# Is the Brain Mechanism for Hierarchical Structure Building Universal Across Languages? An fMRI Study of Chinese and English

**Xiaohan Zhang[1,2], Shaonan Wang[1,2], Nan Lin[3,4], Chengqing Zong[1,2,5]**

[1] National Laboratory of Pattern Recognition, Institute of Automation, CAS

[2] School of Artificial Intelligence, University of Chinese Academy of Sciences

[3] CAS Key Laboratory of Behavioural Sciences, Institute of Psychology

[4] Department of Psychology, University of Chinese Academy of Sciences

[5] CAS Center for Excellence in Brain Science and Intelligence Technology

{xiaohan.zhang, shaonan.wang, cqzong}@nlpr.ia.ac.cn; linn@psych.ac.cn

## Abstract

Evidence from psycholinguistic studies suggests that the human brain builds a hierarchical syntactic structure during language comprehension. However, it is still unknown whether the neural basis of such structures is universal across languages. In this paper, we first analyze the differences in language structure between two diverse languages: Chinese and English. By computing the working memory requirements when applying parsing strategies to different language structures, we find that top-down parsing generates less memory load for the right-branching English and bottom-up parsing is less memory-demanding for Chinese. Then we use functional magnetic resonance imaging (fMRI) to investigate whether the brain has different syntactic adaptation strategies in processing Chinese and English. Specifically, for both Chinese and English, we extract predictors from the implementations of different parsing strategies, i.e., bottom-up and top-down. Then, these predictors are separately associated with fMRI signals. Results show that for Chinese and English, the brain utilizes bottom-up and top-down parsing strategies separately. These results suggest that the brain adopts parsing strategies with less memory load according to different language structures.

## 1 Introduction

A hallmark of human language ability is combining linear sequential word inputs into a hierarchical structure using abstract syntactic rules. This ability enables us to create infinite expressions from finite words. Previous studies have shown that several brain regions are involved in building the hierarchical syntactic structure (Hagoort and Indefrey, 2014; Zaccarella et al., 2017). However, the parsing strategies that the brain uses to build such structures are less discussed. Moreover, most of these studies only have been conducted on one language. It is still unknown whether the neural basis of such structures is universal across languages.

This paper investigates the neural basis of syntactic structure-building, particularly the parsing strategies that the brain uses, in two diverse languages: Chinese and English. We focus on two parsing strategies: top-down and bottom-up. Since existing work has demonstrated that a parsing strategy requires different memory space when processing sentences with different branching directions(Resnik, 1992), we first analyze the branching structures and the memory load generated by two parsing strategies for Chinese and English. Results show that the dominant branching directions and the memory space required by parsing strategies are both different between the two languages. For the right-branching English, the top-down parsing needs less memory space than bottom-up parsing. Whereas for Chinese, which is more mixed in branching directions, the bottom-up parsing is less memory-demanding than the top-down parsing.

Therefore, we have two hypotheses about how the brain processes different languages:

- H1:The brain mechanism of syntactic processing is universal across different languages. Even though Chinese and English have different dominant structures, the brain uses the same parsing strategy no matter what structure of the language they are processing.

- H2: The brain mechanism of syntactic processing is relatively flexible across different languages. The parsing strategy adopted by the brain is regulated by cognitive resources and the strategy with less cognitive load would be preferred. That is, the brain uses different strategies when processing Chinese and English.

To test these two hypotheses, we associate the complexity predictors derived from different parsing strategies with the brain imaging data collected when native speakers were listening to stories. The



*Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 7852 - 7861
December 7-11, 2022 ©2022 Association for Computational Linguistics

complexity predictors are the number of parsing operations when using a parsing strategy to integrate each word into the tree. The key assumption is that brain regions engaged in syntactic structure-building would show increased activity as the number of parsing operations increases. Therefore, if a brain region builds trees following a parsing strategy, then the complexity predictors of this strategy would be able to predict the activation of this brain region. By comparing the prediction performances of different predictors, we can evaluate which parsing strategy better accounts for the brain activity in Chinese and English.

From the comparative study, we have the following interesting findings: the dominant predictor for Chinese is bottom-up but for English it is top-down, which is consistent with the less-memory-demanding strategies for each language. The brain regions with significant effects are also different between Chinese and English. However, in further analysis, we find that the data size gap and the correctness of constituency trees both contribute to the brain-region differences. These results support the second hypothesis that the brain adopts parsing strategies with less cognitive load for different languages.

In conclusion, our main contributions include:

- We investigated the brain mechanism of hierarchical structure building for Chinese and English by exploring the relationship between parsing strategies, language branching directions, and brain activation.
- We found that the processing load of parsing strategies correlates with the branching directions and the brain adopts the less-demanding parsing strategies for each language.
- Our results help to further understand how the brain processes language and would hopefully inspire artificial neural models to process or represent language more efficiently.

