-hemispheric language atlas (see Figure 3a). We also took the semantic network, including the IFGorb, aTL, and AG, from the language atlas as a control. Therefore, both the syntactic and semantic networks were adopted as masks for small volume correction separately for the between-group contrasts ($p < .05$, family-wise error [FWE] corrected). We were further interested in the: (a) contrast of "both groups > implicit baseline" for identifying the regions involved in artificial grammar processing and (b) contrast between the HG and the NG groups to inspect the processing mechanism-specific activation patterns. To note, we included "accuracy" as a covariate of no interest to control for potential processing difficulties.

Given that Broca's area contains BA 44 and BA 45, and these are reported to have functional (e.g., Fedorenko et al., 2011;

Friederici, 2011, 2017; Goucha & Friederici, 2015; Grodzinsky & Friederici, 2006; Hagoort, 2005; Hagoort & Indefrey, 2014; Matchin et al., 2017; Pallier et al., 2011; Tyler et al., 2010; Vigneau et al., 2006), macroanatomical (Amunts et al., 1999) and micro-receptoarchitectonical (Amunts et al., 2010) differences, we further analyzed Broca's area independently by using the anatomical BA 44 and BA 45 masks for contrasting HG and NG.

A whole-brain level analysis under the contrast of "HG > NG" can be found in the Supporting Information 4.4. Additional analyses within the cognitive control network during artificial language processing (see Uddin et al., 2019) can also be found in the Supporting Information 4.5.

2.8 | Signal intensity analysis

The extraction of signal intensities for each grammar in the different regions can provide additional information about subregional activation specificities (e.g., BA 44 vs. BA 45) and directionality of activation (activation vs. deactivation). Regions of interest (ROIs) within the functional language network were thus defined. Peak activity coordinates were localized running one sample *t* tests (regressing out the accuracy), independently for each group, under the contrast of "grammar > implicit baseline" ($p < .05$, FWE-corrected). Centered around these peak coordinates, spherical ROIs with 4 mm radius were defined. It is noteworthy that as mentioned before, we further
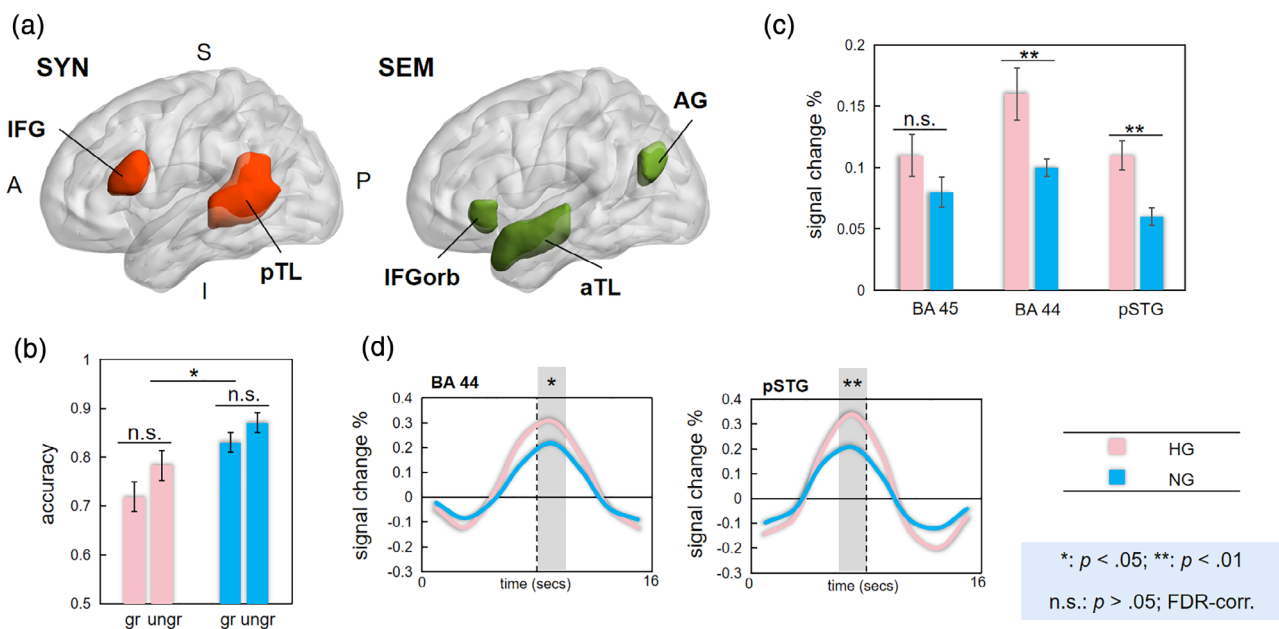

FIGURE 3 (a) Syntactic network (SYN) and semantic network (SEM) from the 220 participant-based functional left-hemisphere language atlas (extended by Fedorenko's lab on the basis of Fedorenko et al., 2010; https://evlab.mit.edu/funcloc/download-parcels). (b) Behavioral results in the scanning session. (c) Signal intensity analysis results. (d) Peak analysis results, in which the gray bar stands for one TR. For subplots (b) and (c), error bar shows the *SEM*. A, anterior; AG, angular gyrus; aTL, anterior temporal lobe; BA, Brodmann area; gr, grammatical condition; HG, Hierarchical syntactic structure-building Grammar; I, inferior; IFG, inferior frontal gyrus; IFGorb, inferior frontal pars orbitalis; NG, Nested associating Grammar; P, posterior; pSTG, posterior superior temporal gyrus; pTL, posterior temporal lobe; S, superior; ungr, ungrammatical condition

adopted two separate anatomical masks for BA 44 and BA 45 (Amunts et al., 1999) to investigate IFG at a finer-grained resolution (see also Martins et al., 2019; Matchin et al., 2017).

Time-series data were extracted from these ROIs, and, within each ROI, percentage signal change was calculated via MarsBaR 0.44 (https://sourceforge.net/projrcts/marsbar/) as signal intensity for each grammar. Two-sample *t* tests were performed to determine whether the signal intensity of HG was different from that of NG within each ROI. The *p*-values of multiple comparisons were "false discovery rate" (FDR)-corrected.

