Modeling Brain Activity Associated with Pronoun Resolution in English and Chinese

Jixing Li
Department of Linguistics
Cornell University
jl2939@cornell.edu

Murielle Fabre
Department of Linguistics
Cornell University
mf684@cornell.edu

Wen-Ming Luh
Cornell MRI Facility
Cornell University
wl358@cornell.edu

John Hale
Department of Linguistics
Cornell University
jthale@cornell.edu

Abstract

Typological differences between English and Chinese suggest stronger reliance on salience of the antecedent during pronoun resolution in Chinese. We examined this hypothesis by correlating a difficulty measure of pronoun resolution derived by the activation-based ACT-R model with the brain activity of English and Chinese participants listening to a same audiobook during fMRI recording. The ACT-R model predicts higher overall difficulty for English speakers, which is supported at the brain level in left Broca's area. More generally, it confirms that computational modeling approach is able to dissociate different dimensions that are involved in the complex process of pronoun resolution in the brain.

1 Introduction

Pronoun resolution has been suggested to be influenced by morpho-syntactic constraints such as gender/number/person agreement, and discourse factors such as salience of the antecedents (e.g., Grosz et al., 1995, for empircical evidence using eye tracking see Arnold, 2000). Yet, unlike English, Chinese pronouns in their spoken form lack morphosyntactic marking and can even be omitted. This leads to a general hypothesis that Chinese speakers may rely more on salience of the antecedent during pronoun resolution compared to English speakers, who have additional morphosyntactic cues to help resolve the referents.

The Adaptive Control of Thought-Rational (ACT-R) model of pronoun resolution (van Rij et al., 2013) offers a way to test this hypothesis. Built within the cognitive architecture of ACT-R (Anderson, 2007), this model represents pronoun resolution as the selection of a most-activated mention from a ranked list of candidates stored in declarative memory. As such, this model calculates the activation level for each mention and selects the most activated one as the antecedent. Given our hypothesis that Chinese has stronger reliance on salience of the antecedent during pronoun resolution in absence of morphosyntactic cues for agreement, we predict that antecedents in Chinese have higher overall activation level than antecedents in English.

We compared the mean ACT-R activation level for each pronoun's antecedent in a same audiobook in English and Chinese translation, and confirmed higher activation level for antecedents in Chinese. We then took the negative of the antecedents' activation levels to represent "effort" during pronoun resolution, and correlated this difficulty measure with brain activity while both English and Chinese participants listened to the pronouns present in an audiobook during fMRI recoding. The results revealed different left-lateralized ACT-R difficulty effects for English and Chinese. Chinese specifically activates the Angular Gyrus where non-grammatical gender information processing was observed (Hammer et al., 2007), while English ACT-R effect encompasses left Broca's area.

This study builds upon the salience-based account of pronoun resolution, which has long been recognized both in computational linguistics (Kantor, 1977; Alshawi, 1987; Brennan et al., 1987; Lappin and Leass, 1994) and in psychology (Grosz et al., 1995; Gordon et al., 1993; Greene et al., 1992; Arnold, 2010; McElree, 2001). Our key contribution is in demonstrating different brain activation maps associated with measures of the ACT-R model in English and Chinese, suggesting that different factors are operative in the system of pronoun resolution in different languages.

In the following sections, we briefly review the typological differences linked to pronominal reference in English and Chinese, and describe the

algorithm of the ACT-R model. We, then compare the model performance on our English and Chinese audiobook texts, and present the methods, results and discussion of the English and Chinese fMRI experiments.

2 English and Chinese Typological Differences in Pronominal Reference

In English, pronouns cannot be omitted in the subject or object position of a tensed clause, even though the referent of the omitted pronoun remains clear.On the contrary, Chinese can have null pronouns in subject or object position in tensed clauses under appropriate contextual licencing (see (1) and (2), data from Huang (1989)). This phenomenon is called *pro*-drop in generative syntax and has been considered a typological parameter that distinguishes *topic-prominent* languages like Chinese and *subject-prominent* languages like English (Li and Thompson, 1976; Xu, 2006).

