- Title: Long-distance linguistic dependencies in Chinese and English brains

- Abbreviated title: Long-distance phenomena in cross-linguistic brains

- Author names and affiliations, including postal codes:

- Donald Dunagan*: Department of Linguistics, University of Georgia, Athens, GA 30602-6205, USA

- Miloš Stanojević: DeepMind, 5 New Street Square, London, EC4A 3TW, UK

- Maximin Coavoux: CNRS, Université Grenoble Alpes, Saint-Martin-d'Hères, 38400, France

- Shulin Zhang: Department of Linguistics, University of Georgia, Athens, GA 30602-6205 USA

- Shohini Bhattasali: Department of Language Studies, University of Toronto Scarborough, Toronto, ON. Canada, M1C 1A4

- Jixing Li: Department of Linguistics and Translation, City University of Hong Kong, 83 Tat Chee Ave, Kowloon Tong, Hong Kong, China

- Jonathan Brennan: Department of Linguistics, University of Michigan, Ann Arbor, MI 48109, USA

- John Hale: Department of Linguistics, University of Georgia, Athens, GA 30602-6205, USA

- Corresponding author email address:

- *dgd45125@uga.edu

- Number of pages: 31

- Number of figures, tables, multimedia, and 3D models (separately):

- figures: 5

- tables: 7

- multimedia: 0

- 3D models: 0

- Number of words for abstract, introduction, and discussion (separately):

- abstract: 249

- introduction: 649

- discussion: 1162

- Conflict of interest statement: The authors declare no competing financial interests.

- Acknowledgments: This material is based upon work supported by the National Science Foundation under Grant No. 1903783. We would like to give special thanks to Christophe Pallier, Benoît Crabbé and Wen-Ming Luh.

Abstract

Words can occur arbitrarily far away from where they contribute their meaning in a sentence. Two examples are WH-questions (WHQs), which begin with a WH-word like *what* and object-extracted relative clauses (ORCs), in which a noun is modified by a sentence-like grammatical unit. While these *long-distance dependencies* have been extensively studied, never before have their brain bases been examined from a multi-lingual, naturalistic perspective. This study fills this gap by analyzing WHQs and ORCs in fMRI data collected while 35 Chinese participants (15 females) and 49 English participants (30 females) listen to translation-equivalent stories. These languages exhibit radical typological differences in word order in these constructions. It remains unknown whether the brain basis for comprehension in these languages is similar or different. Separate general linear model analyses were performed and voxel-level intersections were calculated between the results to identify common regions of selectively increased activation during the comprehension of these linguistic constructions. Further Bayesian region of interest analyses probed whether common increases were truly similar. We found remarkable cross-linguistic commonality for both constructions. WHQs were associated with increased activation in the left middle and superior temporal lobe, left temporoparietal junction, left inferior frontal gyrus, and bilateral medial frontal lobe. ORCs were associated with increased activation in the left middle temporal lobe, left inferior frontal gyrus, bilateral angular gyrus, bilateral posterior cingulate, bilateral precuneus, and left medial frontal lobe. These results support the hypothesis that, regardless of form, the brain bases of higher-level language processing are uniform across languages.

Significance statement

In natural language, words can occur arbitrarily far away from where they make their meaning contribution. These are called long-distance dependencies (LDDs). One example is WH-questions, which are questions that begin with a WH-word like *what*. In this article, we generalize previous work on the brain bases of LDD processing in three way. We: (1) generalize across languages; (2) generalize across LDD construction types; and (3) generalize to naturalistic stimuli. While previous work has shown that the brain processing of language in the context of story comprehension is cross-linguistically highly consistent, we analyze specific sentence types which occur differently in different languages and demonstrate that they are associated with highly similar patterns of brain activation.

1 Introduction

Linguists use the term "long-distance dependency" (LDD) for situations where a word or phrase shows up relatively far away from the position where it contributes to the overall meaning of the sentence (see Chaves & Putnam, 2021, for a comprehensive overview). Two examples are WH-questions (WHQs), which are questions that (in English) begin with a WH-word like *what* or *where* and object-extracted relative clauses (ORCs), in which a noun is modified by a sentence-like grammatical unit. Concrete examples and more details are given in section 2.2. ORCs are a well-known stress-test of the brain's language network (Ben-Shachar et al., 2003; Just et al., 1996; Stromswold et al., 1996) and LDDs are among the most well-theorized constructions in all of natural language (for reviews, see Cheng & Corver, 2006; Levine & Hukari, 2006). This article contributes to the LDD and neuroscience literature by generalizing the brain bases of LDD processing in three ways. We: (1) generalize across languages; (2) generalize across LDD construction types; and (3) generalize to naturalistic stimuli. We analyze the brain activity associated with two LDD constructions: WHQs and ORCs in two different languages: Chinese and English, using publicly available, naturalistic, parallel fMRI data: *The Little Prince Datasets* (Hale et al., 2022; Li et al., 2021, 2022).

Since Chomsky (1977), LDDs have been viewed as related to one another—a coherent class. Previous neuroimaging work with different kinds of LDDs has primarily implicated the inferior frontal gyrus (Ben-Shachar et al., 2003, 2004; Caplan et al., 2008; Just et al., 1996; Stromswold et al., 1996) and superior temporal lobe (Ben-Shachar et al., 2003, 2004; Caplan et al., 2008; Just et al., 1996), suggesting a uniform processing basis despite differences between LDD types. LDDs are of further interest as this coherent class exhibits interesting typological similarities and differences between languages (see Bornkessel-Schlesewsky & Schlesewsky, 2016, for the importance of cross-linguistic investigation in the neurobiology of language). It is underexplored whether similar or different patterns of activation are elicited when phenomena like LDDs have linguistic differences. Previous work, though, has found high degrees of activation similarity in the language network when Russian and English speakers listen to a translation-equvalent narrative (Honey et al., 2012), high degrees of activation similarity for singular vs. plural comprehension across languages which treat this phenomenon differently (Dunagan et al., 2022), and that patterns of activation in the default mode network can identify the distributed representation a story, even when the representation was calculated from a translation of that story in another language (Dehghani et al., 2017). Furthermore, a recent, large-scale project has documented cross-linguistic similarity in the location, lateralization, selectivity, and functional integration of the language network (Malik-Moraleda et al., 2022). This last study, though, probed language processing holistically and did not tease apart granular aspects of processing that have been compellingly linked to cross-linguistic differences.

There is a tension between neural uniformity and typological variability that has not yet been put to the test, but is tested here in the domain of two LDDs. This study tests, using naturalistic stimuli and two different languages: (1) whether the brain bases of ORC processing and WHQ processing are the same or different and (2) whether the brain bases of ORC processing and WHQ processing are the same or different for Chinese and English speakers. We hypothesize that the correlates for processing the two LDD constructions will be the same and implicate at least the left inferior frontal gyrus and left superior temporal lobe. We also hypothesize that the activity associated with the two constructions will be the same for Chinese and English speakers, despite the differences between these two languages. To test these hypotheses, we analyzed WHQs and ORCs in fMRI data collected while 35 Chinese participants (15 females) and 49 English participants (30 females) listened to translation-equivalent audiobooks of the children's story *Le Petit Prince* (*The Little Prince*), by de Saint-Exupéry.

2 Materials and methods

The fMRI data analyzed were *The Little Prince Datasets* (Hale et al., 2022; Li et al., 2021, 2022), a published collection of datasets in which Chinese and English participants were scanned while they engaged in the naturalistic process of listening to a translation-equivalent audiobook in their native language. The participants listened to the children's story *The Little Prince*. General linear models (GLMs) were fit separately for each language and included predictors for each of the LDDs as well as nuisance regressors of non-interest. To probe the cross-linguistic commonality of LDD processing, the voxel-wise intersections of the LDD effects in English and Chinese were calculated. While the GLM analysis allowed for the identification of voxels in which increases or decreases in activation occur in both languages, it did not allow for the ability to ask whether those increases and decreases were *similar*. To answer this question, the BOLD time series from both the Chinese and English participants were entered into Bayesian analyses which estimated posterior distributions over parameter values. The posterior distributions of the LDD parameters were compared with Bayes factor analyses.

2.1 fMRI data

2.1.1 Participants

Chinese participants were 35 healthy, right-handed young adults (15 females, mean age=19.3, SD=1.6). They self-identified as native Chinese speakers, and had no history of psychiatric, neurological, or other medical illness that could compromise cognitive functions. All participants were paid, and gave written informed consent prior to participation, in accordance with the IRB guidelines of Jiangsu Normal University.

English participants were 49 young adults (30 females, mean age=21.3, SD=3.6) with no history of psychiatric, neurological or other medical illness that might compromise cognitive functions. They self-identified as native English speakers and strictly qualified as right-handed on the Edinburgh handedness inventory. All participants were paid, and gave written informed consent prior to participation, in accordance with the IRB guidelines of Cornell University.

2.1.2 Procedure

After giving their informed consent, participants were familiarized with the MRI facility and assumed a supine position on the scanner. They were instructed to not move as best as they could throughout scanning as movement would make the scans unusable. Next, participants were put in the head-coil with pillows under and on the sides of their head and under the knees for comfort and to reduce movement over the scanning session. Participants were given a bulb in their right hand and told to squeeze if something was wrong or they needed a break during scanning. Once in place, participants chose an optimal stimulus volume by determining a level that was loud but comfortable. Auditory stimuli were delivered through MRI-safe, high-fidelity headphones inside the head coil (English: Confon HP-VS01, MR Confon, Magdeburg, Germany; Chinese: Ear Bud Headset, Resonance Technology, Inc, California, USA). The headphones were secured against the plastic frame of the coil using foam blocks.

The English and Chinese participants went through one scanning session, which was divided into 9 runs, with each lasting for about 10 minutes. Participants listened passively to 1 section of the audiobook in each run and completed 4 quiz questions after each run (36 questions in total). These questions were used to confirm their comprehension and were viewed by the participants via a mirror attached to the head coil and they answered through a button box. During scanning, participants were monitored by a camera over their left eye. If they appeared drowsy or seemed to move too much during the movie, the operator of the scanner gave them a warning over the intercom by producing a beep or speaking to them. During breaks between the runs, participants were told that they could relax but not move. Finally, participants were paid and sent home. The entire session lasted for around 2.5 hours.

