Neural correlates of object-extracted relative clause processing across English and Chinese

Donald Dunagan[1], Miloš Stanojević[2], Maximin Coavoux[3], Shulin Zhang[1], Shohini Bhattasali[4], Jixing Li[5], Jonathan Brennan[6], and John Hale[1]

[1]University of Georgia

[2]DeepMind

[3]Université Grenoble Alpes

[4]University of Toronto Scarborough

[5]City University of Hong Kong

[6]University of Michigan

Neural correlates of object-extracted relative clause processing across English and Chinese

Abstract

An under-studied facet of the neurobiology of language is whether, and if so, when, the brain bases of language processing differ across languages. English and Mandarin Chinese present an opportunity to investigate this with object-extracted relative clauses (ORCs), as these constructions exhibit radical typological differences in these two languages. While ORCs have been extensively studied in both languages, it is important that previous experimental results be validated and extended using naturalistic stimuli. ORCs were analyzed in fMRI data collected while English and Chinese participants listened to naturalistic, translation-equivalent stories. A general linear model analysis was performed, calculating effects for ORC processing and the associated contrast between the two languages A voxel-level intersection was calculated between the English and Chinese ORC effects to identify common regions of selectively increased activation. In English, ORC processing was associated with increased activation in the left posterior temporal lobe, angular gyrus, inferior frontal gyrus, precuneus, and posterior cingulate cortex. Similar results were found for Chinese, with the addition of: bilateral mid/anterior superior temporal gyrus and right inferior frontal gyrus. Contrasting the two languages, greater activation was found for Chinese participants than English participants in bilateral superior temporal gyrus and bilateral primary auditory cortex. The results implicate a common collection of brain regions for ORC processing. However, the additionally activated regions for Chinese participants are interpreted as subserving supplemental processing related to structural disambiguation. The results indicate that naturalistic language processing recruits extended language network regions in order to incorporate incoming linguistic information with previous context.

Introduction

An under-studied facet of the neurobiology of language is whether, and if so, when, the brain bases of language processing differ across languages (Bornkessel-Schlesewsky & Schlesewsky, 2016b). With respect to this, an intuitive avenue for investigation is to identify a dimension along which languages differ typologically and compare associated neural activations. English and Mandarin Chinese present such an opportunity in the form of object-extracted relative clauses (ORCs), which differ in structure in these two languages. ORCs are also an interesting cross-linguistic test case because of how extensively they have been investigated, separately, in the English and Chinese language processing literature, often in the case of cleverly designed and carefully controlled experiments. While these sorts of experiments have attempted to shed light on, e.g., the syntactic bases of long-distance dependency processing, it is not the case that language processing happens exclusively in the laboratory. Indeed, it is important that the field of neurolinguistics continue to investigate phenomena in a *naturalistic* context (Hasson et al., 2018; Hasson & Honey, 2012). This article contributes to the neurolinguistic literature by generalizing the brain bases of ORC processing in two ways. We: (1) generalize across languages; and (2) generalize to naturalistic stimuli. We analyze the brain activity associated with the processing of ORCs in English and Chinese using publicly available, naturalistic, parallel fMRI data: *The Little Prince Datasets* (Hale et al., 2022; J. Li et al., 2022; J. Li et al., 2021).

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 (see Chaves & Putnam, 2021, for a comprehensive overview). Object-extracted relative clauses are relative clauses where the head noun is the object of the relative clause verb. This is in contrast to subject-extracted relative clauses (SRCs), in which the head noun is the *subject* of the relative clause verb. In a relative clause, the head noun is known as the *filler*, while the position where the head

noun 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 its 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*. As compared to English, in which the head noun occurs before the relative clause, in Chinese, relative clauses are prenominal. That is, the relative clause occurs before the head noun which it modifies. This is exemplified by the ORC in example 2, where *DE* is a relativizing marker and *planet* is the filler. It is this typological, structural difference which makes the comparison of English and Chinese an intriguing prospect.

(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

In English, results from reading time (Gibson et al., 2005; King & Just, 1991; Staub, 2010; Traxler et al., 2002; Traxler et al., 2005), question answering (Wanner & Maratsos, 1978), event related potential (ERP; King & Kutas, 1995; Müller et al., 1997), and neuroimaging (Caplan et al., 1985; Just et al., 1996; Stromswold et al., 1996) investigations converge on the conclusion that ORCs are more difficult for the human sentence processor than SRCs, thus making them a prime topic for investigation. In Chinese, however, things are not so clear. While some reading time (Chen et al., 2008; Hsiao & Gibson, 2003; Y. Lin & Garnsey, 2011; K. Xu et al., 2020a), ERP (Packard et al., 2011), and neuroimaging (K. Xu & Duann, 2020; K. Xu et al., 2020b) results indicate an ORC advantage (i.e., SRCs are more difficult to process than ORCs), other reading time (Jäger et al., 2015; C.-J. C. Lin & Bever, 2006, 2011; Vasishth et al., 2013), ERP (Xiong et al., 2019), and theoretical (Yun et al., 2015) results indicate a SRC advantage, which corresponds to the case in English. Further, some studies have found either no advantage (Lee & Chan, 2023;

Zhou et al., 2018, fMRI and reading time, respectively), that the advantage is different at different points in the relative clause (Bulut et al., 2018, ERP), or that the advantage changes depending on whether the relative clause is subject- or object-modifying (Xiong & Newman, 2021, fMRI). Rather than further comparing ORCs and SRCs, this article focuses on probing the processing of just ORCs as, in contrast to Chinese, in which the evidence is mixed, in English, this phenomenon is undoubtedly more difficult to process.

The choice to hold constant the grammatical relation of the extracted element reflects a design goal of the study: to compare and contrast the brain bases of processing a construction whose surface properties differ across languages. Alternative goals having to do with working memory or semantic processing have already been pursued in the literature as discussed above.

Previous neuroimaging work in English has associated ORC processing with activation in the left inferior frontal gyrus (IFG; Caplan et al., 2008; Just et al., 1996; Stromswold et al., 1996), right IFG (Just et al., 1996), left (posterior) middle/superior temporal gyrus ((p)M/STG; Caplan et al., 2008; Just et al., 1996), right pM/STG (Just et al., 1996), and left precuneus (Caplan et al., 2008). Previous neuroimaging work in Chinese has associated ORC processing with activation in the left IFG (Lee & Chan, 2023; Xiong & Newman, 2021), right IFG (Xiong & Newman, 2021), left pSTG (K. Xu et al., 2020b), right pSTG(K. Xu et al., 2020b), left mid M/STG (Lee & Chan, 2023), right mid M/STG (Lee & Chan, 2023), left anterior MTG (Xiong & Newman, 2021; K. Xu et al., 2020b), right anterior MTG (K. Xu et al., 2020b), left MTG broadly (Xiong & Newman, 2021), right MTG broadly (Xiong & Newman, 2021), left angular gyrus (AG; Xiong & Newman, 2021), right AG (Xiong & Newman, 2021), left premotor cortex (K. Xu et al., 2020b), right premotor cortex (Lee & Chan, 2023), left precuneus (Xiong & Newman, 2021), left medial frontal gyrus (Xiong & Newman, 2021), left fusiform gyrus (K. Xu et al., 2020b), left posterior cingulate cortex (K. Xu et al., 2020b), left temporal pole (Xiong &

Newman, 2021), and right temporal pole (Xiong & Newman, 2021).