## 2 Related Work

Building hierarchical syntactic structures is an important sub-process of language understanding. Existing work that has investigated this sub-process can be categorized into two groups. One is often called controlled experiments that design artificial stimuli to separate the brain activation, such as comparing structured complex sentences or phrases with word lists (Pallier et al., 2011; Matchin et al., 2017; Sheng et al., 2018; Wu et al., 2019; Matchin and Hickok, 2019). These studies have many valuable findings and have identified several brain regions, including the left temporal lobe and the left frontal lobe, that are involved in syntactic structure building (Zaccarella et al., 2017; Wu et al., 2019). However, as designing artificial stimuli can only separate situations with or without structure-building, these controlled paradigms can hardly be used to further study the parsing strategies that the brain uses to build syntactic structures.

As a complement to controlled experiments, another line of work used naturalistic experimental paradigms and explored the brain mechanism using encoding models(Huth et al., 2016; Wang et al., 2020; Zhang et al., 2022a; Sun et al., 2021). By represent the structure-building process with syntactic predictors, Brennan et al. (2012) explored how the brain builds syntactic structure by using encoding models to predict brain activation with these predictors. This paradigm using syntactic predictors can conveniently incorporate different languages into the same framework. But most of these studies only focused on one language, mainly English(Hale et al., 2015; Brennan et al., 2016; Nelson et al., 2017). Less attention has been paid to other languages and cross-language comparisons. Moreover, little work has investigated the underlying reasons that drive the brain to utilize a parsing strategy.

This paper follows the naturalistic experimental paradigm and aims to explore the structural differences between Chinese and English and whether these differences drive the brain to use different parsing strategies in structure-building.

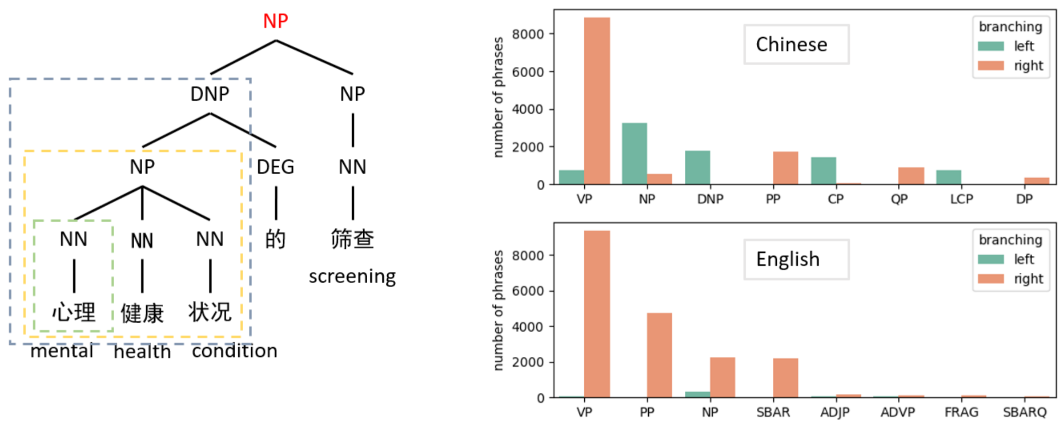
## 3 Materials

The English and Chinese fMRI datasets we use were both collected when native speakers were listening to narrative stories. All these audio stories are naturalistic stimuli and highly representative of the language that humans encounter in everyday life.

**English fMRI data** The English fMRI data we use comes from Zhang et al. (2020)[1], which was collected from 19 human subjects. The stimuli included 52 stories downloaded from the Moth Radio Hour[2] and each story lasts from 4 to 13 minutes.

[1] https://osf.io/eq2ba/

[2] https://themoth.org/radio-hour




(a) (b)

Figure 1: (a) A tree structure example. For the root node NP, the subtrees in the green box and the yellow box are its complete subtrees, but the subtree in the blue box is not a complete subtree because the right child of node DNP is not included in the box. (b) The number of left-branching and right-branching phrases in the Chinese and English stimuli corpora.

| | Chinese | | English | |
|---|---|---|---|---|
| | left | right | left | right |
| embedded | 3882 | 7258 | 300 | 712 |
| total | 8995 | 15072 | 1744 | 24381 |
| percent(%) | 43.16 | 48.16 | 17.20 | 2.92 |

Table 1: The number and percentage of embedded structures in left-branching and right-branching phrases in Chinese and English

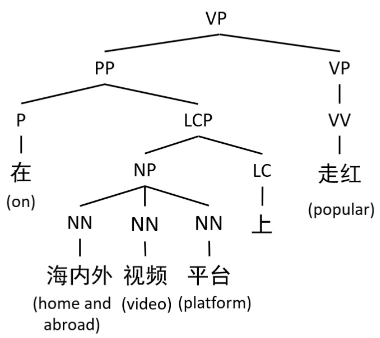



Figure 2: An example of the embedded tree structure in Chinese.