2.1.3 Stimuli

Participants listened to a translation-equivalent audiobook of *The Little Prince* in its entirety. The English audiobook is 94 minutes long, translated by David Wilkinson, and read by Karen Savage. The Chinese audiobook is 99 minutes long and read by a professional female Chinese broadcaster hired by the experimenter.

2.1.4 Acquisition

MRI images were acquired with a 3T MRI GE Discovery MR750 scanner with a 32-channel head coil. Anatomical scans were acquired using a T1-weighted volumetric Magnetization Prepared Rapid Gradient-Echo pulse sequence. Functional scans were acquired using a multi-echo planar imaging sequence with online reconstruction (TR=2000 ms; TEs=12.8, 27.5, 43 ms; FA=77°; matrix size=72 x 72; FOV=240.0 mm x 240.0 mm; 2x image acceleration; 33 axial slices; voxel size=3.75 x 3.75 x 3.8 mm).

2.1.5 Preprocessing

MRI data files were converted from DICOM to NIfTI format and preprocessed using AFNI (Cox, 1996, version 16). *Anatomical.* The anatomical/structural MRI scans were deskulled using *3dSkullStrip*. The resulting anatomical images were nonlinearly aligned to the Montreal Neurological Institute N27 template brain. Resulting anatomical images were used to create grey matter masks. *Functional.* The first 4 volumes in each run were excluded from analyses to allow for T1-equilibration effects. The fMRI timeseries were then corrected for slice-timing differences (*3dTshift*) and despiked (*3dDespike*). Next, volume registration was done by aligning each timepoint to the mean functional image of the center timeseries (*3dvolreg*). Then the volume-registered and anatomically-aligned functional data were nonlinearly aligned to the MNI template brain. Multi-echo independent components analysis (Kundu et al., 2012) was used to denoise data for motion, physiology and scanner artifacts. Images were then resampled at 2 mm cubic voxels (*3dresample*).

2.2 Storybook annotations

Each word in both of the storybooks was annotated for a number of metrics which were expected to be cognitively influential. The first step in automatically annotating the storybook texts was to parse them for syntactic structure. A parser is a natural language processing tool which syntactically annotates sentences (see chapter 10 of Eisenstein, 2019, for a introduction). The syntactic annotations, or trees, were used to calculate a word-by-word syntactic complexity metric and to calculate word-by-word annotations for the LDDs. Some parsing systems work with sentence subpieces whose words are necessarily contiguous whereas others can cope with some limited amount of "discontinuity" i.e., constituents whose words are not all contiguous (for a brief overview and a mathematical theorization of these sorts of formal grammars, see Nederhof & Yli-Jyrä, 2021). The English text was parsed using the discontinuity-capable parser `mtgpy` (Coavoux, 2021) while the Chinese text was parsed using the discontinuity-incapable parser `benepar` (Kitaev et al., 2019). The `benepar` parser is appropriate for Chinese as, in Chinese, WHQs are continuous. In linguistics, this is referred to as *in situ WH*. Following

Brennan et al. (2012), a word-by-word "bottom-up" complexity metric corresponding to the count of reduce operations in an incremental shift-reduce parser was defined using the continuous parse trees. The use of 1-best parses means that the metric is blind to any sort of ambiguity resolution work.

As introduced above, the study concerns two constructions: ORCs and WHQs. A relative clause is a sentence-like grammatical unit that modifies a noun. It differs from a true sentence in that it lacks an element whose meaning is contributed by the head noun that the relative clause itself modifies. ORCs are relative clauses where the head noun is the object of the relative clause verb. In these and other LDDs, the extracted element is known as the *filler*, while the position where it makes its contribution to the meaning of the sentence is known as the *gap* (*site*). Example 1 shows an ORC in English where *the flower* is the filler element which makes it contribution to the overall meaning of the sentence at the coindexed, empty underscore (the gap site). It is an ORC because the head noun, *the flower*, is the object of the relative clause verb, *love*. Fig. 1**A** gives the parse tree for example 1 as found by the `mtgpy` parser. The underlined words, *that you love*, indicate the span between the filler and the gap site where an increase in cognitive demand is expected (e.g., Wanner & Maratsos, 1978). As compared to English ORCs, in which the head noun occurs before the relative clause, Chinese ORCs are prenominal. This is exemplified by example 2, where *DE* is a special relative marker and *planet* is the filler. Fig. 1**B** gives the tree for example 2. The underlined words, *DE planet*, indicate the span from the gap site through the filler and indicate where increased cognitive demand is expected (see Hsiao & Gibson, 2003; Jäger et al., 2015; Yun et al., 2015, for discussion of the differences and potential similarities between processing costs of English and Chinese ORCs).

WHQs are sometimes called *information questions* in order to differentiate them from *yes/no questions*. In English, the spelling of the question word usually starts with the letters *W* and *H* (*who*, *what*, *where*, etc.). In English WH-questions, the WH-word or phrase commonly occurs at the beginning of the sentence, even though that is not the position where it makes its contribution to the overall meaning of the sentence. The underscore in example 3 marks the gap site for the WH-phrase *what planet*. Fig. 1**C** gives the tree for example 3. The underlined words, *what planet are you from*, span through the filler to the gap site and indicate where increased cognitive demand is expected. In contrast to English, Chinese has in situ WH-phrases, as demonstrated by example 4. In these cases, a LDD is formed between the wh-phrase and an unboundedly higher clausal position to its left (Huang, 1982, 1992; Xiang et al., 2015). Fig. 1**D** gives the tree for example 4. The underlined word, *what*, spans the wh-phrase and indicates where increased cognitive demand is expected.

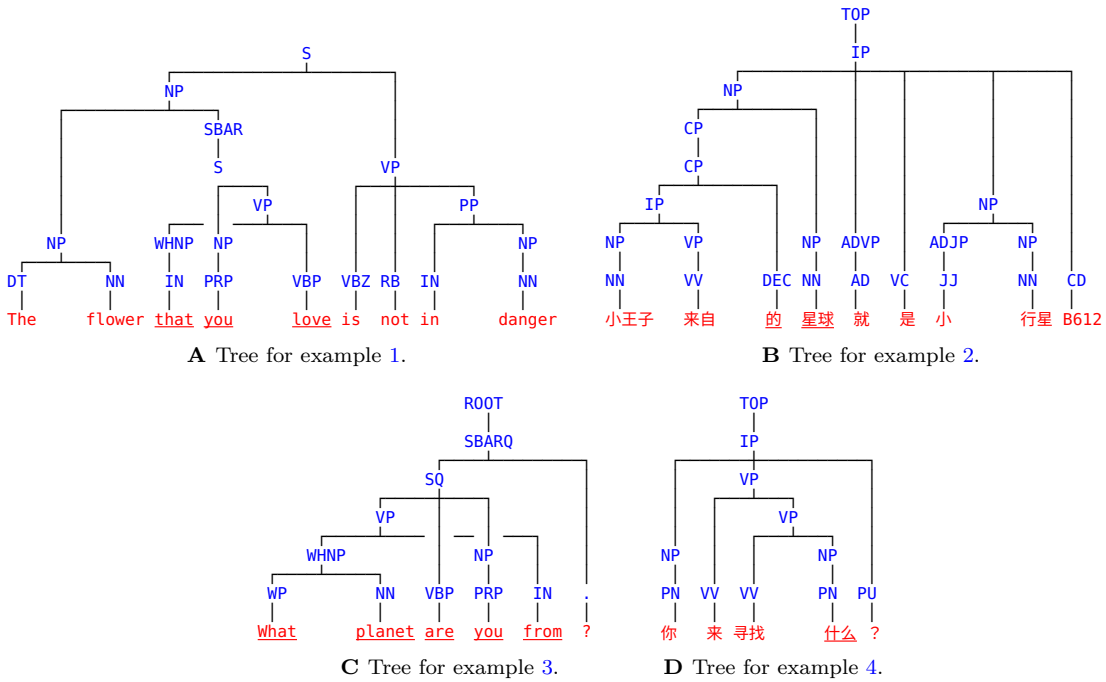
(1) [The flower]$_i$ that you love $__{}_i$ is not in danger

(2) 小王子 来自 $__{}_i$ 的 [星球]$_i$ 就 是 小 行星 B612
the little prince come from $__{}_i$ DE planet$_i$ exactly is little planet B612

The planet that the little prince came from is asteroid B612

(3) [What planet]$_i$ are you from $___i$?

(4) $___i$ 你 来 寻找 [什么]$_i$?
$___i$ you come search what$_i$?
What did you come to search for?



A Tree for example 1. **B** Tree for example 2.

C Tree for example 3. **D** Tree for example 4.

Figure 1: Example phrase structure trees. **A**, The example English sentence with an object-extracted relative clause given in example 1. **B**, The example Chinese sentence with an object-extracted relative clause given in example 2. **C**, The example English WH-question given in example 3. **D**, The example Chinese WH-question given in example 4. Underlined words indicate where increased cognitive demand is expected.

The calculation of the word-by-word ORC and WHQ metrics proceeded as just described, using tree-geometric properties to find the constructions. An example can be given in roughly the notation used by Richard Pito's 1992 `tgrep`, `tgrep2` (Rohde, 2005), and `tregex` (Levy & Andrew, 2006). Example 5 gives the `tregex` rule for identifying ORC constructions in English. For this rule, $A < B$ means that B is a child of A and $A << B$ means that B is a descendant of A. The rule looks for a WH tag that is dominated by an NP that contains an SBAR with an object gap. Notice how this corresponds to the structure in Fig. 1**A**.

(5) NP < (SBAR < (S < NP) < < /WH.*/)

After the filler is found, a gap that it fills or from which it has moved must the posited. The heuristic used to postulate the gap was to first find the sentence-like constituent that governs the filler (depending on the treebank, that can be SBAR, SQ, Srel, SENT or CP) and then posit the gap next to its head verb. The heads are found using head finding rules of Collins (2003) and their equivalents for Chinese borrowed from CoreNLP (Manning et al., 2014).