Similarity and difference results in previous cross-linguistic neurolinguistic studies have hinged on the level of linguistic analysis. Previous work has found high degrees of activation similarity in the language network when Russian and English speakers listen to a translation-equivalent 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). These studies, while finding cross-linguistic similarity, have focused on language processing either holistically (Malik-Moraleda et al., 2022) or at a high level (Dehghani et al., 2017; Dunagan et al., 2022; Honey et al., 2012). In contrast, studies which have contrasted low-level linguistic phenomena like phonological access in Italian vs. English (Paulesu et al., 2000), pitch contour processing in Chinese vs. English (Gandour et al., 2003), and noun and verb representation in Chinese vs. Indo-European (P. Li et al., 2004) have identified differences in processing or representation. ORCs occur in an underexplored middle ground between low- and high-level language processing, and for this reason, their study from a cross-linguistic perspective is valuable.

While much previous work in both English and Chinese has examined ORC processing with carefully controlled stimuli, we examine ORC processing using naturalistic stimuli. As discussed by proponents of naturalistic stimuli (Hasson et al., 2018; Hasson & Honey, 2012; Nastase et al., 2020), the neuroscience of language must move past experimental stimuli in order to become more ecologically valid. Hasson et al. (2010), in review, find that naturalistic stimuli, including audiobooks, can, in cases, evoke more

reliable and functionally selective responses than experimental stimuli. As well, in an investigation of the neural correlates of code switching in bilinguals with MEG, Blanco-Elorrieta and Pylkkänen (2017) find that production and comprehension results vary dramatically depending on whether the stimuli are artificial or naturalistic. Perhaps most important, though, is that experimental stimuli eliminate *context* from the task of language processing. Language, as a communicative tool, inherently involves context. That is, naturalistic language comprehension involves combining incoming linguistic information with previous context to build integrated representations. To analyze language processing in the absence of context is to deny a critical facet of language itself (Hasson et al., 2018).

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 ORCs. This study tests: (1) using naturalistic stimuli, whether the brain bases of ORC processing are identical for English and Chinese, two languages which form this construction radically differently; and (2) the extent to which previous neuroimaging results for ORC processing extend when naturalistic stimuli are used. It is hypothesized that: (1) the correlates for processing ORC constructions will be the same across the two languages and implicate at least the left inferior frontal gyrus and left posterior temporal lobe; and (2) that the naturalistic stimuli will evoke activation in brain areas outside of the traditional language network. To test these hypotheses, we analyzed ORCs in naturalistic fMRI data collected while English and Chinese participants listened to translation-equivalent audiobooks of the children's story *Le Petit Prince* (*The Little Prince*), by de Saint-Exupéry.

Materials and methods

The fMRI data analyzed were *The Little Prince Datasets* (Hale et al., 2022; J. Li et al., 2022; J. Li et al., 2021), 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*. A general linear model (GLM) analysis was performed which included a binary regressor for ORC processing as well as control regressors of non-interest. The binary ORC regressor marks with a 1 only those words where ORC processing occurs. For this binary ORC metric, all other words are marked as 0. This can be visualized in Fig. 1, and more detail will be given below. In order to ensure the fidelity of any effects found for the ORC metric, control regressors were included in the first level of the GLM for speaker amplitude, speaker pitch, spoken word rate, word frequency, syntactic processing, word-by-word surprisal, and lexical semantics. In this way, any effects found for the ORC metric cannot be attributed to these alternate aspects of language processing. The first-level English and Chinese ORC brain maps were entered into a second-level GLM analysis with a two-sample t-test design matrix encoding. Effects of ORC processing in each language were analyzed as was the difference in ORC processing between the two languages. Further, to probe cross-linguistic commonality between the two languages, the voxel-wise intersection was calculated between the ORC processing effects in English and Chinese. That is, if a voxel was found to be activated by ORC processing for both English and Chinese participants, it is recorded in order to construct a new brain map indicating voxels which are commonly activated during ORC processing for both English and Chinese participants.

fMRI data

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.

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, 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.

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).

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*).

Stimuli and storybook annotations

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.

Each word in the storybooks was annotated for a number of metrics which were expected to be cognitively influential. The first step in annotating the storybook texts was to parse them for syntactic structure. The syntactic annotations, or trees, were used to calculate a word-by-word syntactic complexity metric and to calculate a binary word-by-word ORC annotation. The English text was parsed using the `mtgpy` parser (Coavoux, 2021) while the Chinese text was parsed using the `benepar` parser (Kitaev et al., 2019). Following Brennan et al. (2012), a 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 parse trees. This coregressor serves to control for effects of syntactic processing and can be visualized in Fig. 1. The use of 1-best parses means that the metric is blind to any sort of ambiguity resolution work.

Tree-geometric properties were used to find the ORC 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 3 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. 1A.

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

After the filler is found, a gap that it fills or from which it has moved must be 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).

Annotation of the binary ORC metric proceeded in correspondence to areas of expected increased cognitive demand. That is, words where increased cognitive demand associated with ORC processing was expected were annotated with a 1, while all other words were annotated with a 0. For English ORCs, each word between the filler and the gap site was annotated with a 1 (Wanner & Maratsos, 1978). For Chinese ORCs, each word between the gap site and the end of the filler was annotated with a 1. The number of identified ORC constructions for each language is given in Table 1. One may notice that in the English case, the area of expected increased processing demand corresponds to the ORC itself, but in the Chinese case, the area of expected increased demand corresponds to the filler rather than the ORC. While English listeners can be expected to understand that they are processing a relative clause after hearing a relative pronoun, as discussed in the Chinese relative clause processing literature (Bulut et al., 2018; Hsiao & Gibson, 2003; Jäger et al., 2015; C.-J. C. Lin & Bever, 2011; Vasishth et al., 2013), in Chinese, there is main clause/relative clause structural ambiguity during the actual relative clause which is not resolved until at least the relativizer *DE*. Following the relativizer, the filler element must be processed as the head noun which the ORC is modifying. Fig. 1A gives the parse tree for example 1. The underlined words, *that you love*, indicate the span between the filler and the gap site where an increase in cognitive demand is expected in English. Fig. 1B 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 in Chinese.

Additionally, pretrained autoregressive transformer language models—one for each language—were used to calculate word-by-word surprisal (Hale, 2001; Levy, 2008). 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

ORC constructions because, as just mentioned, Chinese relative clauses have main clause/relative clause structural ambiguity at least until the relativizer *DE*. This is in contrast to English relative clauses where comprehenders are informed of the relative clause by a relative pronoun. Recent work (Oh & Schuler, 2022) has shown that as pretrained language model size increases, the ability to use surprisal to predict reading time decreases. For this reason, we use GPT2-caliber models. 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).

Lastly, as a lexical semantic control, word vectors were taken from the pretrained English and Chinese `fastText` models (Bojanowski et al., 2017; Grave et al., 2018). Word vectors are numerical representations which encode the meaning of lexical items. The `fastText` word vectors were trained on `Common Crawl` and `Wikipedia` using Continuous Bag of Words with position-weights, in dimension 300, with character n-grams of length 5, and a window of size 5 and 10 negatives. Native from the model, the vector for each word is of dimensionality 300. That is, each word is associated with 300 real-valued numbers. In order utilize the word vectors in the GLM analysis, they are reduced down to 5 dimensions via principal component analysis with the included utility from the `fastText` toolkit.