While being scanned for fMRI, each subject listened to a subset of the audio stories. In total, the story stimuli include 47,356 words, and the vocabulary size is 5,228 words (duplicates were excluded).

**Chinese fMRI data** We collected the Chinese fMRI data from 12 Chinese native speakers when they were listening to a total of 60 stories. Each of the subjects listened to all 60 stories, and each story was listened to once by one subject. During the scanning of fMRI, subjects were instructed to stay still and pay attention to the story they were hearing. All stories were downloaded from the Renmin Daily Review website[3] and each of them lasts from 4 to 7 minutes. The 60 stories contain 52,269 words, forming a vocabulary of 9,153 words. This fMRI dataset is publicly available at https://openneuro.org/datasets/ds004078. More details about fMRI acquisition and technical validation can be found in Wang et al. (2022).

Both the Chinese and English fMRI data were preprocessed following the HCP pipeline (Glasser et al., 2013).

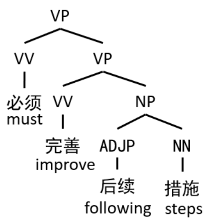
## 4 Language Structure Analysis

Chinese and English are two very diverse languages and differ in many aspects. We focus on the branching direction, which is directly correlated with the tree structure.

### 4.1 The Branching Directions of Languages

The branching direction of language is about the presented order of the head and the modifier in sentences. In English, sentences are largely left-headed and right-branching, which means the heads usually come before the modifiers. Whereas Chinese is more mixed with right-branching and left-branching categories (Levy and Manning, 2003). When the branching direction reflects on the trees, the right-branching structure

[3]https:/www.ximalaya.com/toutiao/30917322/




Top-down

| operations | incomplete nodes | node count |
|---|---|---|
| VP → VV VP | | |
| VV → 必须 (must) | 1 (VP) | 2 |
| VP → VV NP | | |
| VV → 完善 (improve) | 1 (NP) | 2 |
| NP → ADJP NN | | |
| ADJP → 后续 (following) | 1 (NN) | 2 |
| NN → 措施 (steps) | 0 | 1 |

Bottom-up

| operations | incomplete nodes | node count |
|---|---|---|
| VV → 必须 (must) | 1 (VV) | 1 |
| VV → 完善 (improve) | 2 (VV VV) | 1 |
| ADJP → 后续 (following) | 3 (VV VV ADJP) | 1 |
| NN → 措施 (steps) | 0 | 4 |
| NP → ADJP NN | | |
| VP → VV NP | | |
| VP → VV VP | | |

Figure 3: An Chinese example of the implementation of different parsing strategies, as well as the number of incomplete nodes and node count for each word.

| | | top-down | bottom-up | top-down>bottom-up | bottom-up>top-down |
|---|---|---|---|---|---|
| **Chinese** | avg | 4.69 | 4.29 | 65.59% | 32.22% |
| | max | 20 | 20 | 54.16% | 23.74% |
| **English** | avg | 4.59 | 5.32 | 16.21% | 80.58% |
| | max | 33 | 33 | 14.65% | 59.52% |

Table 2: The third and fourth columns are the average and maximum number of incomplete nodes generated during parsing. The fifth and the sixth columns are the percentage of sentences where the top-down or bottom-up parsing has a larger average and maximum number of incomplete nodes.

creates parse trees that grow down to the right and the left-branching structure creates parse trees that grow down to the left. Previous work has suggested that the branching direction of a language affects native speakers in working memory and the way they parse information (Friederici et al., 2017; Amici et al., 2019). Therefore, we analyzed the branching directions of all the language stimuli used in the fMRI collection and quantified the working memory load of top-down and bottom-up parsing.

We first computed the proportion of the left and right branching structures of the stimuli to see whether there is a real branching-direction difference. To better classify the branching direction of a syntactic tree, we define a subtree as *complete* if it has at least two nodes and the children of each node in this subtree are also included in this subtree (see Figure 1a). And a tree is left-branching if its root node only has two children and there are more complete subtrees on its left side. Conversely, a tree is right-branching if its root node only has two children and its right side has more complete subtrees than the left.