(1) Speaker A: Did John see Bill yesterday?
     Speaker B: a. Yes, he saw him.
              b. *Yes, *e* saw him.
              c. *Yes, he saw *e*.
              d. *Yes, *e* saw *e*.
              e. *Yes, I guess *e* saw *e*.
              f. *Yes, John said *e* saw *e*.

(2) Speaker A: Zhangsan kanjian Lisi le ma?
                     Zhangsan see Lisi LE Q?
                     "Did Zhangsan see Lisi?"
     Speaker B: a. Ta kanjian ta le.
                     He see he LE.
                     "He saw him."
              b. *e* kanjian ta le.
                     "[He] saw him."
              c. Ta kanjian *e* le.
                     "He saw [him]."
              d. *e* kanjian *e* le.
                     "[He] saw [him]."
              e. Wo cai *e* kanjian *e* le.
                     I guess see LE.
                     "I guess [he] saw [him]."
              f. Zhangsan shuo *e* kanjian *e* le.
                     Zhangsan say see LE.
                     "Zhangsan said that [he] saw [him]."

Moreover, the spoken form of Chinese pronouns do not mark gender information. Therefore, Chinese pronouns in general provide fewer morpho-syntactic cues than English pronouns. This may additionally lead Chinese speakers to rely more on discourse information than English speakers when resolving the referential linking of pronouns.

3 ACT-R Model for Pronoun Resolution

3.1 Salience-Based Account of Pronoun Resolution

According to the salience-based account on pronoun resolution, the antecedent of the pronoun is a highly prominent entity in the discourse context. The Centering Theory (Grosz et al., 1995), for example, refers to entities that link an utterance to another utterance as "centers". Each utterance (U) has a set of forward-looking centers ($C_f(U)$; i.e., potential antecedents) and a single backward-looking center ($C_b(U)$; i.e., the anaphoric expression). The backward-looking center of utterance U_{n+1} connects with one of the forward-looking center of utterance U_n. The elements of C_f are ordered to reflect their relative prominence in U_n, and the most highly ranked C_f is the C_b of U_{n+1}. The ranking of C_f is mainly determined by its grammatical role in U_n: SUBJECT > OBJECT > OTHERS. Thus the subject of the previous sentence is more likely to be the antecedent of the pronoun in the next sentence.

The Centering Theory has been tested by Gordon et al. (1993) in a number of self-paced reading experiments. They introduced a prominent entity (C_b) and a less prominent entity in a short passage and found that reading time significantly increased when the prominent entity is not pronominalized but repeated. They also showed the repeated-name penalty for C_b only in the grammatical subject position, confirmed the basic notion in the Centering Theory that there is only one C_b in an utterance, and that grammatical subject ranks the highest among the C_fs.

Arnold (2010) prosed a similar salience-based account for pronominalization. She referred to salience of the entity as "accessibility" and suggested that the more accessible entities tend to be pronominalized (see evidence from oral corpus analysis by Arnold et al. 2009). The accessibility of an entity is influenced by whether the entity has been mentioned before, how recent the entity occurs, and whether the entity is syntactically or thematically prominent. McElree (2001), from a memory decay point of view, also suggested that recency, together with frequency, i.e, how many times the entity occurs, contribute to the salience of the entity in the discourse context.

Based on the notion of salience, Greene et al. (1992) proposed a framework for pronouns resolution in which a pronoun is matched automatically against all the entities in the discourse model. The degree of match is determined by the accessibility of the entity, as well as the gender/number features. The entity is automatically identified as the antecedent of the pronoun if its degree of match to the pronoun is better than the match for other entities. If no match or more than one match is identified, the selection process is postponed for more discourse information, or strategic problem-solving can be attempted.