Experimental design and statistical analysis

GLM analysis

The GLM analysis was 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. Low-level regressors of non-interest included spoken word-rate, speaker pitch, and root mean squared amplitude of the spoken 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 log lexical frequency (log unigram frequency from the Google Books Ngram Viewer), the bottom-up syntactic processing metric, large language model surprisal, and word vectors. The regressor of interest was the word-by-word object relative metric.

All predictors were convolved with the `SPM` canonical hemodynamic response function. Following convolution, the lexical frequency, bottom-up syntactic processing, surprisal, word vector, and ORC 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 analysis, a liberal cortical mask (https://surfer.nmr.mgh.harvard.edu/fswiki/CorticalParcellation_Yeo2011) was applied, calculated from the 1,000 participants in Buckner et al. (2011) and Yeo et al. (2011). At the first level of the GLM, linear models were fit to the voxel blood oxygen level dependent (BOLD) timecourses. For the English participants, the English-associated regressors were used. For the Chinese participants, the Chinese-associated regressors were used. The Wilkinson-Rogers formula is given in 4, 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, `bottom_up` marks the offset of every word with its syntactic complexity metric value, `GPT2_surprisal` marks the offset of every word with how surprised the language model is that it encounters the word given the preceding context, **`word_vectors`**$_5$ marks the offset of every word with 5 regressors corresponding to the values of the word's pretrained `fastText` vector, following a model reduction from word vector dimensionality 300 to dimensionality 5, and `obj_relative` marks the offset of every word with the binary ORC regressor value at that word.

(4) $BOLD \sim 1 + word_rate + rms + freq + f0 + bottomup + GPT2_surprisal + \boldsymbol{word_vector}_5 + obj_relative$

The nine scanning sessions/storybook sections were used to compute fixed effects for each participant. Data for two Chinese participants resulted in errors at the first level and were not further analyzed, leaving data from 33 Chinese participants.

The first-level object relative coefficient maps were entered into the second level of the GLM analysis with a two-sample t-test design matrix encoding:

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_{English} \\ \beta_{Chinese} \end{pmatrix}$$

An equivalent number of English and Chinese first-level maps were analyzed: 33 of each language. An 8 mm full width at half maximum Gaussian smoothing kernel was applied to counteract inter-participant anatomical variation. Effects of object relative clause processing were calculated for English and Chinese, as was the contrast between the two languages. Further, the intersection of voxels found to be significant in each of the two languages was calculated in order to identify common voxels associated with object relative processing.

The presented English, Chinese, and contrast second-level maps are z-valued. The main effect second-level maps are thresholded with an expected false discovery rate (FDR) $< .05$ and a cluster threshold of 150 voxels. The contrast effect second-level map is

thresholded with an expected false discovery rate FDR < .05 and a cluster threshold of 10 voxels. The intersection map is the voxel-level intersection of the FDR-thresholded English and Chinese maps with a cluster threshold of 10 voxels. The MNI2TAL tool from the Yale BioImage Suite (Lacadie, Fulbright, et al., 2008; Lacadie, Fulbright, et al., 2008, version 1.4) was referenced for brain region and Brodmann area labels.

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/LPPxORCxEN-CN.

Results

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.

GLM

For the English participants, ORC processing was associated with an increase in activation in left pMTG, extending into the fusiform gyrus, left AG, the left IFG, including the pars opercularis and pars triangularis, left premotor cortex and superior frontal lobe, bilateral posterior cingulate cortex, and left precuneus. Decreases in activation were seen in left primary auditory cortex, bilateral precuneus, right anterior frontal lobe, and right premotor cortex. These increases and decreases in activation can be seen in Fig. 2 and more detail can be found in Table 2.

For the Chinese participants, ORC processing was associated with an increase in activation in left pM/STG, bilateral mid and anterior STG, right temporal pole, left AG, bilateral IFG, including the left pars triangularis and bilateral pars opercularis, left premotor cortex, bilateral premotor and superior frontal cortex, bilateral posterior

cingulate cortex, and bilateral precuneus. Decreases in activation were seen in the right occipital lobe and right fusiform gyrus. These increases and decreases in activation can be seen in Fig. 3 and more detail can be found in Table 3.

In contrasting ORC activation between the two languages, greater activation is found for Chinese than English in bilateral primary auditory cortex, bilateral STG, and left supramarginal gyrus. The increases in activation can be seen in Fig. 4 and more detail can be found in Table 4.

Commonly activated for both Chinese and English participants in response to ORCs were left pMTG, left angular gyrus, left IFG, including the pars opercularis and pars triangularis, left premotor cortex, and left posterior cingulate cortex and precuneus.

The commonalities can be seen in Fig. 5 and more detail can be found in Table 5.

Discussion

The results of this study allow us to answer the questions asked at the beginning, which were: (1) using naturalistic stimuli, are the brain bases of ORC processing identical for English and Chinese, two languages which form this construction radically differently; and (2) to what extent do previous neuroimaging results for ORC processing extend when naturalistic stimuli are used. With respect to the first question, while voxel-level overlap was identified in the left pMTG, left AG, left IFG, including the pars opercularis and pars triangularis, left premotor cortex, and left precuneus and posterior cingulate cortex, for the Chinese participants, the bilateral mid/anterior STG, right IFG, and right premotor cortex were additionally activated. Indeed, in bilateral mid STG and bilteral primary auditory cortex, the contrast between Chinese and English participants was significant. With respect to the second question, ORC processing in a naturalistic context was found to increase activation in brain areas outside of the traditional language network like the angular gyrus, premotor cortex, and the precuneus and posterior cingulate cortex.

The brain areas implicated in ORC processing in this study subsume those which

have previously been found for ORC processing in English: left IFG (Caplan et al., 2008; Just et al., 1996; Stromswold et al., 1996), right IFG (Just et al., 1996), left pM/STG (Caplan et al., 2008; Just et al., 1996), and right pM/STG (Just et al., 1996), although the right hemisphere IFG and pM/STG were found for Chinese speakers only, here, and not English speakers. These regions correspond to the traditionally-considered language network, and have also been found for ORC processing in Chinese: left IFG (Xiong & Newman, 2021), right IFG (Xiong & Newman, 2021), left pSTG (K. Xu et al., 2020b), right pSTG(K. Xu et al., 2020b), left mid M/STG (Lee & Chan, 2023), right mid M/STG (Lee & Chan, 2023), left MTG (Xiong & Newman, 2021; K. Xu et al., 2020b), right MTG (K. Xu et al., 2020b), left MTG broadly (Xiong & Newman, 2021), and right MTG broadly (Xiong & Newman, 2021). Quite interestingly, however, is that in both languages, ORC processing was found to further be associated with an increase in activation in left AG, left premotor cortex, and left precuneus and posterior cingulate cortex. These results are not consistent with previous ORC work in English (although see Caplan et al., 2008, for a left precuneus result). They *are*, however, consistent with more recent neuroimaging results from Chinese, which has identified left AG (Xiong & Newman, 2021), right AG (Xiong & Newman, 2021), left premotor cortex (K. Xu et al., 2020b), left precuneus (Xiong & Newman, 2021), and left posterior cingulate cortex (K. Xu et al., 2020b) for ORC processing.