With this definition, we computed the branching direction of each phrase node in the English and Chinese stimuli corpora. The results, as shown in Figure 1b, are consistent with previous findings (Levy and Manning, 2003). The majority of the phrases in English, regardless of the phrase type, are right-branching. The branching directions of phrases in Chinese are rather mixed, with most of the VP phrases being right-branching and most of the NP and DNP phrases being left-branching. We also found that in Chinese, sometimes the branching structures are embedded, which means that the tree does not consistently grow in one direction. For example, as shown in Figure 2, the root node VP is left-branching because its left side has more complete subtrees. But its left child PP is right-branching and PP has a left-branching child LCP. We then computed the proportion of embedded phrases in Chinese and English. As shown in Table 1, nearly half of the phrases in Chinese are embedded, whereas in English, little phrases are embedded. These results prove the existence of a structural difference between Chinese and English, at least in the experimental stimuli corpus used in the fMRI collection.

### 4.2 The Correlation Between Parsing Strategies and Branching Directions

In the building process of a hierarchical constituency tree, each word in the sentence is parsed following the syntactic rules. A parsing strategy defines the specific parsing directions, whether moving from the words to abstract structures such as phrases and sentences, or starting at the abstract level and working down to the words. Here, we adopt two parsing strategies: top-down parsing, where the parsing begins from the most abstract level (root) to the word level (leaves); and bottom-up parsing, where the parsing begins from the word level to the abstract level. The parsing process of top-down and bottom-up parsing is illustrated in Figure 3 with an example Chinese phrase.

To further investigate the correlation between the branching direction and the parsing strategy, we computed the required working memory space when applying different parsing strategies to the stimuli corpus of the two languages. We used the *incomplete node* defined in Resnik (1992) and computed the number of incomplete nodes during parsing with each parsing strategy. A node is *incomplete* if either its parent or its children have not been established, in which case the brain must store it until it can be attached to a parent node or its child can be attached. Resnik (1992) demonstrated that a top-down parser requires $O(1)$ space for right-branching sentences and $O(n)$ space for left-branching sentences. A bottom-up parser requires $O(n)$space for right-branching sentences and $O(1)$space for left-branching sentences. The branching directions of Chinese and English, although not completely left-branching or right-branching, may affect the memory load of different parsing strategies. As the human brain processes language fast and accurately, a parsing strategy with less memory load may be more psychologically plausible for the brain to use.

The results are shown in Table 2. In general, although the branching directions in Chinese are mixed, the top-down parsing generates more incomplete nodes than the bottom-up parsing, and more than 65% of the sentences have more incomplete nodes when using top-down parsing. This means that top-down parsing is more memory-demanding than bottom-up parsing in the case of Chinese. Whereas in English, bottom-up parsing becomes more memory-demanding than top-down parsing because the average number of incomplete nodes generated by bottom-up parsing is larger than top-down parsing and more than 80% of the sentences have more incomplete nodes when using bottom-up parsing. This finding demonstrates that the distributions of branching directions of languages cause different processing loads for different parsing strategies.

# 5 fMRI Experiments

To test the two hypotheses, whether the brain uses the same parsing strategy regardless of language structures or whether the brain chooses a strategy with less cognitive load when processing different structures, we conducted an fMRI experiment as follows.

The overall framework is shown in Figure 4 We first computed the syntactic predictors from the parsing process of different parsing strategies. Then, we trained the voxel-wise encoding models to predict the fMRI signals from these syntactic predictors and compare their prediction accuracy in different brain regions and languages. The methodology for English and Chinese is the same, which makes the results of these two languages more comparable.

## 5.1 Syntactic Predictors

In the building process of a constituency tree, node count is the number of paring operations needed to integrate each word into the tree structure. Therefore, the syntactic node count is directly related to the process of syntactic structure building (Brennan et al., 2012). Following different parsing strategies, the number of parsing operations for each word would be different. As illustrated in Figure 3, for the four words in this VP-phrase, the top-down parsing performs $[2, 2, 2, 1]$ operations to integrate each word into the tree. Therefore, the top-down node count values for these three words are $[2, 2, 2, 1]$. Similarly, the bottom-up node count values are $[1, 1, 1, 4]$. For each sentence in the Chinese and English stimuli, we computed the node count values of each word with both top-down and bottom-up parsing strategies.

Apart from the node count predictors, we also compute two low-level linguistic features: sound envelope and word rate, to control for the confounding effects. The sound envelope is computed to represent how the amplitude and frequency of speech sound change over time. And the word rate, following the Brennan et al. (2012), identifies the endpoint of each word, where a value of 1 is labeled at the end time of each word and 0 at other times.