3.2 Salience-Based Computational Models for Pronoun Resolution

One early influential pronoun resolution model based on the Centering Theory is proposed by Brennan et al. (1987; henceforth BFP). The BFP algorithm computes the preferred antecedents from relations that hold between the forward and backward looking centers in adjacent sentences. The algorithm first generates all possible $C_b - C_f$ pairs for the pronoun in Utterance U_n. It then filters all pairs by constraints such as the Binding Theory (Chomsky, 1982) and Centering rules, For example, if C_b must be pronominalized if any C_f is pronominalized; C_b is the highest ranked elements in the list of C_fs, etc. Finally, the algorithm ranks the remaining pairs by transition orderings, where maintaining the same C_b (Continue) is preferred over maintaining the same C_b in U_{n+1} but not in U_{n+2} (Retain), which is preferred to changing C_b in U_{n+1} (Shift). The selected $C_b - C_f$ pair is the most preferred relation according to the transition order.

Another influential salience-based model for pronoun resolution is the RAP algorithm proposed by Lappin and Leass (1994). Unlike the BFP algorithm that compares a discrete number of centers, the RAP algorithm assumes a graded activation level for each entities. It also follows a generate-filter-rank procedure and takes as input the output of a full parser and filters entities according to binding constraints and gender/number agreement. It then assigns a salience weight to each entity depending on its recency, syntactic position, grammatical role, etc. The weights are halved for each sentence boundary in between the entity and the pronoun, and weights for all occurrence of the same entity are summed. The entity that receives the highest salience weight is the antecedent of the pronoun.

Both the BFP and the RAP algorithms incorporate claims from the Centering Theory, as well as binding constraints and gender/number agreements. The RAP algorithm is similar to the ACT-R model in that they both assume salience on a graded scale and compute the activation level for each entity. However, the ACT-R model for pronoun resolution is adapted from the cognitive architecture of ACT-R (Anderson, 2007), which is both a computational model and a theory of how different components of the mind worked to produce coherent cognition. ACT-R consists a set of modules, and the ACT-R model for pronoun resolution is adapted from its declarative module for retrieving information from memory. The elements in the ACT-R model for pronoun resolution include only the recency, frequency and grammatical role of the entity, with no syntactic and gender/number information. The following sections describes the formula of the ACT-R model in detail.

3.3 The ACT-R Model for Pronoun Resolution

Built within the cognitive architecture of ACT-R (Anderson, 2007), van Rij et al. (2013) proposed an ACT-R model of pronoun resolution which integrates the major factors that influences the salience of the antecedents – frequency, recency, and the grammatical role of the antecedent. The formula of for the activation level for the antecedent i of a pronoun is

$$A_i = ln(\sum_{k=1}^n t_k^{-0.5}) + \sum_j^m W_j \times 2$$

where the base-level activation $ln(\sum_{k=1}^n t_k^{-0.5})$ represents frequency and recency of each mention k of the antecedent i, and the associative activation $\sum_j^m W_j \times 2$ represents the influence of grammatical role of each mention k. If mention j is a subject, it has a weight (W) of 1; which is divided by the total number of mentions of this antecedent n ($W_j = W/n$), as the total value of associative activation cannot be infinite. W_j is then multiplied by 2, the default value of associative strength in ACT-R (see Anderson, 2007, p.110 for the ACT-R equations).

The effects of frequency and recency is folded into the calculation of the base activation for antecedent i, such that the more mentions it has, and

the more recent the mentions occur, the higher the base activation. Conversely, if antecedent i has been mentioned only once, or if its last mention was a long time ago, its activation level will be low, and it will rank lower on the activation list for all the candidates antecedents. The subjecthood of the mentions of antecedent i gains a spreading activation in addition to the base activation. Overall, the amount of activation value of an antecedent is computed in an attempt to "abstract the impact of neural Hebbian-like learning and spread of activation among neurons" (Anderson, 2007, p.35).