The activation in extended language network areas is interpreted as resulting from the extension to naturalistic stimuli in this study. In a naturalistic setting, ORC comprehension (and language understanding more generally) will involve integrating incoming information into existing context. This is in contrast to previous experiments with carefully controlled stimuli, in which there would be little to no coherence across the individual stimulus items. In a naturalistic storybook text, it is possible for ORC head nouns or subjects to refer to previously introduced people, places, or things which play a narrative role outside of that observation or for ORC verbs to update the state of the

world. The identified extended language network regions have previously been implicated in narrative comprehension: precuneus (Lerner et al., 2011; Maguire et al., 1999; J. 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; J. Xu et al., 2005), and angular gyrus (Lerner et al., 2011; J. 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). This is an important result as it takes up the call to naturalistic action by Hasson and colleagues (Hasson et al., 2018; Hasson & Honey, 2012; Nastase et al., 2020), investigates a specific phenomenon in a naturalistic context, and shows that the results trend as expected towards naturalistic language processing compared to previous, laboratory-style experiments.

The M/STG results are interpreted as locations which build syntactic structure during language processing (Caplan et al., 2008; Flick & Pylkkänen, 2020; Matar et al., 2021; Matchin & Hickok, 2020; Pallier et al., 2011). As ORCs have more complex syntax than, say, declarative statements, this is an intuitive interpretation. More controversial, however, is interpreting the IFG results. While some accounts (e.g., Friederici, 2016; Hagoort, 2016; Pallier et al., 2011) claim that the IFG builds syntactic structure, the interpretation taken here is that IFG is not performing syntactic structure building (e.g., Bornkessel-Schlesewsky & Schlesewsky, 2016a; Caplan et al., 2008; Grodzinsky, 2000; Matchin & Hickok, 2020; Zurif et al., 1993), but is instead, reanalyzing thematic roles assigned by the middle and superior temporal lobe (Caplan et al., 2008). Caplan et al. (2008) ascribe a role to the inferior frontal gyrus in which it carefully checks constructed syntactic representation to verify thematic role assignments. The degree of IFG recruitment in this role is likely dependent on the difficulty or *noncanonicity* of the

material being processed (Traxler et al., 2002). Object relative clauses are noncanonical structures in both English and Chinese and involve complexity in properly assigning thematic roles. While investigating difficulty in thematic role assignment in noncanonical processing in object vs. subject first constructions in German, Meyer et al. (2012) consider "ordering," a mechanisms which in their words, "orders these arguments to avoid a who-whom confusion" (p. 1988). They identify an argument order effect in the left inferior frontal gyrus. In patients with aphasia, Magnusdottir et al. (2013) find that difficulty processing reversible, noncanonical sentences (where a heuristic assignment of thematic roles would lead to an incorrect interpretation) is associated with damage to the pars opercularis and pars triangularis of the IFG. Additional support for thematic role evaluation in the IFG, following assignment in middle and superior temporal lobe (Caplan et al., 1985), comes from the connectivity analysis of Flick and Pylkkänen (2020), who find connectivity from the posterior temporal lobe to the IFG in a syntactic construction condition. One pathway along which the necessary information may flow from the posterior temporal lobe to the IFG is the dorsal language tract. Wilson et al. (2011) observe that, in a syntactic comprehension evaluation (Wilson et al., 2010), for patients with primary progressive aphasia, dorsal language tract damage and gray matter loss in the IFG are associated with difficulty comprehending noncanonical sentences.

One interpretation of the AG results is working memory during sentence understanding (Novais-Santos et al., 2007), generally, and working memory for arguments (Meyer et al., 2012), specifically. The role played by the AG during naturalistic language comprehension seems, however, to include argument processing in a broader sense, related to verb-argument structure (Malyutina & den Ouden, 2017), a critical component of ORC understanding. In the neurocognitive model of Matchin and Hickok (2020), the AG serves as a semantic hub, "involved in event knowledge" (p. 1485); knowledge of an event entails understanding the *who*, *what*, and *whom* arguments of a predicate and how they relate to one another. Indeed, Williams et al. (2017) find that inferior parietal cortex is responsive

to relationality, that is, responsive to the capacity for relationships among arguments and Magnusdottir et al. (2013) find that processing difficulty with reversible, noncanonical sentences is associated with damage to the angular gyrus. In a lexical decision task, Meltzer-Asscher et al. (2015) find that an increase in the number of thematic arguments that a verb takes is associated with an increase in temporoparietal activation.

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, 2016b; 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 high-level language comprehension (Honey et al., 2012; Malik-Moraleda et al., 2022), this domain is advanced, here, by investigating a specific phenomenon, ORCs, 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 (P. Li et al., 2004), it is in this sort of specificity that differences emerge, if they exist. Indeed, both commonality and difference are found between the patterns of activation in English and Chinese. The commonality is interpreted as a common, extended language network for the interpretation of ORCs in naturalistic language. In terms of differences, it was found that for Chinese participants, ORC processing evoked increases in activation in some areas not seen in the English results: bilateral mid/anterior STG, right IFG, and right premotor cortex. These additionally activated regions are interpreted as resulting from increased syntactic, thematic, and representational processing demands for ORCs in Chinese related to reanalysis. In English, while ORCs are associated with a relative increase in processing demand, upcoming ORCs are generally announced by a relative pronoun. In contrast, for Chinese ORCs, main clause/relative clause structural ambiguity is not resolved until at

least the relativizer *DE* (Bulut et al., 2018; Hsiao & Gibson, 2003; Jäger et al., 2015; C.-J. C. Lin & Bever, 2011; Vasishth et al., 2013). It is this additional reanalysis that recruits supplementary, syntactic reanalysis in bilateral STG, thematic role reanalysis in right IFG, and representational reanalysis in right premotor cortex, compared to English (see Lee & Chan, 2023, for a related line of reasoning regarding activation differences between Chinese RCs, which are gap-filler constructions, and Chinese topicalization constructions, which are filler-gap constructions). It was further found that ORC processing evoked significantly greater activation in bilateral STG and bilateral primary auditory cortex for Chinese participants than English participants. As just discussed, the bilateral STG result is interpreted as supplementary syntactic reanalysis. The bilateral primary auditory cortex result is speculatively interpreted as subserving lexical tone processing in the Chinese participants (cf. Gandour et al., 2003), a process which would not occur in English speakers.

In conclusion, it was found that using naturalistic stimuli to investigate the brain bases of object-extracted relative clauses resulted in the implication of brain areas not previously identified (at least in the English literature) for ORC processing. These areas likely subserve the integration of incoming linguistic information into previous linguistic context, a process which would not occur in carefully controlled laboratory stimuli. Comparing English and Chinese, it was found that ORC processing elicits both similarities and differences in brain activation. It is likely that the additional brain activations identified in Chinese are involved in suplemental (re)analysis resulting from differences in surface form.