## 5.2 Constituency Trees

For the English stimuli, only the transcribed text for all the stories is provided in the dataset. To annotate the constituency trees of the story text,

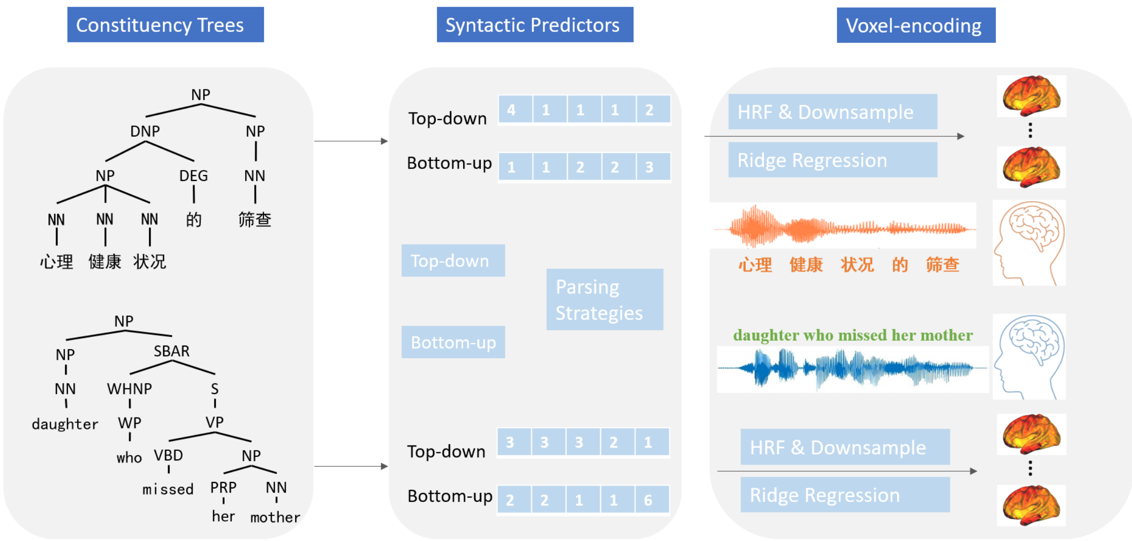



Figure 4: The overall framework of the fMRI experiment.

we used the Stanford CoreNLP parser[4]. For the Chinese stimuli, the text was downloaded from the Renmin Daily Review website and manually corrected to ensure the alignment of the audio and text. The constituency trees for each story were manually annotated.

The node count predictors were extracted from these annotated constituency trees using different parsing strategies.

### 5.3 Voxel-wise Encoding

We investigate the mapping between the structure-building process and brain activation using voxel-wise encoding models; that is, using node count features $x$ to predict brain activation $y$. In practice, fMRI measures the blood-oxygen-level-dependent (BOLD) signal, which changes slowly after the neurons fire. Besides, the frequency of fMRI collection is comparatively slow compared to the speech rate of words. To account for the influence of these two factors, the node count values of words are convolved with a canonical hemodynamic response function (HRF)[5] and then down-sampled to the same sampling rate as the fMRI collection.

To control the low-level linguistic effects represented by the word rate $wr$ and the sound envelope $snd$, we adopt a stepwise ridge regression method as the formalization of encoding models. Specifically, we perform a two-step regression. In the first step, we train the encoding models with word rate $wr$ and sound envelope $snd$ features, which are also convolved with HRF and down-sampled to the fMRI sampling rate.

$$y = \beta_{11} wr + \beta_{12} snd \tag{1}$$

In the second step, the node count predictor $x$ extracted from a parsing strategy, is added to the regression model.

$$y = \beta_{21} wr + \beta_{22} snd + \beta_{23} x \tag{2}$$

The regression weights $\beta$ are trained on the training set. After training, the prediction performance is evaluated on the test set with the Pearson correlation between the predicted and the actual voxel signal.

$$R = Pearson(y_{test}, \hat{y}_{test}) \tag{3}$$

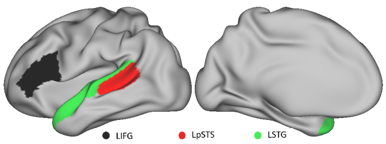
After the two-step training and test, a significance test is conducted to find voxels where the prediction performance is significantly improved by each node count predictor.

### 5.4 Training Details

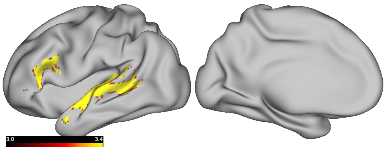
The voxel-wise encoding and subsequent significance test are conducted for the voxels in the following brain regions of interest (ROIs) previously associated with syntactic processing (Hagoort and Indefrey, 2014; Zaccarella et al., 2017): left superior temporal gyrus (LSTG), left posterior superior temporal sulcus (LpSTS), and left inferior frontal lobe (LIFL) (see Figure 5a). The brain atlas we use comes from Fan et al. (2016).