2.9 | Peak intensity analysis

To further inspect the main effect of group within the ROIs which might show significant between-group signal intensity differences, we performed complementary peak intensity analyses (see also Makuuchi, Grodzinsky, Amunts, Santi, & Friederici, 2013). The hemodynamic response was assessed on the basis of the trials per subject for each group within each ROI. The trial time course of these ROIs was constructed by upsampling the signal intensities from the onset of each task trial to the subsequent eighth TR (i.e., 16 s and eight sampling data points in total for each trial). In this way, for each task trial, a hemodynamic response curve and corresponding peak could be estimated. Within each ROI, we then extracted for each participant the signal intensity values at the time-point peak we had retrieved at group-level for each condition. Finally, a two-sample *t* test was performed to compare the peak intensity differences, with a threshold of *p* < .05, FDR-corrected. Both the signal intensity and the peak intensity analyses can provide confirmatory evidence for the robustness of the activation differences detected in the language network-level analysis.

2.10 | Effective connectivity modeling

Given that signal intensity-related analyses could not answer when and how the information was transferred from one ROI to another (e.g., from BA 44 to pSTG), effective connectivity modeling was performed for exploring potential connectivity differences between the two grammar groups.

We employed an "extended unified structural equation modeling" (euSEM) approach, which provides an exploratory and hypothesis-free method to estimate the effective connections between different cortical regions (Gates, Molenaar, Hillary, & Slobounov, 2011). The euSEM also unifies: (a) the lagged effect: a longitudinal effect from a previous time point to the current state, represented in a multivariate autoregressive model; (b) the contemporaneous effect: an effect at the same time point, involving conventional SEM (Gates et al., 2011; Kim, Zhu, Chang, Bentler, & Ernst, 2007). The euSEM was implemented by Group Iterative Multiple Model Estimation (GIMME; Gates et al., 2011; Gates & Molenaar, 2012), a recently developed toolkit outperforming most competing connectivity approaches in large-scale

simulations (Gates & Molenaar, 2012), identifying comparatively reliable connections with respect to conventional effective connectivity analyses (Beltz & Gates, 2017). GIMME has recently been used successfully in language acquisition (Yang, Gates, Molenaar, & Li, 2015) and bilingual language switching studies (Wu et al., 2019).

The time-series data of the ROIs are used in GIMME for effective connectivity modeling. Note that for the euSEM model estimation, each task trial was indexed as "1" to signal the existence of experimental task effect, whereas each baseline trial was indexed as "0." The euSEM was implemented by GIMME with the following steps (Beltz & Gates, 2017; Gates & Molenaar, 2012;Wu, Yang, et al., 2019; Yang et al., 2015):

1. Empty null network models were set up for all the participants.
2. A group-level model was first constructed via Lagrange multiplier tests, and, if a connection improved the model fit significantly (≥75% of the sample), it was added to the model for re-estimation.
3. Such a search-and-add procedure would terminate if no connection improved the model fit significantly with the nonsignificant connections pruned, and then the individual-level network for each participant would be estimated in a similar manner.
4. Finally, a confirmatory model after trimming was fitted. Note that the reliable model fit parameters could be mainly determined by two *a priori* criteria: the comparative fit index (CFI) ≥ 0.90, and the non-normed fit index (NNFI) ≥ 0.90 (Wu, Yang, et al., 2019; Yang et al., 2015).

3 | RESULTS

The results reported below are all based on the data of participants who successfully completed both learning and scanning sessions, with each group having 18 participants (see Section 2.1).

3.1 | Behavioral results

3.1.1 | Learning results

The two grammar groups underwent a comparable number of four-word training blocks to reach the learning criterion of the learning session (HG: *M* = 12.89, *SD* = 7.15; NG: *M* = 8.39, *SD* = 7.17; *t* [34] = 1.89, *p* > .05). Together with the nonsignificant between-group differences in the cognitive test results (see Supporting Information 1.2), these data suggested no group-difference before the fMRI scanning session.

3.1.2 | Behavioral results in the scanning session

For the grammaticality judgment task, the accuracy of the HG group was lower than that of the NG group (HG: grammatical: *M* = .72, *SD* = .15; ungrammatical: *M* = .78, *SD* = .09; NG: grammatical: *M* = .83,

SD = .14; ungrammatical: M = .88, SD = .09; main effect of group: F (1,34) = 8.79, p < .01, η_p^2 = .21). The observation that accuracy rates for both grammatical and ungrammatical conditions were significantly above the chance level (.50) confirmed that both groups were compliant with the tasks. The main effect of "grammaticality" was also significant. For ungrammatical sequences participants responded with higher accuracy than for grammatical sequences (F(1,34) = 7.81, p < .01, η_p^2 = .19). Yet, the interaction of "grammar group × grammaticality" was nonsignificant (F(1,34) = .20, p = .656) (see Figure 3b) (more behavioral information can be found in Supporting Information 3).

3.2 | fMRI results

3.2.1 | Language network-level analysis results

The main effect of "Grammaticality" and the interaction were not significant within either network (see Supporting Information 4.1 for language network-level results). The contrast of both grammars against baseline revealed activity increase in the IFG and pTL within the syntactic network (see Table 1). Note that under the contrast of "implicit baseline > both groups," the IFGorb, aTL, and AG within the semantic network were activated (see Supporting Information 4.2 for details).

Within the syntactic network, compared to the NG group, the results of the HG group revealed higher activation in both the IFG and pTL (see Table 1; see Supporting Information 4.3 for a more sophisticated GLM at the first-level analysis which further provided confirmatory results for the current activation profiles: only the main effect of grammar group was significant). Furthermore, results of the additional analyses within Broca's area are also presented in Table 1. To note, no voxel survived under "NG > HG" even at a liberal threshold (p < .005, uncorrected). Besides, as expected, no activation was detectable within the semantic network for the contrast of "HG > NG" or vice versa (p < .005, uncorrected).

The whole-brain level analysis results, as well as those for the cognitive control network including the executive-function areas can be found in the Supporting Information 4.4 and 4.5.

3.2.2 | Signal intensity analysis results

The peak MNI coordinates identified by masking the one-sample t test results with the whole language area atlas for each grammar group are as follows: (a) HG group: BA 44 (−46, 8, 22), BA 45 (−48, 28, 18), pTL (esp., pSTG) (−56, −36, 10) and (b) NG group: BA 44 (−52, 6, 24), BA 45 (−46, 36, 16), pTL (esp., pSTG) (−44, −34, 6).