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/https://doi.org/10.3389/fninf.2014.00014

Blanco-Elorrieta, E., & Pylkkänen, L. (2017). Bilingual language switching in the laboratory versus in the wild: The spatiotemporal dynamics of adaptive language control. *Journal of Neuroscience*, *37*(37), 9022–9036. https://doi.org/https://doi.org/10.1523/JNEUROSCI.0553-17.2017

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, *5*, 135–146. https://doi.org/https://doi.org/10.48550/arXiv.1607.04606

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016a). The argument dependency model. In *Neurobiology of language* (pp. 357–369). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-407794-2.00030-4

Bornkessel-Schlesewsky, I., & Schlesewsky, M. (2016b). The importance of linguistic typology for the neurobiology of language. *Linguistic Typology*, *20*(3), 615–621. https://doi.org/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*(2), 163–173. https://doi.org/https://doi.org/10.1016/j.bandl.2010.04.002

Brennan, J., Stabler, E. P., Van Wagenen, S. E., Luh, W.-M., & Hale, J. (2016). Abstract linguistic structure correlates with temporal activity during naturalistic comprehension. *Brain and language*, *157*, 81–94. https://doi.org/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*(5), 2322–2345. https://doi.org/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*(1), 38–48. https://doi.org/https://doi.org/10.1002/(SICI)1097-0193(1999)7:1<38::AID-HBM4>3.0.CO;2-Q

Bulut, T., Cheng, S.-K., Xu, K.-Y., Hung, D. L., & Wu, D. H. (2018). Is there a processing preference for object relative clauses in chinese? evidence from erps. *Frontiers in Psychology*, *9*, 995. https://doi.org/10.3389/fpsyg.2018.00995

Caplan, D., Baker, C., & Dehaut, F. (1985). Syntactic determinants of sentence comprehension in aphasia. *Cognition*, *21*(2), 117–175. https://doi.org/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*(4), 643–656. https://doi.org/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. https://doi.org/10.1093/oso/9780198784999.001.0001

Chen, B., Ning, A., Bi, H., & Dunlap, S. (2008). Chinese subject-relative clauses are more difficult to process than the object-relative clauses. *Acta Psychologica*, *129*(1), 61–65. https://doi.org/https://doi.org/10.1016/j.actpsy.2008.04.005

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

Collins, M. (2003). Head-driven statistical models for natural language parsing. *Computational linguistics*, *29*(4), 589–637.

Cox, R. W. (1996). Afni: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research*, *29*(3), 162–173. https://doi.org/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*(12), 6096–6106. https://doi.org/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/https://doi.org/10.1016/j.bandl.2022.105110

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*(5), 581–593. https://doi.org/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*(2), 109–128. https://doi.org/https://doi.org/10.1016/0010-0277(95)00692-R

Flick, G., & Pylkkänen, L. (2020). Isolating syntax in natural language: Meg evidence for an early contribution of left posterior temporal cortex. *Cortex*, *127*, 42–57. https://doi.org/https://doi.org/10.1016/j.cortex.2020.01.025

Friederici, A. D. (2016). The neuroanatomical pathway model of language: Syntactic and semantic networks. In *Neurobiology of language* (pp. 349–356). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-407794-2.00029-8

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*(3), 318–336. https://doi.org/https://doi.org/10.1016/S0093-934X(02)00505-9

Gibson, E., Desmet, T., Grodner, D., Watson, D., & Ko, K. (2005). Reading relative clauses in english. https://doi.org/https://doi.org/10.1515/cogl.2005.16.2.313

Grave, E., Bojanowski, P., Gupta, P., Joulin, A., & Mikolov, T. (2018). Learning word vectors for 157 languages. *arXiv preprint arXiv:1802.06893*. https://doi.org/https://doi.org/10.48550/arXiv.1802.06893

Grodzinsky, Y. (2000). The neurology of syntax: Language use without broca's area. *Behavioral and brain sciences*, *23*(1), 1–21. https://doi.org/https://doi.org/10.1017/S0140525X00002399

Hagoort, P. (2016). Muc (memory, unification, control): A model on the neurobiology of language beyond single word processing. In *Neurobiology of language* (pp. 339–347). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-407794-2.00028-6

Hale, J. (2001). A probabilistic earley parser as a psycholinguistic model. *Second meeting of the north american chapter of the association for computational linguistics*.

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

Hasson, U., Egidi, G., Marelli, M., & Willems, R. M. (2018). Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for

language comprehension. *Cognition*, *180*, 135–157. https://doi.org/https://doi.org/10.1016/j.cognition.2018.06.018

Hasson, U., & Honey, C. J. (2012). Future trends in neuroimaging: Neural processes as expressed within real-life contexts. *NeuroImage*, *62*(2), 1272–1278. https://doi.org/https://doi.org/10.1016/j.neuroimage.2012.02.004

Hasson, U., Malach, R., & Heeger, D. J. (2010). Reliability of cortical activity during natural stimulation. *Trends in cognitive sciences*, *14*(1), 40–48. https://doi.org/https://doi.org/10.1016/j.tics.2009.10.011

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/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*(44), 15277–15283. https://doi.org/https://doi.org/10.1523/JNEUROSCI.1800-12.2012

Hsiao, F., & Gibson, E. (2003). Processing relative clauses in chinese. *Cognition*, *90*(1), 3–27. https://doi.org/https://doi.org/10.1016/S0010-0277(03)00124-0

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/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*(5284), 114–116. https://doi.org/DOI:10.1126/science.274.5284.114

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*(3), 623–634. https://doi.org/https://doi.org/10.1515/lingty-2016-0033

King, J., & Just, M. A. (1991). Individual differences in syntactic processing: The role of working memory. *Journal of memory and language*, *30*(5), 580–602. https://doi.org/https://doi.org/10.1016/0749-596X(91)90027-H

King, J., & Kutas, M. (1995). Who did what and when? using word-and clause-level erps to monitor working memory usage in reading. *Journal of cognitive neuroscience*, *7*(3), 376–395. https://doi.org/https://doi.org/10.1162/jocn.1995.7.3.376

Kitaev, N., Cao, S., & Klein, D. (2019). Multilingual constituency parsing with self-attention and pre-training. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, 3499–3505. https://doi.org/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*(3), 1759–1770. https://doi.org/https://doi.org/10.1016/j.neuroimage.2011.12.028

Lacadie, C. M., Fulbright, R., Arora, J., Constable, R., & Papademetris, X. (2008). Brodmann areas defined in mni space using a new tracing tool in bioimage suite. *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. (2008). More accurate Talairach coordinates for neuroimaging using non-linear registration. *Neuroimage*, *42*(2), 717–725. https://doi.org/https://doi.org/10.1016/j.neuroimage.2008.04.240

Lee, T.-w., & Chan, S.-h. (2023). Better early than late for a filler: An fmri study on the filler-gap order in language. *Journal of Neurolinguistics*, *66*, 101126. https://doi.org/https://doi.org/10.1016/j.jneuroling.2023.101126

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*(8), 2906–2915. https://doi.org/https://doi.org/10.1523/JNEUROSCI.3684-10.2011

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*(3), 1126–1177. https://doi.org/https://doi.org/10.1016/j.cognition.2007.05.006

Levy, R., & Andrew, G. (2006). Tregex and tsurgeon: Tools for querying and manipulating tree data structures. *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*. 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., Yang, Y., Pallier, C., & Hale, J. (2022). Le Petit Prince multilingual naturalistic fMRI corpus. *Scientific Data*, *9*, 530. https://doi.org/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*. OpenNeuro. https://doi.org/doi:10.18112/openneuro.ds003643.v2.0.1

Li, P., Jin, Z., & Tan, L. H. (2004). Neural representations of nouns and verbs in chinese: An fmri study. *Neuroimage*, *21*(4), 1533–1541. https://doi.org/https://doi.org/10.1016/j.neuroimage.2003.10.044

Lin, C.-J. C., & Bever, T. G. (2006). Subject preference in the processing of relative clauses in chinese. *Proceedings of the 25th west coast conference on formal linguistics*, *25*, 254–260.