To give a concrete example of how the activation level for each antecedent is calculated, in the English sentences in Figure 1, the immediate antecedent of the pronoun "they$_{16}$" is "they$_{15}$". The previous mentions of "they$_{15}$" are "their$_{13}$", "boa constrictors$_{12}$" and "boa constrictors$_8$". The time elapsed from these three previous mentions to "they$_{15}$" in the audio are 11.13 s, 3.68 s and 3.02 s respectively. Since "boa constrictor$_{12}$" is a subject of a subordinate clause, it gets an associative weighting W of 1. Therefore, the activation level for the antecedent of "they$_{15}$" is calculated as:

$$A_{they_{15}} = ln(11.13^{-0.5} + 3.68^{-0.5} + 3.02^{-0.5}) + 0/3 \times 2 + 1/3 \times 2 + 0/3 \times 2$$
$$\approx ln(0.3 + 0.52 + 0.58) + 0.67$$
$$\approx 1.01$$

Similarly, for the Chinese sentence in Figure 1, the antecedent of the last pronoun "它们$_{13}$ (tamen)" is "它们$_{11}$", and the the previous mentions of "它们$_{13}$" are "蟒蛇$_7$(mangshe)" and "蟒蛇$_{10}$". The time elapsed from them to "它们$_{13}$" in the audio are 7.94s and 0.44 s, respectively. Therefore, its activation level is calculated as:

$$A_{它们_{11}} = ln(7.94^{-0.5} + 0.44^{-0.5}) + 1/2 \times 2 + 1/2 \times 2$$
$$\approx ln(1.51 + 0.35) + 2$$
$$\approx 2.62$$

4 Current Study

Given the typological difference of pronouns in English and Chinese, it is hypothesized that Chinese speakers rely more on salience of the antecedent to resolve pronouns as they do not have additional morpho-syntactic cues. The current study first tests this hypothesis by comparing the activation level of the antecedent for each pronoun in a same audiobook *The Little Prince* in English and Chinese translation. Our prediction is that antecedents in Chinese have higher overall activation levels than antecedents in English.

To explain how the theoretical predictions and the model performance are specifically brought to bear on brain activity, we further correlated the negative of the activation levels of the antecedents with fMRI time-courses when both English and Chinese participants listened to the pronouns in the audiobook in the scanner. We took the negative of the activation level to indicate difficulty during pronoun resolution: that lower the activation level of the antecedent, the more difficulty to successfully retrieve the antecedents, hence higher hemodynamic response.

Based on the elements in the ACT-R formula, we expected the difficulty of pronoun resolution modeled by activation spread to tease apart the brain areas whose activity is influenced by the frequency, recency and the grammatical role of the antecedent, and to highlight regions where the effort of pronoun resolution is stronger in English and Chinese respectively.

Previous neuroimgaing results on pronoun resolution have implicated the bilateral Inferior Frontal Gyrus (IFG), the left Medial Frontal Gyrus (MFG) and the bilateral Supramarginal/Angular Gyrus in gender mismatch between pronoun and antecedent (Hammer et al., 2007). We therefore expect associated activity in these regions for the ACT-R model of pronoun resolution. We also expect to see activity in the bilateral Superior Temporal Gyrus (STG) as they have been associated with long distance pronoun-antecedent linking (Matchin et al., 2014). The Precuneus cortex activity may also be activated as it has been suggested to track different sorts of story characters (Wehbe et al., 2014).

5 Model Performance on Text Data

Based on van Rij et al.'s (2013) formula, we calculated the activation level for each previously mentioned entities for each pronoun in the English and Chinese audiobook of Antoine de Saint-Exupéry's *The Little Prince*. Within the English audiobook text, 1755 pronouns (excluding possessives, reflexives and dummy pronoun "it") and 3127 non-pronominal entities (4882 mentions in total) are identified. Pronouns with sentential antecedents are removed. For example, in the conversation "That is funny where you live a day only last a minute." "It is not funny at all." "it" in the second sentence is removed from our pronoun set as it refers to the whole sentence "where you live a day

1 Once when I_1 was six years old, I_2 saw a magnificent picture_3 in a book_4 about the primeval forest_5 called real life stories_6.

2 It_7 showed a boa constrictor_8 swallowing a wild animal_9.

3 Here is a copy of the drawing_10.

4 It said in the book_11 boa constrictors_12 swallow their_13 prey_14 whole, without chewing, then they_15 are not able to move and they_16 sleep for the six months it takes for digestion.