Signal intensity analyses showed significant between-group differences in both BA 44 (HG: M = .16, SD = .09; NG: M = .10, SD = .03; t(34) = 2.85, p < .01, d = .89, FDR-corrected), and the pSTG (HG: M = .11, SD = .05; NG: M = .06, SD = .03; t(34) = 3.15, p < .01, d = 1.21; FDR-corrected), whereas BA 45 (HG: M = .11, SD = .07; NG: M = .08, SD = .05) did not exhibited significant group differences (t (34) = 1.58, p > .05, FDR-corrected) (see Figure 3c). Therefore, BA 44 and the pSTG were recruited for further peak analyses separately. (For each ROI, Pearson's correlation tests between signal change and behavioral indexes [accuracy and reaction time] were also performed [all tests: n.s., suggesting neural signal might be independent of behavioral performance; see Supporting Information 5.1 for details].)

3.2.3 | Peak intensity analysis results

For each group, hemodynamic response reached the peak at the fifth TR in BA 44 and at the fourth TR in the pSTG. The trial-based peak analyses confirmed that both BA 44 and the pSTG were significantly more highly activated for the HG group than the NG group (BA 44: HG: M = .31, SD = .17; NG: M = .22, SD = .10; t(34) = 1.99, p < .05, d = .65; pSTG: HG: M = .31, SD = .14; NG: M = .22, SD = .14; t (34) = 2.75, p < .05, d = .93; FDR-corrected) (see Figure 3d). (Within

TABLE 1 Language network-level analyses results (within the left hemisphere)

| Contrasts | Region | BA | K_E | MNI peak coordinates (mm) | | | |
				x	y	z	T
HG + NG > B	IFG	44/45	541	−46	8	22	17.46
	pTL	22	1,161	−52	−36	10	12.81
HG > NG	IFG	44/45	28	−56	14	30	4.06
	pTL	22	78	−50	−48	30	4.79
			110	−54	−46	10	4.73
HG > NG	Broca's area	44	23	−58	14	28	4.15
	Broca's area	45	54	−52	32	18	4.86

Note: Under each contrast, activated regions within the left-hemispheric language network were shown with the correspondence to their BA, the cluster-sizes (K_E), the peak MNI coordinates and the t-values (T). The threshold was set to p < .05, FWE-corrected.

Abbreviations: aTL, anterior temporal lobe; B, implicit baseline; BA, Brodmann areas; FWE, family-wise error; HG, Hierarchical syntactic structure-building Grammar; IFG, inferior frontal gyrus; MFG, middle frontal gyrus; MNI, Montreal Neurological Institute; MNI, Montreal Neurological Institute; NG, Nested associating Grammar; pTL, posterior temporal lobe.

each ROI, peak signal intensity was also independent of the behavioral indexes, and showed between-group adaptation differences. See Supporting Information 5 for illustration.)

3.2.4 | Effective connectivity modeling results

Based on the results above, BA 44 and the pSTG were entered into the effective connectivity analysis as the core regions of the syntactic network for each grammar group. To uncover indirect coupling with other networks (Vakorin, Krakovska, & Mcintosh, 2009), the left caudate nucleus (CN) (see Supporting Information 4.4), MFG and inferior parietal lobule (ParInf) (see Supporting Information 4.5) were also entered as cognitive-general (mainly executive-functioning/cognitive control) regions for model specification. The masks of these regions were extracted from the anatomical automatic labeling template, and the corresponding ROIs were defined as in Section 2.8 for each group. For the HG group, the peak coordinates of each ROI were: CN (−12, 10, 4), MFG (−28, 4, 56), and ParInf (−36, −42, 42); for the NG group they were: CN (−16, −2, 16), MFG (−26, −4, 58), and ParInf (−42, −38, 48).

Model fit indexes showed that the effective connectivity models were reliably built for both grammar groups respectively (HG group: CFI = 1.00, NNFI = 1.00; NG group: CFI = 1.00, NNFI = .94; see Figure 4). As presented in Figure 4, the key differences between the two groups are:

1. In the HG group, BA 44 was the main hub receiving the lagged and contemporaneous experimental task effects, and its own lagged autoregressive ROI effect, whereas in the NG group, pSTG received these task effects and its own lagged autoregressive ROI effect. Here, the experimental task effect refers to the influence of the specific experimental manipulation (i.e., from the trials) on the

neural activity. Conversely, the ROI effect stands for the excitatory influence from a certain ROI. The ROI effect can be both lagged and contemporaneous, and become lagged autoregressive when the previous state of the ROI affects its own current state (see also Gates et al., 2011 for terminology).

2. In the HG group, BA 44 functionally projected a contemporaneous connection to the pSTG, while in the NG group, BA 44 could only affect the pSTG through a lagged connection. These differences, when taking the other executive-functioning ROIs into consideration, indicated that the effective connectivity between BA 44 and pSTG (also including the task effect) is qualitatively distinct for the respective grammar group. Common for both groups was, however, that the nodes (BA 44 or pSTG) of the syntactic network received the experimental task input and sent the information to the other cognitive-general regions within the cognitive control network.