Lin, C.-J. C., & Bever, T. G. (2011). Garden path and the comprehension of head-final relative clauses. In H. Yamashita, Y. Hirose, & J. L. Packard (Eds.), *Processing and producing head-final structures* (pp. 277–297). Springer Netherlands. https://doi.org/10.1007/978-90-481-9213-7_13

Lin, Y., & Garnsey, S. M. (2011). Animacy and the resolution of temporary ambiguity in relative clause comprehension in mandarin. In H. Yamashita, Y. Hirose, &

J. L. Packard (Eds.), *Processing and producing head-final structures* (pp. 241–275). Springer Netherlands. https://doi.org/10.1007/978-90-481-9213-7_12

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*(1), 54–66. https://doi.org/https://doi.org/10.1016/j.neuroimage.2005.07.005

Magnusdottir, S., Fillmore, P., Den Ouden, D., Hjaltason, H., Rorden, C., Kjartansson, O., Bonilha, L., & Fridriksson, J. (2013). Damage to left anterior temporal cortex predicts impairment of complex syntactic processing: A lesion-symptom mapping study. *Human brain mapping*, *34*(10), 2715–2723. https://doi.org/https://doi.org/10.1002/hbm.22096

Maguire, E. A., Frith, C. D., & Morris, R. (1999). The functional neuroanatomy of comprehension and memory: The importance of prior knowledge. *Brain*, *122*(10), 1839–1850. https://doi.org/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/https://doi.org/10.1038/s41593-022-01114-5

Malyutina, S., & den Ouden, D.-B. (2017). Task-dependent neural and behavioral effects of verb argument structure features. *Brain and language*, *168*, 57–72. https://doi.org/https://doi.org/10.1016/j.bandl.2017.01.006

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

Matar, S., Dirani, J., Marantz, A., & Pylkkänen, L. (2021). Left posterior temporal cortex is sensitive to syntax within conceptually matched arabic expressions. *Scientific reports*, *11*(1), 1–14. https://doi.org/https://doi.org/10.1038/s41598-021-86474-x

Matchin, W., & Hickok, G. (2020). The cortical organization of syntax. *Cerebral Cortex*, *30*(3), 1481–1498. https://doi.org/https://doi.org/10.1093/cercor/bhz180

Meltzer-Asscher, A., Mack, J. E., Barbieri, E., & Thompson, C. K. (2015). How the brain processes different dimensions of argument structure complexity: Evidence from fmri. *Brain and language*, *142*, 65–75. https://doi.org/https://doi.org/10.1016/j.bandl.2014.12.005

Meyer, L., Obleser, J., Anwander, A., & Friederici, A. D. (2012). Linking ordering in broca's area to storage in left temporo-parietal regions: The case of sentence processing. *Neuroimage*, *62*(3), 1987–1998. https://doi.org/https://doi.org/10.1016/j.neuroimage.2012.05.052

Müller, H. M., King, J. W., & Kutas, M. (1997). Event-related potentials elicited by spoken relative clauses. *Cognitive Brain Research*, *5*(3), 193–203. https://doi.org/https://doi.org/10.1016/S0926-6410(96)00070-5

Nastase, S. A., Goldstein, A., & Hasson, U. (2020). Keep it real: Rethinking the primacy of experimental control in cognitive neuroscience. *NeuroImage*, *222*, 117254. https://doi.org/https://doi.org/10.1016/j.neuroimage.2020.117254

Novais-Santos, S., Gee, J., Shah, M., Troiani, V., Work, M., & Grossman, M. (2007). Resolving sentence ambiguity with planning and working memory resources: Evidence from fmri. *Neuroimage*, *37*(1), 361–378. https://doi.org/https://doi.org/10.1016/j.neuroimage.2007.03.077

Oh, B.-D., & Schuler, W. (2022). Why does surprisal from larger transformer-based language models provide a poorer fit to human reading times? *arXiv preprint arXiv:2212.12131*. https://doi.org/https://doi.org/10.48550/arXiv.2212.12131

Packard, J. L., Ye, Z., & Zhou, X. (2011). Filler-gap processing in mandarin relative clauses: Evidence from event-related potentials. *Processing and producing head-final structures*, 219–240. https://doi.org/https://doi.org/10.1007/978-90-481-9213-7_11

Pallier, C., Devauchelle, A.-D., & Dehaene, S. (2011). Cortical representation of the constituent structure of sentences. *Proceedings of the National Academy of Sciences*, *108*(6), 2522–2527. https://doi.org/https://doi.org/10.1073/pnas.1018711108

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*(1), 91–96. https://doi.org/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/https://doi.org/10.48550/arXiv.1201.0490

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, *1*(8), 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

Staub, A. (2010). Eye movements and processing difficulty in object relative clauses. *Cognition*, *116*(1), 71–86. https://doi.org/https://doi.org/10.1016/j.cognition.2010.04.002

Stromswold, K., Caplan, D., Alpert, N., & Rauch, S. (1996). Localization of syntactic comprehension by positron emission tomography. *Brain and language*, *52*(3), 452–473. https://doi.org/https://doi.org/10.1006/brln.1996.0024

Traxler, M. J., Morris, R. K., & Seely, R. E. (2002). Processing subject and object relative clauses: Evidence from eye movements. *Journal of memory and language*, *47*(1), 69–90. https://doi.org/https://doi.org/10.1006/jmla.2001.2836

Traxler, M. J., Williams, R. S., Blozis, S. A., & Morris, R. K. (2005). Working memory, animacy, and verb class in the processing of relative clauses. *Journal of Memory and Language*, *53*(2), 204–224. https://doi.org/https://doi.org/10.1016/j.jml.2005.02.010

Vasishth, S., Chen, Z., Li, Q., & Guo, G. (2013). Processing chinese relative clauses: Evidence for the subject-relative advantage. *PloS one*, *8*(10), e77006. https://doi.org/https://doi.org/10.1371/journal.pone.0077006

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.

Williams, A., Reddigari, S., & Pylkkänen, L. (2017). Early sensitivity of left perisylvian cortex to relationality in nouns and verbs. *Neuropsychologia*, *100*, 131–143. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2017.04.029

Wilson, S. M., Dronkers, N. F., Ogar, J. M., Jang, J., Growdon, M. E., Agosta, F., Henry, M. L., Miller, B. L., & Gorno-Tempini, M. L. (2010). Neural correlates of syntactic processing in the nonfluent variant of primary progressive aphasia. *Journal of Neuroscience*, *30*(50), 16845–16854. https://doi.org/https://doi.org/10.1523/JNEUROSCI.2547-10.2010

Wilson, S. M., Galantucci, S., Tartaglia, M. C., Rising, K., Patterson, D. K., Henry, M. L., Ogar, J. M., DeLeon, J., Miller, B. L., & Gorno-Tempini, M. L. (2011). Syntactic processing depends on dorsal language tracts. *Neuron*, *72*(2), 397–403. https://doi.org/https://doi.org/10.1016/j.neuron.2011.09.014

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/https://doi.org/10.48550/arXiv.1910.03771

Xiong, Y., Dekydtspotter, L., & Newman, S. (2019). When embeddedness matters: Electrophysiological evidence for the role of head noun position in chinese relative clause processing. *Journal of Neurolinguistics*, *51*, 236–257. https://doi.org/https://doi.org/10.1016/j.jneuroling.2019.03.005

Xiong, Y., & Newman, S. (2021). Both activation and deactivation of functional networks support increased sentence processing costs. *Neuroimage*, *225*, 117475. https://doi.org/https://doi.org/10.1016/j.neuroimage.2020.117475

Xu, J., Kemeny, S., Park, G., Frattali, C., & Braun, A. (2005). Language in context: Emergent features of word, sentence, and narrative comprehension. *Neuroimage*, *25*(3), 1002–1015. https://doi.org/https://doi.org/10.1016/j.neuroimage.2004.12.013