[4] `https://nlp.stanford.edu/software/stanford-dependencies.html`

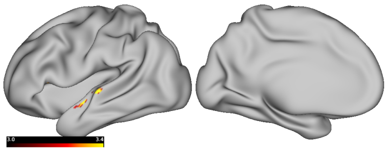
[5] The canonical HRF describes how BOLD signals would theoretically respond to a neural impulse



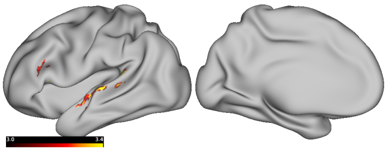

(a) ROIs



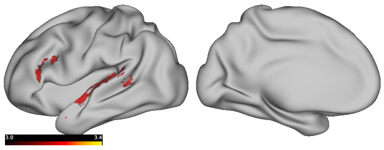
(b) Chinese bottom-up



(c) English top-down



(d) Chinese-small bottom-up



(e) Chinese-parser bottom-up

Figure 5: (a). Brain regions of interest. (b)-(e). Significant brain regions under different circumstances, $-log(q)$, FDR corrected, $q < 0.05$

All fMRI signals and the linguistic predictors of each story are z-scored before training and testing. Both the training and the significance testing are followed by the advice in Zhang et al. (2022b). Specifically, for both two steps of ridge regression, we run a nested cross-validation training which contains two loops: the inner loop and the outer loop. Both the inner-loop and the outer loop are 10-fold standard cross-validation. The inner loop chooses the best hyper-parameters (uniformly selected from the log-space from $10^{-5}$ to $10^{5}$) and computes the regression weights for each outer loop, and the outer loop tests the computed regression weight. After the two-step regression, we conduct a paired t-test on the outer loop results to extract voxels where adding each parsing node count can significantly improve the prediction accuracy.

## 6 Results

After brain encoding and the significance test, Figure 5b and Figure 5c are the brain regions that can be significantly predicted by different node count predictors.

### 6.1 Parsing Strategies with Significant Effects

In both Chinese and English, only one parsing strategy has significant effects, and the significant parsing strategies are different between the two languages. For Chinese, bottom-up parsing involves significant brain regions in the left temporal lobe and the left frontal lobe. Whereas for English, only the top-down parsing shows significant effects. The significant parsing strategy for each language is consistent with the one of less working-memory load as described in section 4.2. These results of fMRI experiments support the second hypothesis that the brain adopts parsing strategies with less cognitive load during the hierarchical structure-building process.

The memory constraint during language understanding has been discussed in existing work. As a sentence unfolds, new words rapidly obliterate previous words (Christiansen and Chater, 2015). The brain must process language efficiently under time and memory constraints. Liu et al. (2017) pointed out that to reduce the memory burden, human languages may have evolved to minimize the dependency distance. Futrell et al. (2015) provide quantitative evidence for dependency length minimization in 37 languages. The results of our experiments suggest that the brain utilizes the parsing strategies with less memory load for different branching languages, which can be seen as a complement to how the brain processes constituency structures under memory constraints.

### 6.2 Brain Regions for Structure-building

As shown in Figure 5b and Figure 5c, for both languages, the node count predictors show significant effects in the left cortex, which is consistent with previous findings. However, the involved brain regions, although overlapped to some extent, are different between Chinese and English.

For Chinese, the bottom-up predictor shows significant effects in the LSTG, the LpSTS, and the LIFG, as shown in Figure 5b. The brain regions for English top-down parsing, as shown in Figure 5c, only include a small area in the LSTG. However, we do not take these results to indicate that the LpSTS and the LIFG are not involved in syntactic computation. In fact, much of the previous work also finds that the LIFG and the LpSTS are involved in hierarchical syntactic computation in English(Zaccarella et al., 2017). In the next section, we conducted a detailed analysis of the possible explanations for these results.

# 7 Analysis

In this section, we investigate the possible reasons for the cross-language differences in the correlated brain regions. Our analysis is conducted on two experimental aspects, including the data size and the correctness of constituency trees between Chinese and English.

## 7.1 The Effects of Data Size

As a data-driven method, the results of encoding models would inevitably be affected by data size. The size of our Chinese fMRI data is remarkably larger than the English fMRI data. Therefore, we tested whether the cross-language brain-region differences related to the gap in data size by reducing the size of the Chinese fMRI data to the same level as the English data.