4 | DISCUSSION

Our study sought to explore the involvement of the cortical syntactic network for hierarchical processing on the basis of category relations. In order to avoid any semantic processing bias, we designed a novel artificial grammar labeled as HG, composed of a set of natural syntax-like structure-building rules. Here, we focused on how participants applied these syntactic rules—which they had successfully learnt in a previous learning session—to solve a grammaticality judgment task during the scanning session. The processing of HG structures was compared to a grammar requiring the processing of nested structures (labeled as NG), which were conversely based on mere multilevel associations between the elements. Compared to NG sequences, the successful processing of HG sequences could therefore only be
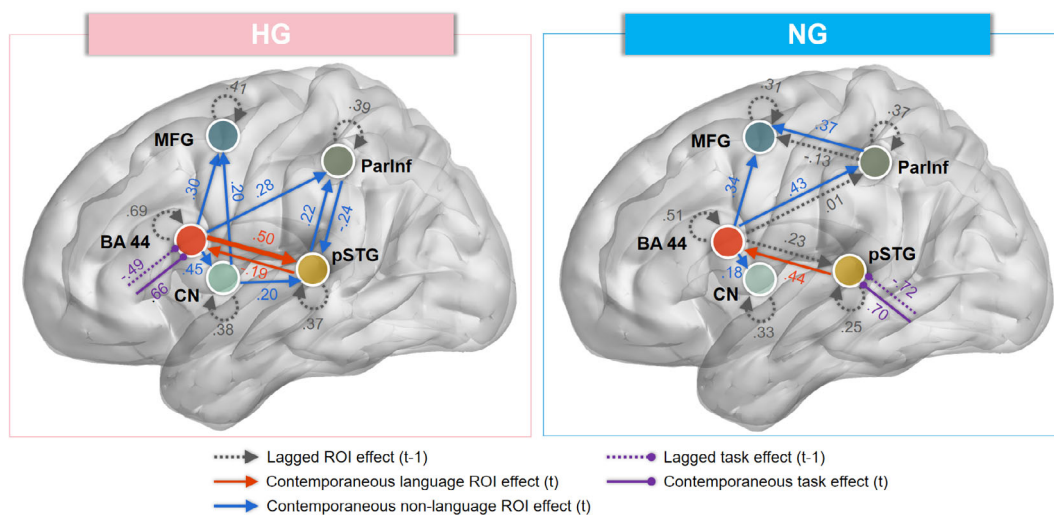

FIGURE 4 Effective connectivity modeling results. Group-mean connectivity strength (beta value) was also presented for each connection. BA, Brodmann area; CN, caudate nucleus; HG, Hierarchical syntactic structure-building Grammar; MFG, middle frontal gyrus; NG, Nested associating Grammar; ParInf, inferior parietal lobule; pSTG, posterior superior temporal gyrus

accomplished by establishing consecutive hierarchical syntactic relations among the categories of the elements. Overall, the fMRI results show that the involvement of BA 44, the pSTG, and their effective connectivity within the left-hemispheric language network might play a pivotal role in hierarchical syntactic processing, compared to the multilevel associative processing of the NG. We will discuss these results in turn below.

4.1 | The left BA 44: Syntactic hierarchy building on the basis of category relations

Previous studies testing natural language processing have assigned a possible syntactic role to Broca's area (Fedorenko, Duncan, & Kanwisher, 2012; Just et al., 1996; Makuuchi et al., 2009; Skeide, Brauer, & Friederici, 2016). At the neuroanatomical level, Broca's area has been shown to be receptoarchitectonically divided into two distinct subregions, approximately encompassing BA 45 and BA 44 (Amunts et al., 2010), whose corresponding functional specifications in language processing have long been debated (Fedorenko et al., 2011; Friederici, 2011, 2017; Grodzinsky & Friederici, 2006; Hagoort, 2005; Hagoort & Indefrey, 2014; Tyler et al., 2010; Vigneau et al., 2006). Activity in BA 45 has been shown to parametrically increase with the number of words forming a constituent, regardless of semantic information (Pallier et al., 2011). Moreover, a role for BA 45 during syntactic prediction has been proposed (Matchin et al., 2017; Matchin, Sprouse, & Hickok, 2014; Santi & Grodzinsky, 2012), yet a recent study (Goucha & Friederici, 2015) on the involvement of Broca's area in syntactic processing reported that only BA 44 was found active when lexical-semantic information and derivational morphology were suppressed in the stimulus material, but additional activation in BA 45 also appeared once lexical-semantic and derivational information was available. These data suggest that BA 45 might serve for semantic processing, and BA 44 for syntactic processing.

Our study specifies this syntactic hypothesis because we found that within Broca's area, only BA 44 reliably showed higher activation for hierarchical syntactic processing compared to associative processing. The effective connectivity modeling results further showed that, specifically during hierarchical syntactic processing in the HG group, BA 44 (a) receives the task effects and then spreads the processed information to other regions, and (b) it has a contemporaneous ROI effect upon the pSTG. According to Gates et al. (2011), this indicates that task input indirectly influenced the significant BOLD signal changes of the pSTG via its lagged and contemporaneous effects upon BA 44, thus highlighting BA 44 as a hub for online hierarchical syntactic structure construction on the basis of category relations. The critical involvement of BA 44 is consistent with the findings of Zaccarella and Friederici (2015), where the processing of basic combination of two elements to form a hierarchical phrase activated a sub-part within BA 44.

In contrast to these findings, Pallier et al. (2011) failed to detect activation of BA 44 when studying the processing of the increased size of word combinations of natural language. In that study, participants were required to detect probe sentences with button press, and to perform word memory tests. The authors ascribed the insensitivity of BA 44 to their task, that is, natural language processing without grammaticality judgments might not rely on BA 44. However, even when a simple low-level probe-word task was adopted, BA 44 still responds to syntactic operations (Goucha & Friederici, 2015). Moreover, in the present study, both grammars required participants to judge the grammaticality of the target structures (with the task difficulty controlled), so the higher activation of BA 44 for hierarchical syntactic processing could not be simply ascribed to task differences. Alternatively, the control stimuli of Pallier et al. (2011), containing pseudo-words and the real functional words, could also maintain the syntactic category relations and trigger the hierarchical syntactic structure build-up processes, which might have drastically diminished the detection power of BA 44 (see also Zaccarella, Schell, & Friederici, 2017). It is also noteworthy that we ran supplementary analyses to further exclude task difficulty effects from our data, which might have been reflected in the between-group difference of the accuracy rate in the scanning session. We found that the left IFG and pTL were still highly involved during HG processing even when a more sophisticated first-level GLM was adopted, which took the incorrectly responded and missing trials as regressors of no interest (see Supporting Information 4.3). Furthermore, we correlated signal intensities from different ROIs (including BA 44, 45, and pSTG) and the behavioral indexes (see Supporting Information 5) and found no tendency toward significance. In addition, supplementary involvement of regions commonly associated to executive function processes, and correlations between behavioral indexes and signal intensity in these areas might reflect task demands during the artificial grammar processing (see Supporting Information 4.5). This also seems to be confirmed by the effective connectivity results with the pattern of the connections projected from BA 44 to the executive-functioning regions. Together, these results suggest that BA 44 plays a critical role during hierarchical syntactic processing, independently of task demands.