Xu, K., & Duann, J.-R. (2020). Brain connectivity in the left frontotemporal network dynamically modulated by processing difficulty: Evidence from chinese relative clauses. *Plos one*, *15*(4), e0230666. https://doi.org/https://doi.org/10.1371/journal.pone.0230666

Xu, K., Wu, D. H., & Duann, J.-R. (2020a). Dynamic brain connectivity attuned to the complexity of relative clause sentences revealed by a single-trial analysis. *NeuroImage*, *217*, 116920. https://doi.org/https://doi.org/10.1016/j.neuroimage.2020.116920

Xu, K., Wu, D. H., & Duann, J.-R. (2020b). Enhanced left inferior frontal to left superior temporal effective connectivity for complex sentence comprehension: Fmri evidence from chinese relative clause processing. *Brain and Language*, *200*, 104712. https://doi.org/https://doi.org/10.1016/j.bandl.2019.104712

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/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*(2), 113–148. https://doi.org/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*, 241. https://doi.org/https://doi.org/10.48550/arXiv.1909.05658

Zhou, P., Yao, Y., & Christianson, K. (2018). When structure competes with semantics: Reading chinese relative clauses. *Collabra: Psychology*, *4*(1). https://doi.org/https://doi.org/10.1525/collabra.131

Zurif, E., Swinney, D., Prather, P., Solomon, J., & Bushell, C. (1993). An on-line analysis of syntactic processing in broca s and wernicke s aphasia. *Brain and language*, *45*(3), 448–464. https://doi.org/https://doi.org/10.1006/brln.1993.1054

	Object-extracted relative clauses
English	32
Chinese	33

Table 1

The number of identified English and Chinese object-extracted relative clauses in The Little Prince.

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
L Posterior Cingulate Cortex (BA 31)	36832	-8.0	-62.0	24.0	5.86
L Angular Gyrus (BA 39)		-38.0	-70.0	36.0	5.74
		-36.0	-58.0	38.0	5.25
L Posterior Cingulate Cortex (BA 23)		-6.0	-36.0	36.0	5.23
L Premotor Cortex (BA 6)	6632	-30.0	14.0	56.0	4.85
		-24.0	14.0	68.0	4.56
L Fusigorm Gyrus (BA 37)	6992	-52.0	-48.0	-20.0	4.74
		-58.0	-50.0	-6.0	4.01
L Middle Temporal Gyrus (BA 21)		-58.0	-34.0	-12.0	3.07
L Fusigorm Gyrus (BA 37)	2256	-30.0	-34.0	-22.0	4.66
L Pars Triangularis (BA 45)	2688	-48.0	40.0	6.0	4.28
L Superior Frontal (BA 8)	1352	-34.0	14.0	32.0	3.49
L Pars Opercularis (BA 44)		-50.0	14.0	20.0	3.07
R Precuneus (BA 7)	2792	16.0	-56.0	72.0	-4.61
		18.0	-40.0	66.0	-3.42
L Primary Auditory Cortex (BA 41)	2072	-34.0	-26.0	12.0	-4.09
		-42.0	-18.0	0.0	-3.92
L Precuneus (BA 7)	1496	-16.0	-44.0	66.0	-3.84
R Anterior Frontal (BA 10)	1832	30.0	52.0	-4.0	-3.80
R Premotor Cortex (BA 6)	2184	34.0	-14.0	72.0	-3.52
		44.0	-10.0	64.0	-3.44
		22.0	-4.0	74.0	-3.41
		28.0	-4.0	56.0	-2.97

Table 2

English object-extracted relative clause GLM results, thresholded with an expected false discovery rate < .05 and a cluster threshold of 150 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
L Precuneus (BA 7)	10296	-8.0	-70.0	34.0	5.54
L Posterior Cingulate Cortex (BA 23)		-6.0	-34.0	24.0	5.36
R Posterior Cingulate Cortex (BA 23)		2.0	-30.0	26.0	4.78
R Posterior Cingulate Cortex (BA 31)		18.0	-64.0	32.0	3.94
L Premotor Cortex (BA 6)	14456	-52.0	2.0	48.0	5.00
L Pars Triangularis (BA 45)		-56.0	26.0	0.0	4.50
L Pars Opercularis (BA 44)		-46.0	20.0	20.0	4.13
L Pars Triangularis (BA 45)		-44.0	20.0	12.0	4.01
R Superior Frontal (BA 8)	5496	52.0	10.0	48.0	4.93
		34.0	4.0	34.0	4.37
		42.0	8.0	38.0	4.34
L Superior Temporal Gyrus (BA 22)	13688	-62.0	-18.0	6.0	4.69
L Middle Temporal Gyrus (BA 21)		-60.0	-44.0	6.0	4.40
L Superior Temporal Gyrus (BA 22)		-50.0	-2.0	-8.0	4.23
		-50.0	-40.0	22.0	3.73
R Superior Temporal Gyrus (BA 22)	13248	54.0	-2.0	-6.0	4.52
R Pars Opercularis (BA 44)		44.0	16.0	18.0	4.09
R Temporal Pole (BA 38)		52.0	20.0	-20.0	4.08
R Superior Temporal Gyrus (BA 22)		56.0	-24.0	4.0	4.01
L Angular Gyrus (BA 39)	4008	-44.0	-72.0	48.0	4.41
		-36.0	-60.0	42.0	4.40
L Premotor Cortex (BA 6)	1680	-2.0	4.0	60.0	4.25
L Superior Frontal (BA 8)		-10.0	16.0	48.0	3.29
R Occipital (BA 19)	1440	56.0	-72.0	10.0	-4.08
R Fusiform Gyrus (BA 37)		56.0	-72.0	-2.0	-3.68
R Occipital (BA 19)		46.0	-84.0	14.0	-3.60

Table 3

Chinese object-extracted relative clause GLM results, thresholded with an expected false discovery rate < .05 and a cluster threshold of 150 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z	Peak stat (z)
R Superior Temporal Gyrus (BA 22)	3704	56.0	-2.0	-6.0	4.72
R Primary Auditory Cortex (BA 41)		48.0	-24.0	6.0	4.16
L Primary Auditory Cortex (BA 41)	1160	-56.0	-16.0	6.0	4.29
L Superior Temporal Gyrus (BA 22)		-68.0	-16.0	6.0	4.28
L Supramarginal Gyrus (BA 40)		-46.0	-26.0	10.0	3.78
R Primary Auditory Cortex (BA 41)	120	38.0	-32.0	12.0	4.20

Table 4

Chinese > English object-extracted relative clause GLM results, thresholded with an expected false discovery rate < .05 and a cluster threshold of 10 voxels. Region, cluster size, MNI coordinates, and peak statistic (z).

Region	Cluster size (mm^3)	MNI coordinates x	y	z
L Middle Temporal Gyrus (BA 21)	480	-60.0	-46.0	2.0
L Pars Opercularis (BA 44)	952	-40.0	16.0	28.0
L Pars Triangularis (BA 45)	416	-46.0	30.0	12.0
L Angular Gyrus (BA 39)	3736	-40.0	-64.0	44.0
L Premotor Cortex (BA 6)	408	-38.0	4.0	54.0
L Posterior Cingulate Cortex (BA 31)	5640	-8.0	-56.0	30.0

Table 5

Object-extracted relative clause voxel-level intersection of the false discovery rate-thresholded Chinese and English results, thresholded with a cluster threshold of 10 voxels. Region, cluster size, and MNI coordinates.