For English ORCs, each word between the filler and the gap site was annotated with a 1. These are the words which are underlined in Fig. 1**A**. For Chinese ORCs, each word between the gap site and the end of the filler was annotated with a 1. These are the words which are underlined in Fig. 1**B**. For English WHQs, each word from the WH-word to the gap site was annotated with a 1. These are the words which are underlined in Fig. 1**C**. For Chinese WHQs, each word in the WH-phrase was annotated with a 1. These are the words which are underlined in Fig. 1**D**. For these two LDD metrics, all other words were annotated with a value of 0. The number of identified LDD constructions for each language is given in Table 1.

	WH-question	Object-relative	Total
Chinese	85	33	118
English	77	32	109

Table 1: Long-distance dependency attestations in *The Little Prince*. The number of WH-question attestations identified in Chinese and English, the number of object-extracted relative clause attestations identified in Chinese and English, and the total number of long-distance dependencies identified.

Additionally, GPT2-caliber autoregressive transformer language models—one for each language—were used to calculate word-by-word surprisal. Surprisal has been shown to be associated with activity in the language network (Brennan et al., 2016; Henderson et al., 2016) and was included to control for incremental sequential processing. Furthermore, word-by-word lexical surprisal is particularly pertinent to Chinese WH-constructions as comprehenders in this language do not know that a WH-construction is coming up. This is in contrast to English, in which WH-constructions are fronted and immediately recognized by the perceiver at the beginning of the sentence. As measured by number of parameters, the two models are reasonably similar in complexity. For English, GPT2 (Radford et al., 2019, 1.5 billion parameters and trained on 40 gigabytes of data) was used. For Chinese, Chinese GPT2 (Zhao et al., 2019, 102 million parameters and trained on 14 gigabytes of data) was used. The Chinese model was accessed via HuggingFace Transformers (Wolf et al., 2019).

2.3 Experimental design and statistical analysis

2.3.1 GLM analysis

The GLM analyses were carried out using `Nilearn` (Abraham et al., 2014; Pedregosa et al., 2011, version 0.9.1), a package for the statistical analysis of neuroimaging data in Python. Separate analyses were carried out for Chinese and English. Regressors of non-interest included spoken word-rate, log lexical frequency, speaker pitch, and root mean squared amplitude of the speaker narration (RMS). These coregressors served to ensure that any results of interest were not the result of low-level language processing (cf. Bullmore et al. 1999; Lund et al. 2006). Word-by-word regressors of non-interest included the bottom-up processing metric and large language model surprisal. The regressors of interest were the word-by-word object-relative and WH-question metrics.

All predictors were convolved with the SPM canonical hemodynamic response function. Following convolution, the bottom-up, surprisal, and long-distance dependency word-by-word regressors were orthogonalized with respect to word-rate to remove correlations resulting from their common timing. Before modeling, the predictors were standardized (shifted to mean 0 and scaled to standard deviation 1) by scanning session/storybook section.

Throughout the GLM analyses, a liberal cortical mask calculated using the 1,000 Buckner et al. (2011) and Yeo et al. (2011) participants was applied (`https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011`). At the first level of the GLM, linear models were fit to the voxel BOLD timecourses. The Wilkinson-Rogers formula is given in 6, where `word_rate` marks the offset of every word with a 1, `RMS` is root mean squared amplitude of the spoken narration and is marked every 10 ms, `freq` marks the offset of every word with its log lexical frequency, `f0` is the speaker pitch and is marked every 10 ms, `gpt2_surprisal` marks the offset of every word with how surprised the language model is that it encounters the word given the preceding context, `bottom_up` marks the offset of every word with the syntactic complexity metric at that word, `WH` marks the offset of every word with the binary WHQ regressor value at that word, and `object_relative` marks the offset of every word with the binary ORC regressor value at that word.

(6) $$BOLD \sim 1 + word_rate + RMS + freq + f0 + gpt2_surprisal + bottom_up + WH + object_relative$$

The nine scanning sessions/storybook sections were used to compute fixed effects for each participant. The participant coefficient maps were entered into the second level, where one-sample t-tests were performed. An 8 mm full width at half maximum Gaussian smoothing kernel was applied to counteract inter-participant anatomical variation. The presented Chinese and English second-level maps are z-valued and thresholded with an expected false discovery rate (FDR) = .05 and a cluster threshold of

100 voxels. The presented intersection maps are the voxel-level intersections of the FDR thresholded Chinese and English maps with no cluster thresholding. Nothing of interest survived a $p < .05$ Bonferroni correction. The MNI2TAL tool from the Yale BioImage Suite (Lacadie et al., 2008a,b, version 1.4) was referenced for brain region and Brodmann area labels.

2.3.2 Bayesian ROI analysis

Bayesian mixed-effects linear models were fit to the joint Chinese and English data using the `brms` (Bürkner, 2017, version 2.16.3) package in R (R Core Team, 2020, version 4.0.0). Due to the computational cost of fitting each model, it would be impractical to fit a model for each voxel. For this reason, models were fit to extracted BOLD time series from selected ROIs of the Human Connectome Project Multi-Modal Parcellation 1 (Glasser et al., 2016), a cortical atlas consisting of 360 parcellations (180 in each hemisphere). The Human Connectome Project Multi-Modal Parcellation 1 was selected for its spatial resolution and for being well known. The volumetric version of the atlas was used, which was prepared by Dianne Patterson (`https://neuroimaging-core-docs.readthedocs.io/en/latest/pages/atlases.html`). Based on the mass univariate results, ROIs were selected encompassing the bilateral extended language network: lateral and medial prefrontal lobe, lateral temporal lobe, and lateral and medial parietal lobe.

The predictors and their preparation were the same as described in 2.2 and 2.3.1. Additionally, the dependent BOLD data were grouped by language, participant, and scanning session and then standardized. Given this preprocessing, the fitted models had population-level intercept 0 and population-level standard deviation 1; by-language, by-participant, and by-session, group-level intercepts would have been 0 if they were estimated. Random intercepts were not estimated in order to reduce computation time. By-language, population-level interaction effects for the WH and object-relative metrics were included. That is, separate coefficient estimates could be extracted for each language, the posterior distributions of which can be compared. The fitted models were maximally random with respect to participant. That is, each applicable population-level effect was also included as a by-participant group-level effect. The Wilkinson-Rogers formula for the regression model is given in 7. Different from 6, in 7, `language` is included as a factor because the Chinese and English data are jointly modeled, `WH:language` specifies an interaction term between `WH` and `language`, `obj_relative:language` specifies an interaction term between `obj_relative` and `language`, and the applicable population-level effects are included as by-participant random effects.

(7) $$BOLD \sim 1 + word_rate + RMS + freq + f0 + gpt2_surprisal + bottomup + WH + obj_relative + language + WH : language + obj_relative : language + (0 + word_rate + RMS + freq + f0 + gpt2_surprisal + bottomup + WH + obj_relative | participant)$$

Bayesian models require the specification of prior distributions over parameter values. For the population-level intercept, a normal distribution with mean 0 and standard deviation 1 was used. For the population-level effects, a normal distribution with mean 0 and standard deviation 2.5 was used. For the residual standard deviation, a half student-t with 3 degrees of freedom, scale parameter 2.5, and location parameter 0 was used. For the standard deviations of group-level effects, an exponential distribution with rate parameter 1 was used. Lastly, for the correlations of group-level effects, the default prior was retained which specifies that, a priori, all correlation matrices are equally likely.

The models were fit with 4 chains, with each chain running for 2,000 iterations. Each chain was run with a 1,000 iteration warmup. That is, only the final 1,000 iteration following the warmup period were kept for analysis. The fitted models returned no convergence warnings and Rhat values for all estimated parameters for all models were < 1.05.

Having fitted models which provide long-distance dependency regressor coefficients for each language, the estimated slopes were probed for similarity with Bayesian hypothesis testing. By positive, it is meant that a coefficient estimate is positive and that the 95% credible interval excludes 0. By negative, it is meant that a coefficient estimate is negative and that the 95% credible interval excludes 0. By similar, it is meant that a Bayesian hypothesis test for equality returns a Bayes factor > 3.2 (Kass & Raftery, 1995). For both the WH-construction coefficient and the object-relative coefficient, ROIs were categorized based on coefficient direction and posterior distribution similarity: positive in both Chinese and English and similar, negative in both Chinese and English and similar, or neither of these.

2.4 Data and code availability

The fMRI data are available through the associated OpenNeuro repository (https://openneuro.org/datasets/ds003643/versions/2.0.1). The analysis code is available at https://github.com/dgd45125/LPPxLDDxCN-EN.

3 Results

3.1 Behavioral results

Participants answered four four-choice comprehension questions after each section (36 questions in total). Participants performed well with a mean accuracy of 86.4% (SD=2.7) for Chinese participants and 89.5% (SD=3.8) for English participants.

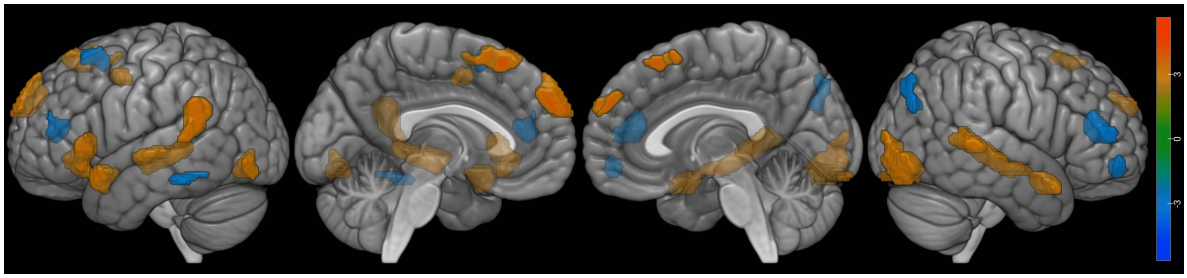
3.2 GLM

3.2.1 WH-question processing

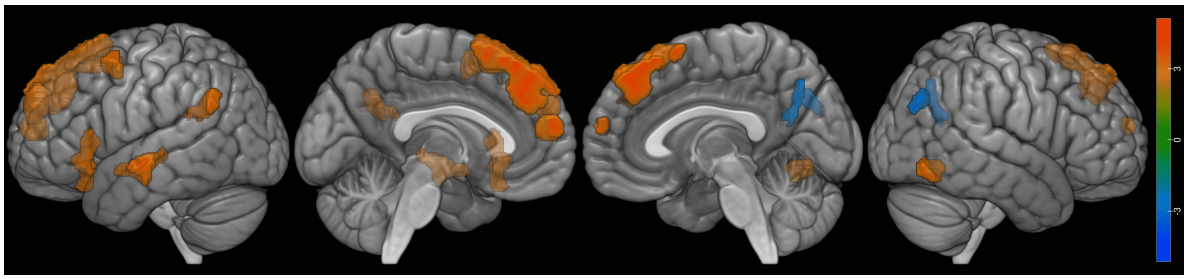
For the Chinese participants, WHQ processing was associated with an increase in activation in the bilateral superior temporal gyrus, bilateral middle temporal gyrus, and bilateral temporal pole, left inferior frontal gyrus, including the pars triangularis and pars orbitalis, bilateral medial frontal lobe, left premotor cortex, and bilateral occipital lobe. Decreases in activation were seen in bilateral middlle frontal lobe, left premotor cortex, left middle temporal gyrus, and the right angular gyrus. These increases and decreases in activation can be seen in Fig. 2**A** and more detail can be found in Table 2.