Regarding the artificial grammar learning/processing research, activation of BA 44 has been reported by several studies adopting A^nB^n grammars to generate multilevel associations (Bahlmann et al., 2008, 2009; Friederici et al., 2006). The processing of multilevel associations in the current study replicated this result, but BA 44 was less involved in the NG group than in the HG group. Moreover, according to the effective connectivity profiles, BA 44 in the NG group did not receive the task effects. This suggested that BA 44 might be less sensitive to the NG during multilevel associating. Even though nesting of multiple associations might result in similar nonadjacent dependencies found in human language (Petkov & ten Cate, 2019), this similarity is only superficial, as it does not reflect the underlying internal hierarchical structures among the constituents (see Goucha et al., 2017 for a recent review). Thus, the associative processing specifically tested in previous artificial grammar studies (see Section 1) might be insufficient to account for human syntactic processing. It is worth noting that BA44 involvement has previously

been observed in the processing of simple linear grammar systems independent of processing requirements related to hierarchy (or the sense of "multilevel" as in NG) (e.g., Petersson et al., 2012). Yet, if it were true that the inferior frontal region (including BA 44) played a role in unifying various pieces of information regardless of whether hierarchical processing is needed, as claimed by Petersson et al. (2012), we would expect BA 44 to respond to both HG and NG with equal intensity and similar effective connectivity coupling, but this is not borne out in the present study. At the same time, previous evidence shows that multilevel association processing may itself lead to higher activation in BA 44 than the simpler adjacent associations (e.g., Bahlmann et al., 2008, 2009; Friederici et al., 2006). We therefore propose that BA44 may engage in binding processes of different syntactic complexities, but that it may be further specialized to handle syntactic hierarchies based on category relations, for which it is maximally involved.

4.2 | The left pSTG: Integration and storage of linguistic information

The pSTG (sometimes with an extension to pSTS) is also reported to support certain aspects of syntactic processing (Brennan, Stabler, Wagenen, Luh, & Hale, 2016; Ding et al., 2016; Hagoort & Indefrey, 2014; Matchin et al., 2014; Matchin et al., 2017; Matchin et al., 2019; Nelson et al., 2017). For instance, in adult language processing, Brennan et al. (2016) found that complex grammar models describing abstract hierarchical structures well captured the activation profiles of the left pTL (including pSTG) during passive natural language listening. Similarly, via intracranial recording, Nelson et al. (2017) further confirmed that the increase in the number of syntactic nodes to be combined (i.e., indicating the increase of the hierarchical syntactic structure construction load) led to higher activation in the pSTG during an auditory story comprehension task. Ontogenetically, Skeide et al. (2016) further found that higher activation of the pSTG was positively correlated with children's increased grammatical performance, indicating that the activation level (reflecting the functional maturity) of the pSTG could predict children's syntactic competence development. In addition, previous studies suggested that the pSTG might function as an integration component for mapping thematic relations onto syntactic structures, or more generally, for the integration of syntactic and semantic information (Ben-Shachar, Hendler, Kahn, Ben-Bashat, & Grodzinsky, 2003; Bornkessel et al., 2005; Constable et al., 2004; den Ouden et al., 2012; Matchin et al., 2014; Röder, Stock, Neville, Bien, & Rösler, 2002; see also recent reviews: Goucha et al., 2017; Zaccarella, Schell, & Friederici, 2017). However, even when semantic information was deprived, the posterior superior temporal cortex still responded to syntactic processing (Pallier et al., 2011). In line with this, both grammars of the current study activated the pSTG, providing extra evidence that the pSTG might be involved in the integration of various types of linguistic information (Boeckx, Martinez-Alvarez, & Leivada, 2014; Hagoort, 2005, 2013; Herrmann, Maess, Hasting, &

Friederici, 2009; Kuhnke, Meyer, Friederici, & Hartwigsen, 2017; Meyer, Obleser, Anwander, & Friederici, 2012; Xiang, Fonteijn, Norris, & Hagoort, 2010). Moreover, our effective connectivity modeling results revealed that activation of the pSTG was likely indirectly involved in hierarchical syntactic processing, given that it did not receive the experimental task effects. In contrast, the associative processing of the NG directly influenced the neural activity of the pSTG through both lagged and contemporaneous task effects. This could suggest that, depending on the stimulus items, multilevel associations might be integrated and consolidated in the pSTG, which can be easily accessible to the experimental input for retrieval and revision.

4.3 | Toward a syntactic network for hierarchical syntactic processing

Within each grammar group's model, the nodes of the syntactic network (BA 44 or pSTG) received the task effects and then send the processed information to the other cognitive-general regions, which are not specific to language. This implies that in principle, language-specific (i.e., here, syntactic processing-specific) regions might functionally modulate these cognitive-general regions. The different modulation patterns between the language ROIs and the executive-function ROIs in the two grammar groups might reflect grammar complexity differences, leading to task demand differences (see also Xiang et al., 2010 for a similar structural connectivity pattern).

A key difference regarding the syntactic network results between grammar groups is that the HG group relied on a contemporaneous connection from BA 44 to pSTG, whereas in the NG group, BA 44 only had a lagged effect on pSTG. Moreover, BA 44 in the HG group received the experiment task inputs, implying that the syntactic information was checked and processed for building up syntactic hierarchies. This is in line with the sense of "rule-based binding" from Opitz (2010). However, here, we highlight such a binding's hierarchical nature by providing additional neurobiological evidence from our novel artificial grammar processing. The processed information has to be transferred to pSTG in time for integration and storage, and feedback might be simultaneously sent from pSTG to BA 44. Such a real-time communication between these two core syntactic areas also provides neurobiological details for the possible computational syntactic parsing models like the one proposed by Battaglia, Borensztajn, and Bod (2012).

In contrast to the HG group, the experimental task inputs directly went to pSTG in the NG group model, indicating that multilevel associations might be used as a template for direct matching. Such a process was further modulated by the lagged ROI effect from BA 44, which is distinct from the contemporaneous connection from BA 44 to pSTG in the HG group model. Thus, the two key syntactic regions studied here might exchange different kinds of information within the left fronto-temporal network (see also Hagoort, 2005).