As described in section 3, the English fMRI data includes 51 naturalistic stories and 19 subjects, with each subject listening to a subset of the audio stories, and each story being listened to twice. To reduce the Chinese fMRI data to a similar size as the English fMRI data, we randomly divided all subjects into 6 groups with 2 subjects in each group. The fMRI response to each story is averaged within each group. Then, we randomly chose the averaged fMRI response of 55 stories across all groups to form a reduced fMRI dataset, which is approximately the size of the English fMRI data. The same voxel-encoding and significance test were conducted on this Chinese-small dataset.

The results are shown in Figure 5d. As shown, only the bottom-up parsing shows significant effects in a small area in the LIFG and the LSTG, which is very similar to the English brain areas significant in top-down parsing. Therefore, it is possible that the difference in data size results in the brain-region difference between Chinese and English. However, this needs to be tested with a larger English fMRI dataset in the future.

## 7.2 The Effects of the Correctness of Constituency Trees

Apart from the data size, the correctness of constituency trees may also influence the encoding results of node count predictors. As mentioned in section 5.2, the constituency trees are manually-labelled for Chinese stimuli but annotated by the trained Stanford CoreNLP parser for English stimuli. Therefore, the Chinese trees are correct, and the English trees inevitably have mistakes. These mistakes may further affect the encoding performance of node count predictors.

To test whether the correctness of constituency trees affects the encoding results, we conduct encoding for Chinese with node count values extracted from the constituency trees annotated by the Stanford CoreNLP parser. Results are shown in Figure 5e. As shown, it indeed affects the significant brain regions, for the significant brain regions are much smaller than those in Figure 5b.

In conclusion, the size and quality of data both affect the significant brain regions that the encoding models can find, which also highlights the importance of large-scale high-quality data. Reducing the data size or quality of Chinese data makes the significant brain regions more similar to the significant brain regions in English top-down parsing. However, none of these experimental factors affects the dominant parsing strategy for Chinese, which further supports that the different branching directions are the reason for the different dominant parsing strategies between Chinese and English.

# 8 Conclusion

To investigate whether the brain mechanism for hierarchical structure building is universal across languages, this work investigated the correlation between language branching directions, parsing strategies, and brain activation. By comparing the fitness of the complexity metrics extracted from different parsing strategies in two diverse languages, i.e., Chinese and English, we find experimental results supporting the hypothesis that the language structure may play an important role in determining the parsing strategy that the brain uses. That is, the brain may use different parsing strategies for different language structures to reduce the cognitive load. Our results demonstrate the flexibility of the brain mechanism for language processing and highlight the importance of cross-language studies in studying the brain language comprehension.

## Limitations

This work has several limitations, which may restrict the generalization of our findings.

Although we speculate that the language branching direction affects the parsing strategy the brain uses and try to prove it through working memory demand, we cannot directly verify it using an encoding framework. Because the Chinese experimental stimuli are rather mixed, the fMRI response of left-branching phrases can hardly be separated from the right ones. Future research can carefully design language stimuli with left-branching and right-branching structures separated, or use a metric other than node count to study the relationship between the brain parsing strategy and the language branching direction.

In addition, node count is only associated with parsing difficulty. More detailed information during the tree-building process, such as the phrase nodes to be generated, or the specific parsing operation to be performed, cannot be represented by such a simple metric. Therefore, more powerful representations of the parsing process and the information in the hierarchical tree are needed if we wish to further uncover the mechanism of bran syntactic computation. Future work can use neural language models like BERT to generate more powerful representations.

## Ethics Statement

The collection of Chinese fMRI data was conducted under the approval of the Institutional Review Board of Peking University.

## Acknowledgements

This work is supported by the National Key Research and Development Program of China under grant 2021ZD0204105, as well as the Natural Science Foundation of China under Grant 62036001 and 61906189. This work is also supported by the independent research project of National Laboratory of Pattern Recognition.

## References

Federica Amici, Alex Sánchez-Amaro, Carla Sebastián-Enesco, Trix Cacchione, Matthias Allritz, Juan Salazar-Bonet, and Federico Rossano. 2019. The word order of languages predicts native speakers' working memory. *Scientific Reports*, 9.

Jonathan Brennan, Yuval Nir, Uri Hasson, Rafael Malach, and Liina Pylkkänen. 2012. Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and Language*, 120:163–173.

Jonathan Brennan, E. Stabler, Sarah E. Van Wagenen, Wen-Ming Luh, and John Tracy Hale. 2016. Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and Language*, 157:81–94.

Morten H. Christiansen and Nick Chater. 2015. The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39.