For the English participants, WHQ processing was associated with an increase in activation in the left middle and superior temporal gyri, left inferior frontal gyrus, including the pars triangularis and pars orbitalis, bilateral medial frontal lobe, left premotor cortex, left superior frontal lobe, left angular gyrus, and the right fusiform gyrus. Decreases in activation were seen in the right angular gyrus and right posterior cingulate cortex. These increases and decreases in activation can be seen in Fig. 2**B** and more detail can be found in Table 3.

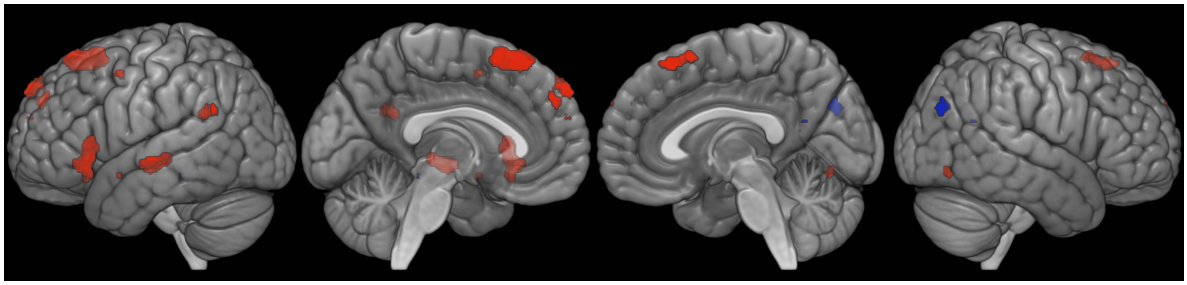
Commonly activated for both Chinese and English participants in response to WHQs were the left middle and superior temporal gyri, left pars orbitalis, bilateral medial frontal lobe, left superior frontal lobe, left premotor cortex, left angular gyrus, and right occipital lobe. Commonly deactivated were the left parahippocampal gyrus, right posterior cingulate cortex, and right angular gyrus. These commonalities can be seen in Fig. 2**C** and more detail can be found in Table 4.



A Chinese



B English



C Intersection

Figure 2: WH-question GLM results. **A**, Chinese WH-question coefficient GLM results (z-valued), thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. **B**, English WH-question coefficient GLM results (z-valued), thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. **C**, WH-question voxel-level intersection of the false discovery rate-thresholded Chinese and English maps with no cluster thresholding.

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
L Superior Temporal Gyrus (BA 22)	10056	-56.0	-18.0	-4.0	5.66
		-46.0	-28.0	-6.0	5.16
L Middle Temporal Gyrus (BA 21)		-60.0	-46.0	18.0	4.61
R Middle Temporal Gyrus (BA 21)	11168	50.0	-24.0	-10.0	5.16
R Superior Temporal Gyrus (BA 22)		46.0	-30.0	-2.0	4.92
R Middle Temporal Gyrus (BA 21)		44.0	-40.0	4.0	4.47
R Temporal Pole (BA 38)		50.0	10.0	-20.0	4.31
L Medial Frontal (BA 10)	4192	-6.0	64.0	26.0	5.00
L Temporal Pole (BA 38)	6032	-52.0	12.0	-14.0	4.54
L Pars Triangularis (BA 45)		-50.0	24.0	-2.0	4.34
L Pars Orbitalis (BA 47)		-40.0	26.0	-10.0	3.86
		-40.0	34.0	-6.0	3.63
R Occipital (BA 19)	6856	40.0	-80.0	-14.0	4.53
		26.0	-88.0	-18.0	3.86
R Occipital (BA 18)		38.0	-92.0	8.0	3.60
L Medial Frontal (BA 8)	3408	-2.0	26.0	54.0	4.49
L Medial Frontal (BA 6)		-8.0	4.0	58.0	3.50
L Occipital (BA 19)	2416	-48.0	-80.0	-12.0	4.29
		-40.0	-78.0	-12.0	4.11
		-30.0	-76.0	-8.0	3.39
L Frontal (BA 6)	1048	-36.0	-4.0	42.0	4.13
L Frontal (BA 46)	1488	-38.0	38.0	10.0	-4.30
L Frontal (BA 10)		-46.0	44.0	22.0	-3.15
R Frontal (BA 46)	2520	50.0	48.0	8.0	-4.10
		44.0	38.0	14.0	-4.03
R Angular Gyrus (BA 39)	1568	40.0	-80.0	40.0	-3.80
		50.0	-76.0	30.0	-3.52
R Frontal (BA 10)	920	40.0	50.0	-10.0	-3.75
L Frontal (BA 6)	1136	-24.0	12.0	56.0	-3.68
L Middle Temporal Gyrus (BA 21)	1192	-64.0	-36.0	-18.0	-3.52

Table 2: Chinese WH-question coefficient GLM results, thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
L Medial Frontal (BA 6)	15664	-8.0	18.0	58.0	5.14
L Medial Frontal (BA 8)		-10.0	36.0	32.0	5.09
		-2.0	32.0	54.0	4.52
R Medial Frontal (BA 6)		12.0	22.0	56.0	4.34
L Middle Temporal Gyrus (BA 21)	3720	-64.0	-14.0	-8.0	4.98
L Superior Temporal Gyrus (BA 22)		-50.0	-2.0	-10.0	3.79
L Pars Orbitalis (BA 47)	4824	-44.0	22.0	-6.0	4.39
L Pars Triangularis (BA 45)		-44.0	18.0	6.0	4.22
L Pars Orbitalis (BA 47)		-28.0	24.0	-12.0	4.21
		-32.0	22.0	-20.0	3.93
L Medial Frontal (BA 10)	2552	-4.0	56.0	16.0	4.22
L Angular Gyrus (BA 39)	2088	-50.0	-50.0	24.0	4.06
		-42.0	-42.0	22.0	3.78
L Frontal (BA 10)	2392	-22.0	50.0	30.0	4.03
		-16.0	56.0	20.0	3.52
R Fusiform Gyrus (BA 37)	2344	54.0	-66.0	-14.0	3.88
		44.0	-68.0	-10.0	3.81
L Frontal (BA 6)	1680	-40.0	4.0	60.0	3.64
		-46.0	2.0	50.0	3.55
R Angular Gyrus (BA 39)	1064	46.0	-68.0	32.0	-3.81
		52.0	-74.0	26.0	-3.48
R Posterior Cingulate Cortex (BA 31)	1144	10.0	-62.0	34.0	-3.73
		12.0	-60.0	26.0	-3.49

Table 3: English WH-question coefficient GLM results, thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Directionality
L Middle Temporal Gyrus (BA 21)	1360	-58.0	-20.0	-6.0	1.0
L Medial Frontal (BA 9)	8	-4.0	62.0	34.0	1.0
L Medial Frontal (BA 10)	16	-4.0	58.0	22.0	1.0
L Medial Frontal (BA 9)	8	-6.0	62.0	36.0	1.0
L Medial Frontal (BA 9)	352	-4.0	54.0	40.0	1.0
L Medial Frontal (BA 9)	288	-2.0	48.0	34.0	1.0
L Medial Frontal (BA 10)	16	-8.0	56.0	22.0	1.0
L Medial Frontal (BA 8)	2720	-2.0	22.0	58.0	1.0
L Frontal (BA 10)	8	-14.0	56.0	24.0	1.0
L Frontal (BA 6)	136	-38.0	2.0	48.0	1.0
L Frontal (BA 6)	8	-44.0	0.0	46.0	1.0
L Superior Temporal Gyrus (BA 22)	32	-50.0	2.0	-12.0	1.0
L Pars Orbitalis (BA 47)	1792	-44.0	22.0	-4.0	1.0
L Angular Gyrus (BA 39)	8	-52.0	-48.0	20.0	1.0
L Angular Gyrus (BA 39)	208	-56.0	-50.0	26.0	1.0
L Angular Gyrus (BA 39)	96	-62.0	-54.0	26.0	1.0
R Medial Frontal Cortex (BA 9)	16	4.0	62.0	32.0	1.0
R Occipital (BA 19)	224	44.0	-70.0	-10.0	1.0
L Parahippocampal Gyrus (BA 36)	8	-32.0	-34.0	-12.0	-1.0
L Parahippocampal Gyrus (BA 36)	16	-28.0	-34.0	-14.0	-1.0
L Parahippocampal Gyrus (BA 36)	8	-26.0	-34.0	-12.0	-1.0
R Posterior Cingulate Cortex (BA 23)	8	12.0	-52.0	14.0	-1.0
R Posterior Cingulate Cortex (BA 23)	56	18.0	-56.0	20.0	-1.0
R Angular Gyrus (BA 39)	368	48.0	-74.0	30.0	-1.0

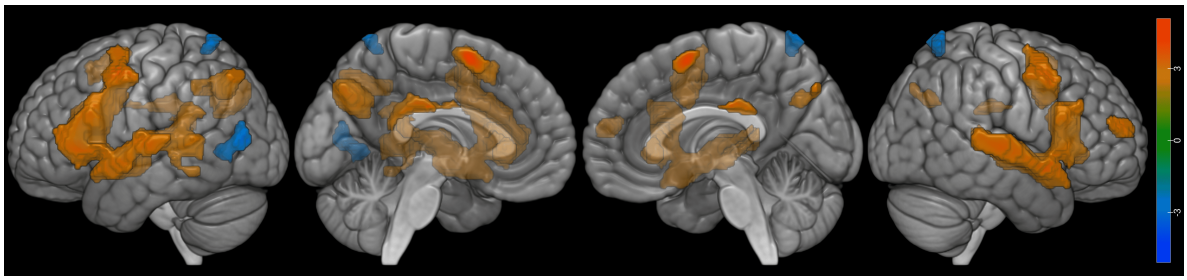
Table 4: WH-question voxel-level intersection of the false discovery rate-thresholded Chinese and English results with no cluster thresholding. Region, cluster size, MNI coordinates, and directionality.