Overall, the present connectivity results converge on the existence of a functionally distinctive left fronto-temporal syntactic

network involved in hierarchical syntactic processing, with respect to the bi-directional contemporaneous connectivity between BA 44 and pSTG. Such a real-time efficient functional syntactic network is reminiscent of the potentially corresponding anatomical fiber tract, the arcuate fasciculus (AF). It has been reported to play a critical role in syntactic processing in ontogenetic, second language syntactic learning, phylogenetic, and aphasia studies (Brauer, Anwander, Perani, & Friederici, 2013; Dick, Bernal, & Tremblay, 2014; Friederici, 2011; Friederici, 2012; Friederici, 2017; Friederici, 2018; Friederici & Gierhan, 2013; Friederici & Singer, 2015; Meyer et al., 2012; Skeide et al., 2016; Wilson et al., 2011; Yamamoto & Sakai, 2017). Phylogenetically, human beings are equipped with a well-developed AF, when compared with other nonhuman primates (Balezeau et al., 2020; Rilling et al., 2008), highlighting its human-specific nature for language processing (Friederici, 2016, 2017, 2018; Goucha et al., 2017). This connectivity between BA 44 and the pSTG for category-relation-based hierarchical syntactic processing observed in the current study might structurally be supported by the AF, highlighting the capacity of using syntactic category relations to construct hierarchical syntactic structures as a remarkable human language faculty.

5 | CONCLUSION

The current study set out to provide neurobiological evidence for hierarchical syntactic processing by using a novel artificial grammar, the HG. Unlike previous association grammars such as A^nB^n grammar, this grammar contained a set of category-relation-based structure-building rules that participants were explicitly instructed to follow in order to correctly judge the grammaticality of artificially created syllabic sequences. We found a distinctive functional fronto-temporal network composed of BA 44 and the pSTG, which is crucial for hierarchical syntactic processing, beyond simpler multilevel associating. In particular, our data suggest that the left BA 44 is involved in the construction of hierarchical syntactic structures on the basis of category relations, and that the left pSTG might be involved in information integration and storage. The present novel grammar reliably reflects the hierarchical nature of human language, and can be flexibly adapted to ontogenetic and phylogenetic studies, thus paving a promising way to more delicate and informed neurobiological investigations into the uniqueness of the human language faculty.

CONFLICT OF INTEREST

All authors approved the final version of the manuscript for submission and declared no conflict of interest.

AUTHOR CONTRIBUTIONS

Luyao Chen and **Tomás Goucha**: Came up with the original idea of the Hierarchical syntactic structure-building Grammar. **Luyao Chen**, **Emiliano Zaccarella**, **Tomás Goucha**, **Claudia Männel**, and **Angela D. Friederici**: Designed the experiment. **Luyao Chen** and **Emiliano Zaccarella**: Conducted the pilot behavioral study and performed the actual fMRI experiment. **Luyao Chen** and **Emiliano Zaccarella**: Analyzed the data. All authors participated in the discussion of the results. **Luyao Chen** and **Emiliano Zaccarella**: Completed the manuscript, which was further revised by **Claudia Männel** and **Angela D. Friederici**.

DATA AVAILABILITY STATEMENT

Anonymized data will be made available upon reasonable requests addressed to the coauthors.

ORCID

Luyao Chen https://orcid.org/0000-0002-0796-3323
Emiliano Zaccarella https://orcid.org/0000-0002-5703-1778

REFERENCES

Abe, K., & Watanabe, D. (2011). Songbirds possess the spontaneous ability to discriminate syntactic rules. *Nature Neuroscience, 14*, 1067–1074.

Adger, D. (2013). *A syntax of substance*. Cambridge, MA: MIT Press.

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*(9), e1000489.

Amunts, K., Schleicher, A., Bürgel, U., Mohlberg, H., Uylings, H. B., & Zilles, K. (1999). Broca's region revisited: Cytoarchitecture and inter-subject variability. *Journal of Comparative Neurology, 412*(2), 319–341.

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.

Balezeau, F., Wilson, B., Gallardo, G., Dick, F., Hopkins, W., Anwander, A., … Petkov, C. I. (2020). Primate auditory prototype in the evolution of the arcuate fasciculus. *Nature Neuroscience, 23*, 611–614.

Battaglia, F. P., Borensztajn, G., & Bod, R. (2012). Structured cognition and neural systems: From rats to language. *Neuroscience and Biobehavioral Reviews, 36*(7), 1626–1639.

Beckers, G. J., Bolhuis, J. J., Okanoya, K., & Berwick, R. C. (2012). Birdsong neurolinguistics: Songbird context-free grammar claim is premature. *Neuroreport, 23*, 139–145.

Beltz, A. M., & Gates, K. M. (2017). Network mapping with GIMME. *Multivariate Behavioral Research, 52*(6), 789–804.

Ben-Shachar, M., Hendler, T., Kahn, I., Ben-Bashat, D., & Grodzinsky, Y. (2003). The neural reality of syntactic transformations: Evidence from functional magnetic resonance imaging. *Psychological Science, 14*(5), 433–440.

Berwick, R. C., Friederici, A. D., Chomsky, N., & Bolhuis, J. J. (2013). Evolution, brain, and the nature of language. *Trends in Cognitive Sciences, 17*(2), 89–98.

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

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*(1), 221–233.

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., & Pylkkänen, L. (2012). The time-course and spatial distribution of brain activity associated with sentence processing. *NeuroImage*, *60*, 1139–1148.

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 comprehension. *Brain and Language*, *157–158*, 81–94.

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.

Chen, L., Wu, J. G., 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.

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

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.

Dapretto, M., & Bookheimer, S. Y. (1999). Form and content: Dissociating syntax and semantics in sentence comprehension. *Neuron*, *24*, 2427–2432.

Davey, J., Cornelissen, P. L., Thompson, H. E., Sonkusare, S., Hallam, G., Smallwood, J., & Jefferies, E. (2015). Automatic and controlled semantic retrieval: TMS reveals distinct contributions of posterior middle temporal gyrus and angular gyrus. *Journal of Neuroscience*, *35*, 15230–15239.

de Vries, M. H., Monaghan, P., Knecht, S., & Zwitserlood, P. (2008). Syntactic structure and artificial grammar learning: The learnability of embedded hierarchical structures. *Cognition*, *107*(2), 763–774.