Lingzhong Fan, Hai Li, Junjie Zhuo, Yu Zhang, Jiaojian Wang, Liangfu Chen, Zhengyi Yang, Congying Chu, Sangma Xie, Angela R. Laird, Peter T. Fox, Simon B. Eickhoff, Chunshui Yu, and Tianzi Jiang. 2016. The human brainnetome atlas: A new brain atlas based on connectional architecture. *Cerebral Cortex (New York, NY)*, 26:3508 – 3526.

Angela D. Friederici, Noam Chomsky, Robert C. Berwick, Andrea Moro, and Johan J Bolhuis. 2017. Language, mind and brain. *Nature Human Behaviour*, 1:713–722.

Richard Futrell, Kyle Mahowald, and Edward Gibson. 2015. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112:10336 – 10341.

Matthew F. Glasser, Stamatios N. Sotiropoulos, J. Anthony Wilson, Timothy S. Coalson, Bruce R. Fischl, Jesper L. R. Andersson, Junqian Xu, Saâd Jbabdi, Matthew A. Webster, J. Polimeni, David C. Van Essen, and Mark Jenkinson. 2013. The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80:105–124.

Peter Hagoort and Peter Indefrey. 2014. The neurobiology of language beyond single words. *Annual review of neuroscience*, 37:347–62.

John Hale, David Lutz, Wen-Ming Luh, and Jonathan Brennan. 2015. Modeling fmri time courses with linguistic structure at various grain sizes. In *CMCL@NAACL-HLT*.

Alexander G. Huth, Wendy A. de Heer, Thomas L. Griffiths, Frédéric E. Theunissen, and Jack L. Gallant. 2016. Natural speech reveals the semantic maps that tile human cerebral cortex. *Nature*, 532:453 – 458.

R. Levy and Christopher D. Manning. 2003. Is it harder to parse chinese, or the chinese treebank? In *ACL*.

Haitao Liu, Chunshan Xu, and Junying Liang. 2017. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of life reviews*, 21:171–193.

William Matchin, Christopher Mathias Hammerly, and Ellen F. Lau. 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.

William Matchin and Gregory Hickok. 2019. The cortical organization of syntax. *Cerebral cortex*.

Matthew J. Nelson, Imen El Karoui, Kristóf Giber, Xiaofang Yang, Laurent D. Cohen, Hilda J Koopman, Sydney S. Cash, Lionel Naccache, John Tracy Hale, Christophe Pallier, and Stanislas Dehaene. 2017. Neurophysiological dynamics of phrase-structure building during sentence processing. *Proceedings of the National Academy of Sciences*, 114:E3669 – E3678.

Christophe Pallier, Anne-Dominique Devauchelle, and Stanislas Dehaene. 2011. Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, 108(6):2522–2527.

Philip Resnik. 1992. Left-corner parsing and psychological plausibility. In *COLING 1992 Volume 1: The 14th International Conference on Computational Linguistics*.

Jingwei Sheng, Li Zheng, Bingjiang Lyu, Zhehang Cen, Lang Qin, Li Hai Tan, Mingxiong Huang, Nai Ding, and Jia-Hong Gao. 2018. The cortical maps of hierarchical linguistic structures during speech perception. *Cerebral cortex*.

Jingyuan Sun, Shaonan Wang, Jiajun Zhang, and Chengqing Zong. 2021. Neural encoding and decoding with distributed sentence representations. *IEEE Transactions on Neural Networks and Learning Systems*, 32:589–603.

Shaonan Wang, Jiajun Zhang, Nan Lin, and Chengqing Zong. 2020. Probing brain activation patterns by dissociating semantics and syntax in sentences. In *AAAI*.

Shaonan Wang, Xiaohan Zhang, Jiajun Zhang, and Chengqing Zong. 2022. A synchronized multimodal neuroimaging dataset for studying brain language processing. *Scientific Data*, 9.

Chiao-Yi Wu, Emiliano Zaccarella, and Angela D. Friederici. 2019. Universal neural basis of structure building evidenced by network modulations emerging from broca's area: The case of chinese. *Human Brain Mapping*, 40:1705 – 1717.

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

Xiaohan Zhang, Shaonan Wang, Nan Lin, Jiajun Zhang, and Chengqing Zong. 2022a. Probing word syntactic representations in the brain by a feature elimination method. In *AAAI*.

Xiaohan Zhang, Shaonan Wang, and Chengqing Zong. 2022b. How does the experimental setting affect the conclusions of neural encoding models? In *LREC*.

Yizhen Zhang, Kuan Han, Robert M. Worth, and Zhongming Liu. 2020. Connecting concepts in the brain by mapping cortical representations of semantic relations. *Nature Communications*, 11(1):1877.