3.2.2 Object-extracted relative clause processing

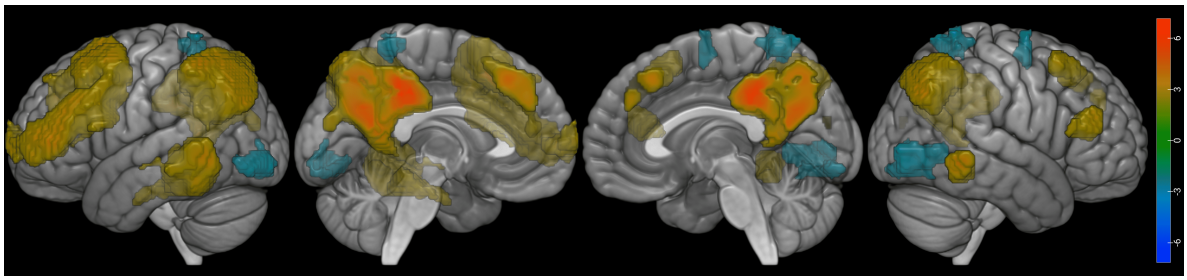
For the Chinese participants, ORC processing was associated with an increase in activation in bilateral superior temporal gyrus, right premotor cortex, left premotor cortex, extending into the inferior frontal gyrus, right pars opercularis, right middle frontal lobe, bilateral medial frontal lobe, bilateral precuneus, bilateral posterior cingulate cortex, and left angular gyrus. Decreases in activation were seen in the left occipital lobe and bilateral parietal lobe. These increases and decreases in activation can be seen in Fig. 3**A** and more detail can be found in Table 5.

For the English participants, ORC processing was associated with an increase in activation in the bilateral angular gyrus, bilateral posterior cingulate cortex, bilateral precuneus, bilateral medial frontal lobe, left frontal lobe, broadly, extending from premotor cortex to inferior frontal gyrus, to anterior middle frontal lobe, right superior frontal lobe, right middle temporal gyrus, and right inferior frontal gyrus. Decreases in activation were seen in the bilateral occipital lobe, bilateral superior parietal lobe, and right premotor cortex. These increases and decreases in activation can be seen in Fig. 3**B** and more detail can be found in Table 6.

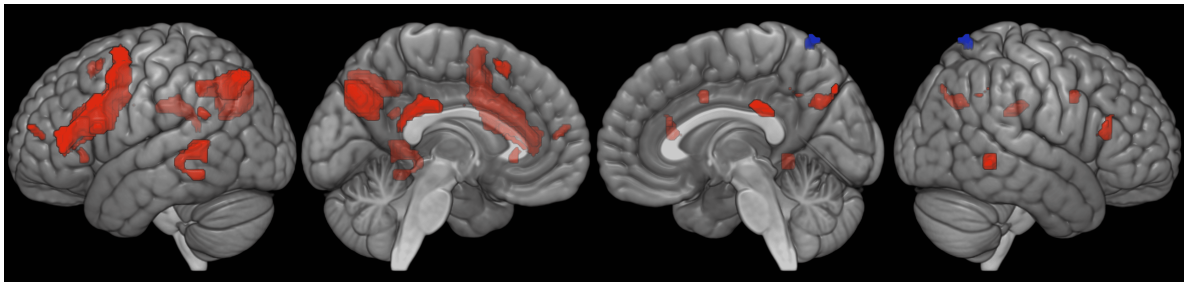
Commonly activated for both Chinese and English participants in response to ORCs were the left middle temporal gyrus, bilateral angular gyrus, right supramarginal gyrus, left medial frontal lobe, left inferior frontal gyrus, including the pars opercularis and pars triangularis, right pars opercularis, left superior frontal lobe, bilateral middle frontal lobe, left premotor cortex, bilateral posterior cingulate cortex, bilateral precuneus, and right fusiform gyrus. Commonly deactivated was right parietal lobe. The commonalities can be seen in 3C and more detail can be found in Table 7.



A Chinese



B English



C Intersection

Figure 3: Object-extracted relative clause GLM results. **A**, Chinese Object-extracted relative clause coefficient GLM results (z-valued), thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. **B**, English Object-extracted relative clause coefficient GLM results (z-valued), thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. **C**, Object-extracted relative clause voxel-level intersection of the false discovery rate-thresholded Chinese and English maps with no cluster thresholding.

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
R Superior Temporal Gyrus (BA 22)	25440	54.0	-2.0	-6.0	5.12
R Pars Opercularis (BA 44)		48.0	18.0	22.0	4.83
		54.0	14.0	-4.0	4.40
R Superior Temporal Gyrus (BA 22)		70.0	-24.0	4.0	4.33
L Frontal Cortex (BA 6)	44656	-52.0	2.0	52.0	5.02
		-48.0	10.0	28.0	4.76
L Primary Auditory Cortex (BA 41)		-60.0	-18.0	6.0	4.58
L Superior Temporal Gyrus (BA 22)		-50.0	0.0	-8.0	4.53
L Precuneus (BA 7)	5968	-8.0	-66.0	36.0	5.02
R Occipital (BA 19)		18.0	-64.0	34.0	3.64
R Precuneus (BA 7)		8.0	-74.0	38.0	3.42
L Posterior Cingulate Cortex (BA 23)	2880	-6.0	-30.0	26.0	4.64
R Posterior Cingulate Cortex (BA 23)		4.0	-24.0	26.0	4.28
L Medial Frontal (BA 6)	3240	-1.0	4.0	58.0	4.61
L Medial Frontal (BA 8)		-8.0	16.0	46.0	3.47
R Frontal (BA 6)	7520	52.0	8.0	46.0	4.51
R Frontal (BA 8)		42.0	8.0	36.0	4.11
L Angular Gyrus (BA 39)	4144	-38.0	-66.0	44.0	4.36
L Angular Gyrus (BA 39)		-38.0	-52.0	38.0	3.67
L Angular Gyrus (BA 39)		-46.0	-48.0	40.0	3.29
R Frontal (BA 10)	1272	42.0	54.0	16.0	3.93
L Occipital (BA 19)	2488	-52.0	-76.0	8.0	-3.88
		-48.0	-64.0	2.0	-3.57
L Parietal (BA 7)	920	-24.0	-58.0	68.0	-3.50
R Parietal (BA 7)	1472	20.0	-60.0	68.0	-3.46

Table 5: Chinese object-extracted relative clause coefficient GLM results, thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
L Angular Gyrus (BA 39)	79048	-34.0	-58.0	50.0	7.20
L Posterior Cingulate Cortex (BA 31)		-6.0	-38.0	40.0	6.85
		-10.0	-62.0	26.0	6.57
L Angular Gyrus (BA 39)		-32.0	-68.0	40.0	6.34
L Frontal (BA 6)	46408	-26.0	16.0	52.0	6.66
L Frontal (BA 8		-34.0	22.0	46.0	5.24
L Frontal (BA 10)		-18.0	66.0	6.0	5.13
L Frontal (BA 46)		-44.0	44.0	8.0	5.08
R Angular Gyrus (BA 39)	16024	38.0	-58.0	42.0	5.73
		48.0	-68.0	36.0	5.48
L Fusiform Gyrus (BA 37)	20480	-54.0	-50.0	-18.0	5.23
		-60.0	-48.0	-4.0	5.20
		-30.0	-36.0	-20.0	4.74
L Parahippocampal Gyrus (BA 36)		-38.0	-14.0	-26.0	3.48
R Middle Temporal Gyrus (BA 21)	3872	70.0	-44.0	-6.0	4.86
L Medial Frontal (BA 8)	6416	-4.0	28.0	44.0	4.64
		-8.0	36.0	42.0	4.49
L Anterior Cingulate Cortex (BA 32)		-10.0	44.0	14.0	3.31
R Frontal (BA 8)	4664	30.0	22.0	48.0	4.22
R Frontal (BA 46)	3600	48.0	28.0	14.0	4.14
R Frontal (BA 9)		52.0	38.0	24.0	3.62
		60.0	24.0	14.0	2.93
R Occipital (BA 19)	5904	40.0	-68.0	-2.0	-3.89
		38.0	-78.0	0.0	-3.71
		44.0	-58.0	-2.0	-3.35
		24.0	-84.0	-12.0	-3.25
L Occipital (BA 18)	3824	-28.0	-90.0	-10.0	-3.57
		-28.0	-84.0	6.0	-2.99
R Parietal (BA 7)	3560	14.0	-52.0	66.0	-3.52
		22.0	-50.0	74.0	-3.11
R Parietal (BA 5)		20.0	-40.0	68.0	-2.84
L Parietal (BA 7)	1912	-16.0	-48.0	66.0	-3.36
		-30.0	-46.0	72.0	-3.12
		-18.0	-38.0	62.0	-2.78
R Frontal (BA 6)	1400	24.0	-4.0	64.0	-3.18
		22.0	-4.0	74.0	-2.88

Table 6: Chinese object-extracted relative clause coefficient GLM results, thresholded with an expected false discovery rate = .05 and a cluster threshold of 100 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Directionality
L Middle Temporal Gyrus (BA 21)	2464	-58.0	-42.0	0.0	1.0
R Supramarginal Gyrus (BA 40)	16	44.0	-46.0	40.0	1.0
R Frontal (BA 9)	696	50.0	26.0	16.0	1.0
R Angular Gyrus (BA 39)	8	40.0	-48.0	38.0	1.0
R Angular Gyrus (BA 39)	40	36.0	-52.0	36.0	1.0
R Frontal (BA 8)	256	36.0	6.0	36.0	1.0
L Anterior Cingulate Cortex (BA 24)	8	-6.0	-6.0	30.0	1.0
L Medial Frontal (BA 8)	328	-6.0	18.0	52.0	1.0
L Posterior Cingulate Cortex (BA 23)	2304	-2.0	-32.0	28.0	1.0
L Precuneus (BA 7)	5296	-6.0	-68.0	34.0	1.0
L Frontal (BA 6)	104	-28.0	14.0	54.0	1.0
L Insula (BA 13)	208	-30.0	24.0	0.0	1.0
L Frontal (BA 10)	208	-34.0	52.0	14.0	1.0
L Pars Triangularis (BA 45)	8	-42.0	32.0	2.0	1.0
L pars Triangularis (BA 45)	8	-48.0	26.0	10.0	1.0
L Angular Gyrus (BA 39)	4144	-40.0	-62.0	42.0	1.0
L Angular Gyrus (BA 39)	224	-54.0	-48.0	24.0	1.0
L Pars Opercularis (BA 44)	14688	-44.0	16.0	28.0	1.0
R Pars Opercularis (BA 44)	8	56.0	20.0	12.0	1.0
R Fusiform Gyrus (BA 37)	608	62.0	-44.0	-4.0	1.0
R parietal (BA 7)	256	16.0	-60.0	68.0	-1.0

Table 7: Object-extracted relative clause voxel-level intersection of the false discovery rate-thresholded Chinese and English results with no cluster thresholding. Region, cluster size, MNI coordinates, and directionality.