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*(1), 2–19.

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.

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

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.

Epstein, S., Kitahara, H., & Seely, D. (2014). Labeling by minimal search: Implications for successive-cyclic A-movement and the conception of the postulate 'phase'. *Linguistic Inquiry*, *45*, 463–481.

Everaert, M. B. H., Huybregts, M. A. C., Chomsky, N., Berwick, R. C., & Bolhuis, J. J. (2015). Structures, not strings: Linguistics as part of the cognitive sciences. *Trends in Cognitive Science*, *19*, 729–743.

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*(39), 16428–16433. https://doi.org/10.1073/pnas.111293710

Fedorenko, E., Duncan, J., & Kanwisher, N. (2012). Language-selective and domain-general regions lie side by side within Broca's area. *Current Biology*, *22*(21), 2059–2062.

Fedorenko, E., Hsieh, P.-J., Nieto-Castañon, A., Whitfield-Gabrieli, S., & Kanwisher, N. (2010). New method for fMRI investigations of language: Defining ROIs functionally in individual subjects. *Journal of Neurophysiology*, *104*(2), 1177–1194.

Ferrigno, S., Cheyette, S. J., Piantadosi, S. T., & Cantlon, J. F. (2020). Recursive sequence generation in monkeys, children, and native Amazonians. *Science Advances*, *6*, eaaz1002.

Fitch, W. T. (2014). Toward a computational framework for cognitive biology: Unifying approaches from cognitive neuroscience and comparative cognition. *Physics of Life Reviews*, *11*(3), 329–364.

Fitch, W. T., & Friederici, A. D. (2012). Artificial grammar learning meets formal language theory: An overview. *Philosophical transactions of the Royal Society of London. Series B, Biological Sciences*, *367*, 1933–1955.

Fitch, W. T., & Hauser, M. D. (2004). Computational constraints on syntactic processing in a nonhuman primate. *Science*, *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. (2012). The cortical language circuit: From auditory perception to sentence comprehension. *Trends in Cognitive Sciences*, *16*, 262–268.

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

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

Friederici, A. D. (2018). The neural basis for human syntax: Broca's area and beyond. *Current Opinion in Behavioral Sciences*, *21*, 88–92.

Friederici, A. D. (2019). Hierarchy processing in human neurobiology: How specific is it? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *375*(1789), 20180391.

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., & Gierhan, S. M. (2013). The language network. *Current Opinion in Neurobiology*, *23*(2), 250–254.

Friederici, A. D., Makuuchi, M., & Bahlmann, J. (2009). The role of the posterior superior temporal cortex in sentence comprehension. *Neuroreport*, *20*(6), 563–568.

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

Fujita, K. (2014). Recursive merge and human language evolution. In T. Roeper & M. Speas (Eds.), *Recursion: Complexity in cognition. Studies in theoretical psycholinguistics* (Vol. 43, pp. 243–264). Switzerland: Springer International Publishing.

Gates, K. M., & Molenaar, P. C. (2012). Group search algorithm recovers effective connectivity maps for individuals in homogeneous and heterogeneous samples. *NeuroImage*, *63*(1), 310–319.

Gates, K. M., Molenaar, P. C., Hillary, F. G., & Slobounov, S. (2011). Extended unified SEM approach for modeling event-related fMRI data. *NeuroImage*, *54*(2), 1151–1158.

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

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 and Biobehavioral Reviews*, *81*, 213–224.

Grodzinsky, Y., & Friederici, A. D. (2006). Neuroimaging of syntax and syntactic processing. *Current Opinion in Neurobiology*, *16*(2), 240–245.

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

Hartwigsen, G., Weigel, A., Schuschan, P., Siebner, H. R., Weise, D., Classen, J., & Saur, D. (2016). Dissociating parieto-frontal networks for phonological and semantic word decisions: A condition-and-perturb TMS study. *Cerebral Cortex*, 26, 2590–2601.

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

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.

Hornstein, N. (2009). *A theory of syntax: Minimal operations and universal grammar*. Cambridge: Cambridge University Press.

Hornstein, N., & Nunes, J. (2008). Adjunction, labeling, and bare phrase structure. *Biolinguistics*, 2(1), 57–86.

Hornstein, N., & Pietroski, P. (2009). Basic operations: Minimal syntax-semantics. *Catalan Journal of Linguistics*, 8, 113–139.

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 relation among categorization, merge and labeling, and their nature. *Biolinguistics*, 13, 1–21.

Huettel, S., Song, A., & McCarthy, G. (2004). *Functional magnetic resonance imaging*. Sunderland, MA: Sinauer Associates.

Jeon, H.-A., & Friederici, A. D. (2013). Two principles of organization in the prefrontal cortex are cognitive hierarchy and degree of automaticity. *Nature Communications*, 4, 2041.

Jiang, X., Long, T., Cao, W., Li, J., Dehaene, S., & Wang, L. (2018). Production of supra-regular spatial sequences by macaque monkeys. *Current Biology*, 28, 1851–1859.

Jobard, G., Vigneau, M., Mazoyer, B., & Tzourio-Mazoyer, N. (2007). Impact of modality and linguistic complexity during reading and listening tasks. *NeuroImage*, 34, 784–800.

Jung, J., & Lambon Ralph, M. A. (2016). Mapping the dynamic network interactions underpinning cognition: A cTBS-fMRI study of the flexible adaptive neural system for semantics. *Cerebral Cortex*, 26, 3580–3590.

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.

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(2), 85–93.

Kinno, R., Kawamura, M., Shioda, S., & Sakai, K. L. (2008). Neural correlates of noncanonical syntactic processing revealed by a picture-sentence matching task. *Human Brain Mapping*, 29, 1015–1027.

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.

Lau, E. F., Phillips, C., & Poeppel, D. A. (2008). Cortical network for semantics: (de)Constructing the N400. *Nature Reviews Neuroscience*, 9, 920–933.

Lenneberg, E. (1967). *Biological foundations of language*. New York, NY: Wiley.

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.

Makuuchi, M., Grodzinsky, Y., Amunts, K., Santi, A., & Friederici, A. D. (2013). Processing noncanonical sentences in Broca's region: Reflections of movement distance and type. *Cerebral Cortex*, 23(3), 694–702.