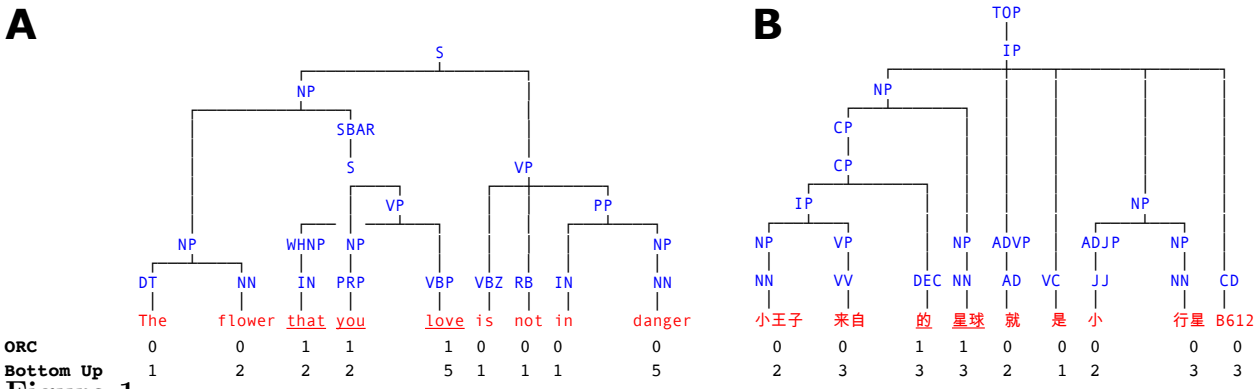

Figure 1

Example phrase structure trees with annotations for the corresponding word-by-word object-extracted relative clause (ORC) and bottom up syntactic processing metric values. ***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.*

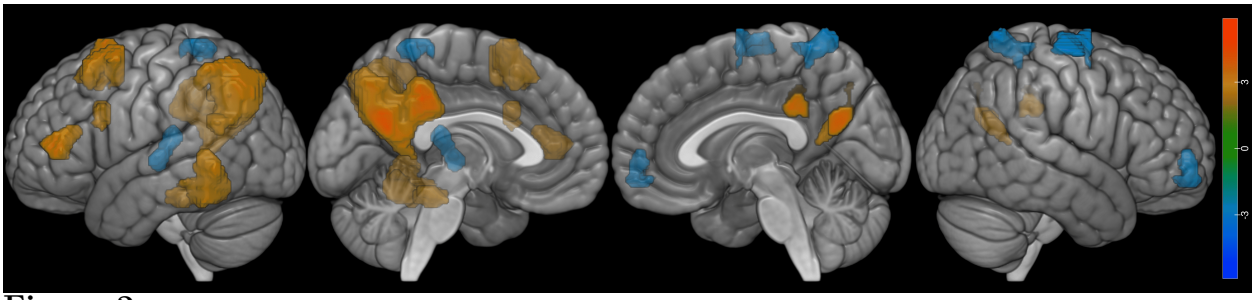

Figure 2

English object-extracted relative clause GLM results (z-valued), thresholded with an expected false discovery rate < .05 and a cluster threshold of 150 voxels.

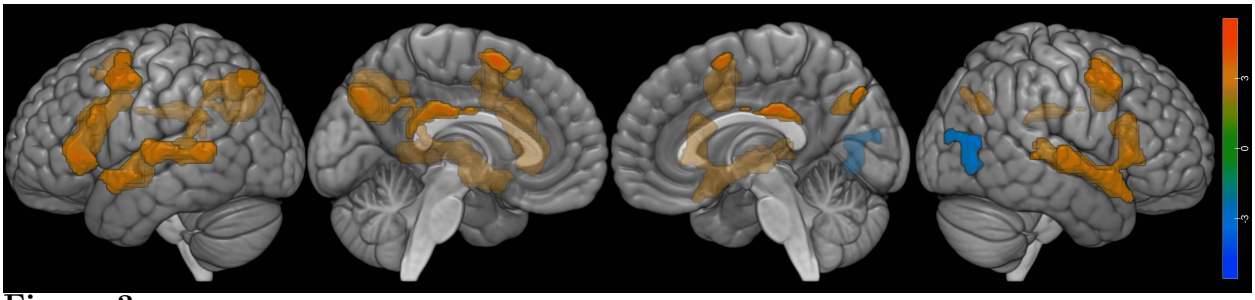

Figure 3

Chinese object-extracted relative clause GLM results (z-valued), thresholded with an expected false discovery rate < .05 and a cluster threshold of 150 voxels.

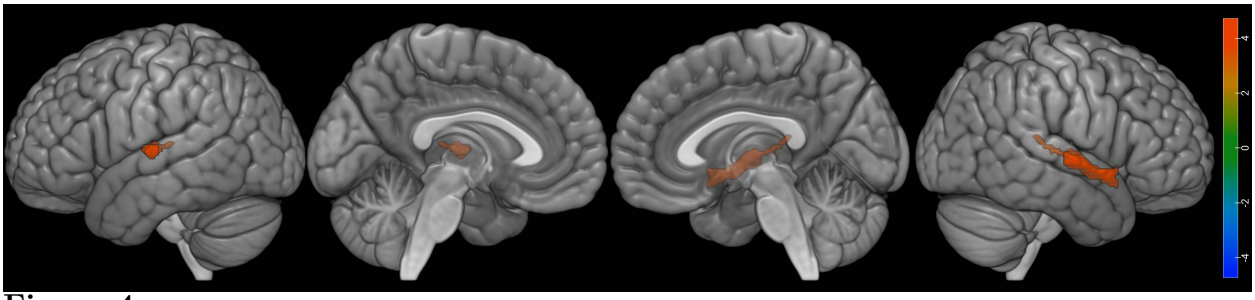

Figure 4

Chinese > English object-extracted relative clause GLM results (z-valued), thresholded with an expected false discovery rate < .05 and a cluster threshold of 10 voxels.

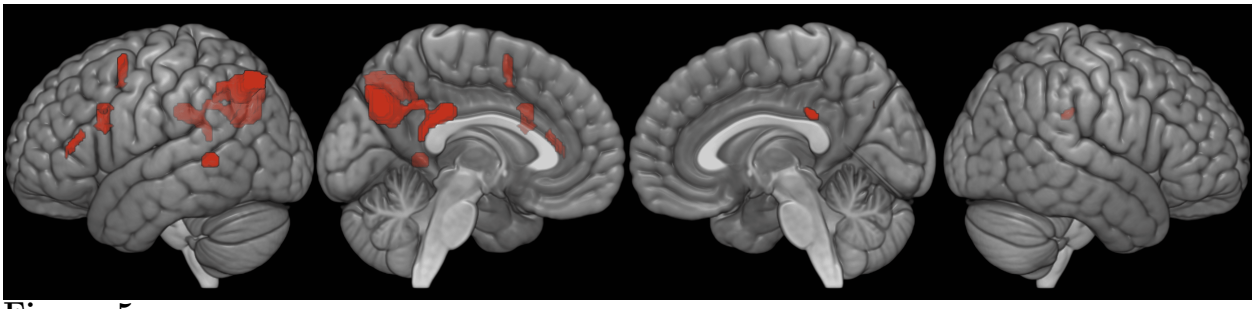

Figure 5

Object-extracted relative clause voxel-level intersection of the false discovery rate-thresholded Chinese and English maps, thresholded with a cluster threshold of 10 voxels.