3.3 Bayesian ROI

3.3.1 WH-question processing

In Chinese, WHQ processing was associated with an increase in activation in the bilateral superior and anterior temporal lobe, left inferior frontal gyrus, left superior frontal lobe, bilateral medial frontal lobe, bilateral premotor cortex, and bilateral superior parietal lobe. A decrease in activation was seen in bilateral superior frontal lobe, bilateral posterior parietal lobe, and right middle frontal lobe. The coefficients can be visualized in Figure 4**A**.

In English, WHQ processing was associated with an increase in activation in the left superior temporal lobe, left tomporoparietal junction/angular gyrus, right superior temporal gyrus, left frontal lobe, broadly, including premotor cortex, the inferior frontal gyrus, and superior frontal lobe, and bilateral dorso- and ventro-medial frontal lobe. A decrease in activation was seen in the precuneus, bilaterally. The coefficients can be visualized in Figure 4**B**.

As can be visualized in Figure 4**C**, both languages had positive and similar distributions for the WHQ coefficient in the left superior temporal lobe, right superior temporal gyrus, left inferior frontal gyrus, bilateral premotor cortex, left superior frontal lobe, and bilateral medial frontal lobe.

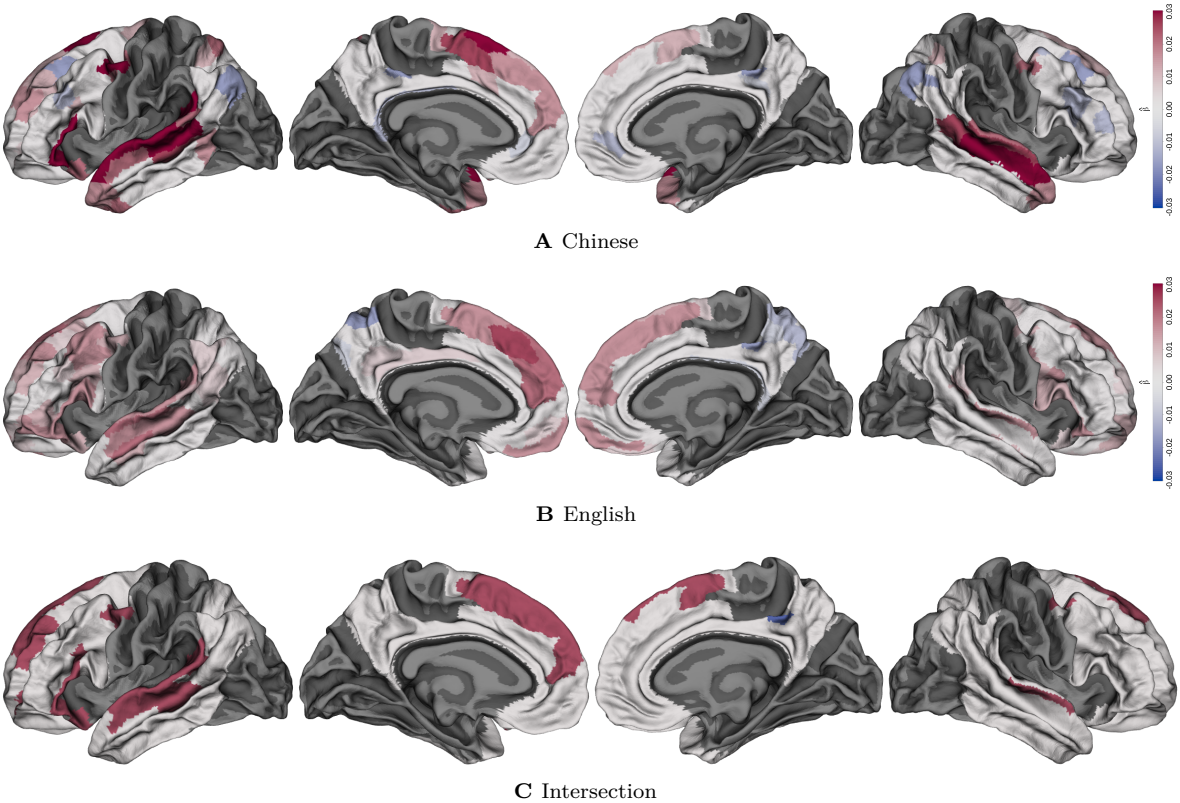


Figure 4: WH-question Bayesian ROI results. Dark areas were not analyzed. **A**, Chinese WH-question coefficient results. Marked ROIs indicate a posterior distribution for which the 95% credible interval excludes 0. **B**, English WH-question coefficient results. Marked ROIs indicate a posterior distribution for which the 95% credible interval excludes 0. **C**, ROI-level intersection of the Chinese and English results with the further condition that a Bayesian hypothesis test for equality return a Bayes factor > 3.2.

3.3.2 Object-extracted relatives clause processing

In Chinese, ORC processing was associated with an increase in activation in the left posterior and inferior temporal lobe, right posterior inferior temporal lobe, bilateral angular gyrus, bilateral medial frontal lobe, left (and right to a lesser extent) posterior cingulate and precuneus, bilateral frontal lobe, broadly, including premotor cortex and the inferior frontal gyrus, and left superior frontal lobe. The coefficients can be visualized in Fig. 5**A**.

In English, ORC processing was associated with an increase in activation in the left middle temporal lobe, bilateral parietal lobe, including the left angular gyrus, bilateral posterior cingulate cortex and precuneus, left dorso- and ventromedial frotal lobe, right ventromedial frontal lobe, and left frontal lobe, broadly, including premotor cortex, middle, and superior frontal lobe. The coefficients can be visualized in Fig. 5**B**.

As can be visualized in Fig. 5**C**, both languages had positive and similar distributions for the ORC

coefficient in the left posterior temporal lobe, left angular gyrus, left frontal lobe, broadly, including premotor cortex and middle frontal lobe, left medial frontal lobe, and left (and to a lesser extent right) posterior cingulate cortex and precuneus.

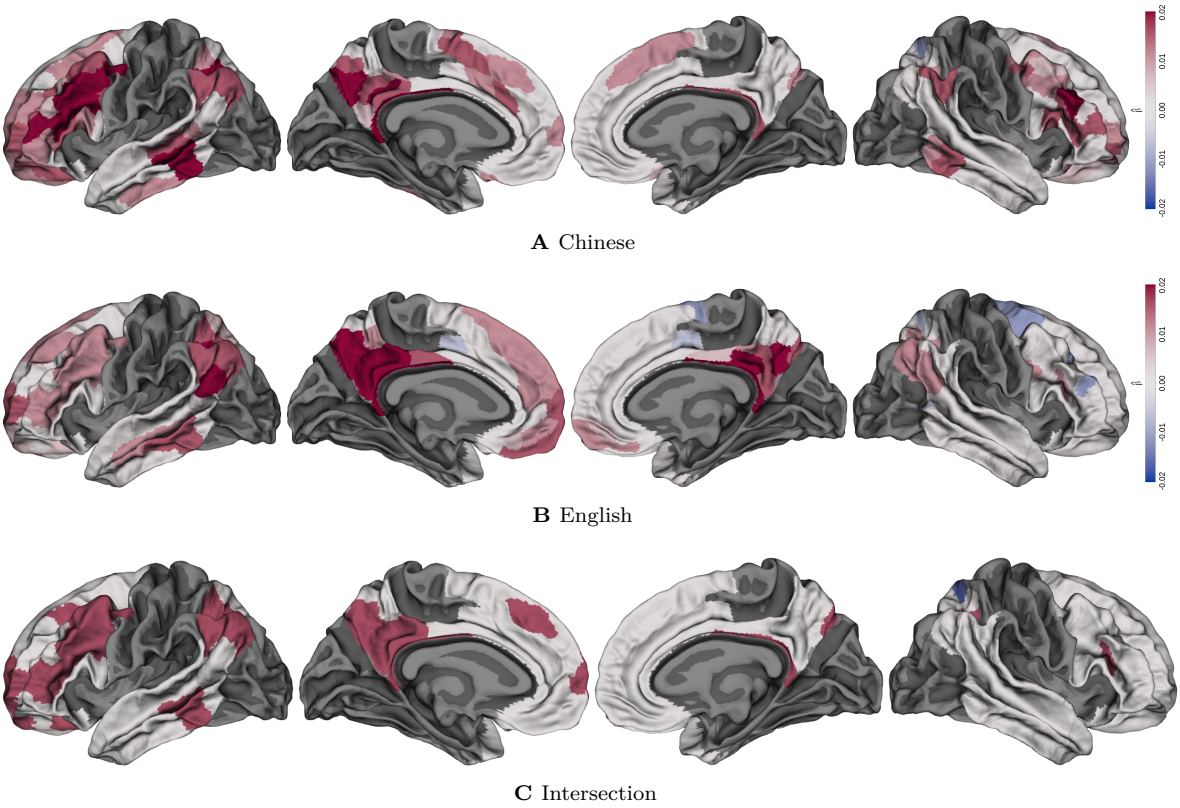


Figure 5: Object-extracted relative clause Bayesian ROI results. Dark areas were not analyzed. **A**, Chinese Object-extracted relative clause coefficient results. Marked ROIs indicate a posterior distribution for which the 95% credible interval excludes 0. **B**, English Object-extracted relative clause coefficient results. Marked ROIs indicate a posterior distribution for which the 95% credible interval excludes 0. **C**, ROI-level intersection of the Chinese and English results with the further condition that a Bayesian hypothesis test for equality return a Bayes factor > 3.2.