Martins, M. J. D., Bianco, R., Sammler, D., & Villringer, A. (2019). Recursion in action: An fMRI study on the generation of new hierarchical levels in motor sequences. *Human Brain Mapping*, 40(9), 2623–2638.

Matchin, W., Brodbeck, C., Hammerly, C., & Lau, E. (2019). The temporal dynamics of structure and content in sentence comprehension: Evidence from fMRI-constrained MEG. *Human Brain Mapping*, 40, 663–678.

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 tempo-parietal regions: The case of sentence processing. *NeuroImage*, 62, 1987–1998.

Misyak, J. B., Christiansen, M. H., & Tomblin, J. B. (2010). On-line individual differences in statistical learning predict language processing. *Frontiers in Psychology*, 1, 31.

Miyagawa, S., Berwick, R. C., & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, 4, 71.

Morgan-Short, K. (2020). Insights into the neural mechanisms of becoming bilingual: A brief synthesis of second language research with artificial linguistic systems. *Bilingualism: Language and Cognition*, 23(1), 87–91.

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.

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(2), e56230. https://doi.org/10.1371/journal.pone.0056230

Opitz, B. (2010). Neural binding mechanisms in learning and memory. *Neuroscience & Biobehavioral Reviews*, 34(7), 1036–1046.

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). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6), 2522–2527.

Petersson, K. M., Folia, V., & Hagoort, P. (2012). What artificial grammar learning reveals about the neurobiology of syntax. *Brain and Language*, 120, 83–95.

Petkov, C. I., & ten Cate, C. (2019). Structured sequence learning: Animal abilities, cognitive operations, and language evolution. *Topics in Cognitive Science*, 00, 1–15.

Pylkkännen, L. (2019). The neural basis of combinatory syntax and semantics. *Science*, 366(6461), 62–66.

Rey, A., Perruchet, P., & Fagot, J. (2012). Centre-embedded structures are a by-product of associative learning and working memory constraints: Evidence from baboons (Papio Papio). *Cognition*, 123, 180–184.

Rilling, J. K., Glasser, M. F., Preuss, T. M., Ma, X., Zhao, T., Hu, X., & Behrens, T. E. (2008). The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience*, 11, 426–428.

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.

Santi, A., & Grodzinsky, Y. (2012). Broca's area and sentence comprehension: A relationship parasitic on dependency, displacement or predictability? *Neuropsychologia*, 50(5), 821–832.

Sheng, J., Zheng, L., Lyu, B., Cen, Z., Qin, L., Tan, L. H., ... Gao, J. H. (2019). The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral Cortex*, 29(8), 3232–3240.

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

Snijders, T. M., Vosse, T., Kempen, G., van Berkum, J. J., 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.

Stobbe, N., Westphal-Fitch, G., Aust, U., & Fitch, W. T. (2012). Visual artificial grammar learning: Comparative research on humans, kea (*Nestor notabilis*) and pigeons (*Columba livia*). *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 1995–2006.

Tagarelli, K. M., Shattuck, K. F., Turkeltaub, P. E., & Ullman, M. T. (2019). Language learning in the adult brain: A neuroanatomical meta-analysis of lexical and grammatical learning. *NeuroImage*, 193, 178–200.

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(2), 352–364.

Uddén, J., Araujo, S., Forkstam, C., Ingvar, M., Hagoort, P., & Petersson, K. M. (2009). A matter of time: Implicit acquisition of recursive sequence structures. *Proceedings of the Cognitive Science Society*, 2009, 2444–2449.

Uddén, J., & Männel, C. (2018). Artificial grammar learning and its neurobiology in relation to language processing and development. In *The Oxford handbook of psycholinguistics* (2nd ed., pp. 755–783). Oxford, England: Oxford University Press. https://doi.org/10.1093/oxfordhb/9780198786825.013.33

Uddin, L. Q., Yeo, B. T. T., & Spreng, R. N. (2019). Towards a universal taxonomy of macro-scale functional human brain networks. *Brain Topography*, 32(6), 926–942.

Vakorin, V. A., Krakovska, O. A., & Mcintosh, A. R. (2009). Confounding effects of indirect connections on causality estimation. *Journal of Neuroscience Methods*, 184(1), 152–160.

Vigneau, M., Beaucousin, V., Herve, P. Y., Duffau, H., Crivello, F., Houde, O., ... Tzourio-Mazoyer, N. (2006). Meta-analyzing left hemisphere language areas: Phonology, semantics, and sentence processing. *NeuroImage*, 30(4), 1414–1432.

Whitney, C., Kirk, M., O'Sullivan, J., Lambon Ralph, M. A., & Jefferies, E. (2012). Executive semantic processing is underpinned by a large-scale neural network: Revealing the contribution of left prefrontal, posterior temporal, and parietal cortex to controlled retrieval and selection using TMS. *Journal of Cognitive Neuroscience*, 24, 133–147.

Wilson, B., Spierings, M., Ravignani, A., Mueller, J. L., Mintz, T. H., Wijnen, F., ... Rey, A. (2018). Non-adjacent dependency learning in humans and other animals. *Topics in Cognitive Science*, 12, 843–858. https://doi.org/10.1111/tops.12381

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.

Winkler, M., Mueller, J. L., Friederici, A. D., & Männel, C. (2018). Infant cognition includes the potentially human-unique ability to encode embedding. *Science Advances*, 4, eaar8334.

Wu, C., Zaccarella, E., & Friederici, A. D. (2019). Universal neural basis of structure building evidenced by network modulations emerging from Broca's area: The case of Chinese. *Human Brain Mapping*, 40(6), 1705–1717.

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

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.

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.

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.

Zaccarella, E., & Friederici, A. D. (2015). Merge in the human brain: A subregion based functional investigation in the left pars opercularis. *Frontiers in Psychology*, 6, 1818.

Zaccarella, E., Meyer, L., Makuuchi, M., & Friederici, A. D. (2017). Building by syntax: The neural basis of minimal linguistic structures. *Cerebral Cortex*, 27(1), 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 and Biobehavioral Reviews*, 80, 646–656.

SUPPORTING INFORMATION

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