4 Discussion

The results of this study allow us to answer the questions asked at the beginning, which were: (1) whether the brain bases of ORC processing and WHQ processing are the same or different and (2) whether the brain bases of ORC processing and WHQ processing are the same or different for Chinese and English speakers. With respect to the first question, there were both similarities and differences in the patterns of brain activity associated with the processing of WHQs and ORCs. Common to both construction types was increased activation in left posterior middle/superior temporal lobe, left inferior frontal gyrus, left premotor cortex, and left medial frontal lobe. This is indicative of a common, core brain network for the processing of LDDs. The most notable differences were that ORC, but not WHQ processing was associated with bilateral increases in activation in the precuneus, posterior cingulate, and angular gyrus. It appears to be the case that ORC processing requires different and/or greater computational resources than WHQ processing. Interestingly, while processing both types of LDDs was associated with an increase in activation in the left middle temporal lobe and left inferior frontal gyrus, the WHQ-associated left temporal activation tended to be more anterior and superior than that of the ORC-associated activation and the WHQ-associated left inferior frontal activation tended to be more inferior than the ORC-associated activation. With respect to the second question, the degree of similarity between the results for the Chinese speakers and the English speakers was truly remarkable. The greatest differences, though, were that, in Chinese, WHQ processing seemed to additionally recruit right superior and middle temporal lobe and ORC processing seemed to additionally recruit left superior temporal lobe, patterns not seen in the associated English activations.

The brain areas implicated in LDD processing in this study subsume those which have traditionally been identified: left inferior frontal gyrus (Ben-Shachar et al., 2003, 2004; Caplan et al., 2008; Just et al., 1996; Stromswold et al., 1996) and left superior temporal lobe (Ben-Shachar et al., 2003, 2004; Caplan et al., 2008; Just et al., 1996). While other areas like left precuneus, left dorsolateral prefrontal cortex, and left premotor cortex have been associated with LDD processing (Caplan et al., 1985), further strong effects were found in the angular gyrus, medial frontal lobe, and posterior cingulate cortex. The most straightforward explanation for this difference is in the extension to naturalistic stimuli. One criticism raised toward naturalistic experiments is that there are a greater number of factors which can covary with the variable(s) of interest. While this aspect cannot be denied, steps were taken to alleviate this problem by including covariates of non-interest such as spoken word rate, word-by-word syntactic complexity, and word-by-word lexical surprisal. As has been persuasively argued (Armeni et al., 2017), the ability to study language processing in a naturalistic context outweighs this concern.

Another interpretation is that the processing of LDDs involves not just syntactic processing, as

traditionally thought, but higher-level linguistic processing at the level of narrative comprehension. As compared to previous experiments where there would be little to no coherence across the stimuli, in a naturalistic storybook text, it is possible for the WH-words or phrases in a WHQ (e.g., *which teacher*, *what planet*) to refer to previously introduced people, places, or things which play a narrative role outside of that sentence. Likewise for the nouns which ORCs modify (e.g. *the boy who the wolf ate*, *the flower that the man loved*). The unexpected regions that we identified have previously been implicated in narrative comprehension: the precuneus (Lerner et al., 2011; Maguire et al., 1999; Xu et al., 2005), posterior cingulate (Ferstl et al., 2008; Fletcher et al., 1995; Maguire et al., 1999), medial frontal lobe (Ferstl et al., 2008; Fletcher et al., 1995; Lerner et al., 2011), lateral frontal lobe (Lerner et al., 2011; Maguire et al., 1999; Xu et al., 2005), and angular gyrus (Lerner et al., 2011; Xu et al., 2005). Cross-linguistically, these regions have been shown to be able to decode latent space representations of translation-equivalent stories in a different language than the story presented to the participant (Dehghani et al., 2017). They have also demonstrated shared responses to narrative content when English and Russian speakers listen to translation-equivalent stories Honey et al. (2012).

One of the themes addressed in this paper is the study of the brain bases of language from a cross-linguistic perspective (Bornkessel-Schlesewsky & Schlesewsky, 2016; Kemmerer, 2016). This is a crucial endeavor for cognitive neuroscience as different languages have different structures and a comprehensive model of language processing in the brain must account for all languages. While previous work has demonstrated cross-linguistic similarity in patterns of activation elicited by general language comprehension (Honey et al., 2012; Malik-Moraleda et al., 2022), we advance this domain by investigating specific phenomena which occur differently in the two studied languages. As previously demonstrated, for example, in Italian vs. English phonological access (Paulesu et al., 2000), Chinese vs. English pitch contour processing (Gandour et al., 2003), and Chinese vs. Indo-European noun and verb representation (Li et al., 2004), it is in this sort of specificity that differences emerge, if they exist. More work needs to be done probing the typological dimensions along which languages differ.

As discussed in the Introduction, theoreticians consider LDDs to be a coherent class (Chomsky, 1977). What binds these phenomena together is a linguistic rule called *Move* (for more discussion regarding the cognitive status of this rule, see Phillips & Lewis, 2013; Stabler, 1983). *Move* is the mechanism by which the displaced elements come to be *Move*-ed away from the location where they contribute their meaning in a sentence, hence, the long-distance dependency. The results, here, though, demonstrate that the neural processing of two of the items in this class, WHQs and ORCs, are not as consistent as might be expected from such similarly considered phenomena. More work needs to be done probing why these two construction types elicit different patterns of activation. In this study, WHQs and ORCs were modeled as block phenomena due to the temporal resolution of fMRI. Future

work should separately model filler storage, filler maintenance, and filler retrieval/integration, as it may be along one or more of these dimensions that the neural processing of different LDD types differs. This may not be possible with fMRI, though, and neuroimaging modalities with greater temporal resolution like MEG may be needed.

In conclusion, we found a remarkable degree of cross-linguistic similarity in the brain activations associated with the processing of WH-question and object-extracted relative clause long-distance dependencies in Chinese and English. This is despite the fact that these constructions manifest striking typological differences and supports a model in which the brain bases of higher-level language processing are uniform across languages. It was also shown, in not just one, but both languages, that WH-questions and object-relative clauses processing elicit different patterns of activation. This serves to remind us that grammatical similarity need not translate into similarity of brain basis.

References

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., Gramfort, A., Thirion, B., & Varoquaux, G. (2014). Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, *8*, 14. https://doi.org/10.3389/fninf.2014.00014.

Armeni, K., Willems, R. M., & Frank, S. L. (2017). Probabilistic language models in cognitive neuroscience: Promises and pitfalls. *Neuroscience & Biobehavioral Reviews*, *83*, 579–588. https://doi.org/10.1016/j.neubiorev.2017.09.001.

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*, 433–440. https://doi.org/10.1111/1467-9280.01459.

Ben-Shachar, M., Palti, D., & Grodzinsky, Y. (2004). Neural correlates of syntactic movement: converging evidence from two fmri experiments. *Neuroimage*, *21*, 1320–1336. https://doi.org/10.1016/j.neuroimage.2003.11.027.

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016). The importance of linguistic typology for the neurobiology of language. *Linguistic Typology*, *20*, 615–621. https://doi.org/10.1515/lingty-2016-0032.

Brennan, J., Nir, Y., Hasson, U., Malach, R., Heeger, D. J., & Pylkkänen, L. (2012). Syntactic structure building in the anterior temporal lobe during natural story listening. *Brain and language*, *120*, 163–173. https://doi.org/10.1016/j.bandl.2010.04.002.

Brennan, J. R., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. T. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and language*, *157*, 81–94. https://doi.org/10.1016/j.bandl.2016.04.008.

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of neurophysiology*, *106*, 2322–2345. https://doi.org/10.1152/jn.00339.2011.

Bullmore, E. T., Brammer, M. J., Rabe-Hesketh, S., Curtis, V. A., Morris, R. G., Williams, S. C., Sharma, T., & McGuire, P. K. (1999). Methods for diagnosis and treatment of stimulus-correlated motion in generic brain activation studies using fMRI. *Human Brain Mapping*, *7*, 38–48. https://doi.org/10.1002/(SICI)1097-0193(1999)7:1<38::AID-HBM4>3.0.CO;2-Q.

Bürkner, P.-C. (2017). brms: An r package for bayesian multilevel models using stan. *Journal of statistical software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01.

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*, 117–175. https://doi.org/10.1016/0010-0277(85)90048-4.

Caplan, D., Stanczak, L., & Waters, G. (2008). Syntactic and thematic constraint effects on blood oxygenation level dependent signal correlates of comprehension of relative clauses. *Journal of Cognitive Neuroscience*, *20*, 643–656. https://doi.org/10.1162/jocn.2008.20044.

Chaves, R. P., & Putnam, M. T. (2021). *Unbounded dependency constructions: Theoretical and experimental perspectives*. Oxford University Press, USA. 10.1093/oso/9780198784999.001.0001.

Cheng, L., & Corver, N. (2006). Lines of inquiry into wh-movement. In *Wh-movement: moving on* (pp. 1–18). Mit Press. https://doi.org/10.7551/mitpress/7197.003.0004.

Chomsky, N. (1977). On wh-movement. In P. W. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal Syntax* (pp. 71–133).

Coavoux, M. (2021). BERT-proof syntactic structures: Investigating errors in discontinuous constituency parsing. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (pp. 3259–3272). Online: Association for Computational Linguistics. https://aclanthology.org/2021.findings-acl.288. 10.18653/v1/2021.findings-acl.288.

Collins, M. (2003). Head-driven statistical models for natural language parsing. *Computational linguistics*, *29*, 589–637.

Cox, R. W. (1996). Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*, 162–173. https://doi.org/10.1006/cbmr.1996.0014.

Dehghani, M., Boghrati, R., Man, K., Hoover, J., Gimbel, S. I., Vaswani, A., Zevin, J. D., Immordino-Yang, M. H., Gordon, A. S., Damasio, A. et al. (2017). Decoding the neural representation of story meanings across languages. *Human brain mapping*, *38*, 6096–6106. https://doi.org/10.1002/hbm.23814.

Dunagan, D., Zhang, S., Li, J., Bhattasali, S., Pallier, C., Whitman, J., Yang, Y., & Hale, J. (2022). Neural correlates of semantic number: A cross-linguistic investigation. *Brain and Language*, *229*, 105110. https://doi.org/10.1016/j.bandl.2022.105110.

Eisenstein, J. (2019). *Introduction to natural language processing*. MIT press.

Ferstl, E. C., Neumann, J., Bogler, C., & Von Cramon, D. Y. (2008). The extended language network: a meta-analysis of neuroimaging studies on text comprehension. *Human brain mapping*, *29*, 581–593. https://doi.org/10.1002/hbm.20422.

Fletcher, P. C., Happe, F., Frith, U., Baker, S. C., Dolan, R. J., Frackowiak, R. S., & Frith, C. D. (1995). Other minds in the brain: a functional imaging study of "theory of mind" in story comprehension. *Cognition*, *57*, 109–128. https://doi.org/10.1016/0010-0277(95)00692-R.

Gandour, J., Dzemidzic, M., Wong, D., Lowe, M., Tong, Y., Hsieh, L., Satthamnuwong, N., & Lurito, J. (2003). Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language*, *84*, 318–336. https://doi.org/10.1016/S0093-934X(02)00505-9.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M. et al. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*, 171–178. https://doi.org/10.1038/nature18933.

Hale, J. T., Campanelli, L., Li, J., Bhattasali, S., Pallier, C., & Brennan, J. R. (2022). Neurocomputational models of language processing. *Annual Review of Linguistics*, *8*, 427–446. https://doi.org/10.1146/annurev-linguistics-051421-020803.

Henderson, J. M., Choi, W., Lowder, M. W., & Ferreira, F. (2016). Language structure in the brain: A fixation-related fmri study of syntactic surprisal in reading. *Neuroimage*, *132*, 293–300. https://doi.org/10.1016/j.neuroimage.2016.02.050.

Honey, C. J., Thompson, C. R., Lerner, Y., & Hasson, U. (2012). Not lost in translation: Neural responses shared across languages. *Journal of Neuroscience*, *32*, 15277–15283. https://doi.org/10.1523/JNEUROSCI.1800-12.2012.

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in chinese. *Cognition*, *90*, 3–27. https://doi.org/10.1016/S0010-0277(03)00124-0.

Huang, C. J. (1982). Move wh in a language without wh movement, . https://doi.org/10.1515/tlir.1982.1.4.369.

Huang, C.-R. (1992). Reverse long-distance dependency and functional uncertainty-the interpretation of mandarin questions. In *Proceedings of the Korean Society for Language and Information Conference* (pp. 111–120). Korean Society for Language and Information. https://koreascience.kr/article/CFKO199211920165879.page.

Jäger, L., Chen, Z., Li, Q., Lin, C.-J. C., & Vasishth, S. (2015). The subject-relative advantage in chinese: Evidence for expectation-based processing. *Journal of Memory and Language*, *79*, 97–120. https://doi.org/10.1016/j.jml.2014.10.005.

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. DOI:10.1126/science.274.5284.114.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*, 773–795. https://www.tandfonline.com/doi/abs/10.1080/01621459.1995.10476572.

Kemmerer, D. (2016). Do language-specific word meanings shape sensory and motor brain systems? the relevance of semantic typology to cognitive neuroscience. *Linguistic Typology*, *20*, 623–634. https://doi.org/10.1515/lingty-2016-0033.

Kitaev, N., Cao, S., & Klein, D. (2019). Multilingual constituency parsing with self-attention and pre-training. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (pp. 3499–3505). Florence, Italy: Association for Computational Linguistics. https://aclanthology.org/P19-1340. 10.18653/v1/P19-1340.

Kundu, P., Inati, S. J., Evans, J. W., Luh, W.-M., & Bandettini, P. A. (2012). Differentiating bold and non-bold signals in fMRI time series using multi-echo EPI. *Neuroimage*, *60*, 1759–1770. https://doi.org/10.1016/j.neuroimage.2011.12.028.

Lacadie, C., Fulbright, R., Arora, J., Constable, R., & Papademetris, X. (2008a). Brodmann areas defined in mni space using a new tracing tool in bioimage suite. In *Proceedings of the 14th annual meeting of the organization for human brain mapping*.

Lacadie, C. M., Fulbright, R. K., Rajeevan, N., Constable, R. T., & Papademetris, X. (2008b). More accurate Talairach coordinates for neuroimaging using non-linear registration. *Neuroimage*, *42*, 717–725. https://doi.org/10.1016/j.neuroimage.2008.04.240.

Lerner, Y., Honey, C. J., Silbert, L. J., & Hasson, U. (2011). Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. *Journal of Neuroscience*, *31*, 2906–2915. https://doi.org/10.1523/JNEUROSCI.3684-10.2011.

Levine, R., & Hukari, T. E. (2006). *The unity of unbounded dependency constructions*. Stanford University Center for the Study of Language and Information.

Levy, R., & Andrew, G. (2006). Tregex and tsurgeon: tools for querying and manipulating tree data structures. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. Genoa, Italy: European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2006/pdf/513_pdf.pdf.

Li, J., Bhattasali, S., Zhang, S., Franzluebbers, B., Luh, W. M., Spreng, R. N., Brennan, J. R., Yang, Y., Pallier, C., & Hale, J. T. (2022). Le Petit Prince multilingual naturalistic fMRI corpus. *Scientific Data*, *9*, 530. https://doi.org/10.1038/s41597-022-01625-7.

Li, J., Hale, J., & Pallier, C. (2021). Le Petit Prince: A multilingual fMRI corpus using ecological stimuli. 10.18112/openneuro.ds003643.v1.0.3.

Li, P., Jin, Z., & Tan, L. H. (2004). Neural representations of nouns and verbs in chinese: an fmri study. *Neuroimage*, *21*, 1533–1541. https://doi.org/10.1016/j.neuroimage.2003.10.044.

Lund, T. E., Madsen, K. H., Sidaros, K., Luo, W., & Nichols, T. E. (2006). Non-white noise in fMRI: Does modelling have an impact? *Neuroimage*, *29*, 54–66. https://doi.org/10.1016/j.neuroimage.2005.07.005.

Maguire, E. A., Frith, C. D., & Morris, R. (1999). The functional neuroanatomy of comprehension and memory: the importance of prior knowledge. *Brain*, *122*, 1839–1850. https://doi.org/10.1093/brain/122.10.1839.

Malik-Moraleda, S., Ayyash, D., Gallée, J., Affourtit, J., Hoffman, M., Mineroff, Z., Jouravlev, O., & Fedorenko, E. (2022). An investigation across 45 languages and 12 language families reveals a universal language network. *Nature Neuroscience*, . https://doi.org/10.1038/s41593-022-01114-5.

Manning, C. D., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S. J., & McClosky, D. (2014). The Stanford CoreNLP natural language processing toolkit. In *Association for Computational Linguistics (ACL) System Demonstrations* (pp. 55–60). http://www.aclweb.org/anthology/P/P14/P14-5010.

Nederhof, M.-J., & Yli-Jyrä, A. (2021). A derivational model of discontinuous parsing. *Information and Computation*, *279*, 104619. https://doi.org/10.1016/j.ic.2020.104619.

Paulesu, E., McCrory, E., Fazio, F., Menoncello, L., Brunswick, N., Cappa, S. F., Cotelli, M., Cossu, G., Corte, F., Lorusso, M. et al. (2000). A cultural effect on brain function. *Nature Neuroscience*, *3*, 91–96. https://doi.org/10.1038/71163.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V. et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830. https://doi.org/10.48550/arXiv.1201.0490.

Phillips, C., & Lewis, S. (2013). Derivational order in syntax: Evidence and architectural consequences. *Studies in Linguistics*, *6*.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria. https://www.R-project.org/.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I. et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*, 9.

Rohde, D. L. (2005). Tgrep2 user manual. *Unpublished manuscript*, . https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.70.9846&rep=rep1&type=pdf.

Stabler, E. P. (1983). How are grammers represented? *Behavioral and Brain Sciences*, *6*, 391–402. https://doi.org/10.1017/S0140525X00016678.

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and language*, *52*, 452–473. https://doi.org/10.1006/brln.1996.0024.

Wanner, H., & Maratsos, M. (1978). An atn approach to comprehension. In M. Halle, J. Bresnan, & G. Miller (Eds.), *Linguistic theory and psychological reality*. M.I.T. Press.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M. et al. (2019). Huggingface's transformers: State-of-the-art natural language processing. *arXiv preprint arXiv:1910.03771*, . https://doi.org/10.48550/arXiv.1910.03771.

Xiang, M., Wang, S., & Cui, Y. (2015). Constructing covert dependencies—the case of mandarin wh-in-situ dependency. *Journal of Memory and Language*, *84*, 139–166.

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: emergent features of word, sentence, and narrative comprehension. *Neuroimage*, *25*, 1002–1015. https://doi.org/10.1016/j.neuroimage.2004.12.013.

Yeo, B. T., Krienen, F. M., Sepulcre, J., Sabuncu, M. R., Lashkari, D., Hollinshead, M., Roffman, J. L., Smoller, J. W., Zöllei, L., Polimeni, J. R. et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of neurophysiology*, . https://doi.org/10.1152/jn.00338.2011.

Yun, J., Chen, Z., Hunter, T., Whitman, J., & Hale, J. (2015). Uncertainty in processing relative clauses across east asian languages. *Journal of East Asian Linguistics*, *24*, 113–148. https://doi.org/10.1007/s10831-014-9126-6.

Zhao, Z., Chen, H., Zhang, J., Zhao, X., Liu, T., Lu, W., Chen, X., Deng, H., Ju, Q., & Du, X. (2019). Uer: An open-source toolkit for pre-training models. *EMNLP-IJCNLP 2019*, (p. 241). https://doi.org/10.48550/arXiv.1909.05658.