1 当 我_1 还只有 六 岁 的 时候, 在 一 本 描写 原始 森林_3 的 名 叫 真实 的 故事_4 的 书_5 中 看到 了 一幅 精彩 的 插画_6.

2 画 的 是 一 条 蟒蛇_7 正在 吞食 一 只 大 野兽_7.

3 页头 上 就 是 那 幅 画_8 的 摹本.

4 这 本 书_9 中 写道, 这些 蟒蛇_10 把 它们_11 的 猎获物_12 不 加 咀嚼 地 囫囵 吞 下, 尔后 就 不 能 再 动弹 了.

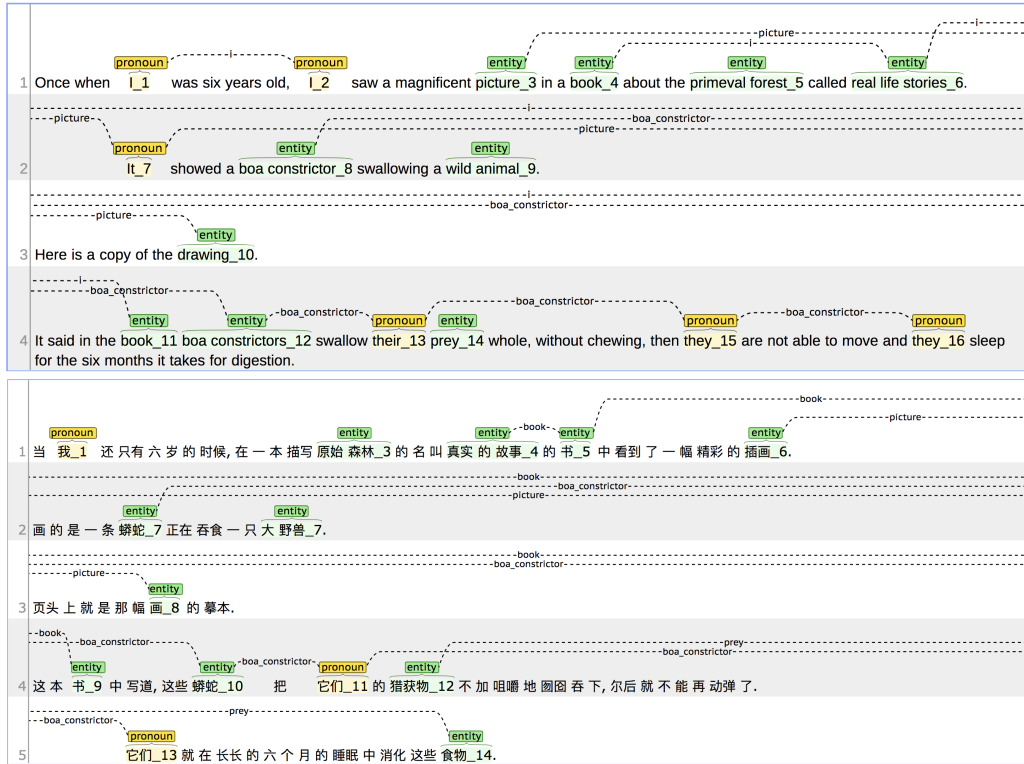5 它们_13 就 在 长长 的 六 个 月 的 睡眠 中 消化 这些 食物_14.

Figure 1: Sample annotations of pronouns and non-pronoun mentions in English and Chinese, visualized using the annotation tool brat (Stenetorp et al., 2012).

only last a minute". The resulting English data-set contains 645 1st person pronouns, 302 2nd person pronouns and 675 3rd person pronouns (see Table 1).

The Chinese audiobook text contains 1785 pronouns (excluding possessives and reflexives) and 2947 non-pronominal mentions (4732 mentions in total). We further pruned the pronoun set to exclude possessives, reflexives, and pronouns with sentential antecedents. Comparably to the English pronoun set, the resulting Chinese pronoun set contains 639 1st person pronouns, 298 2nd person pronouns and 529 3rd person pronouns (see Table 1). All pronouns and non-pronominal mentions in the English and Chinese texts were annotated using the annotation tool brat (Stenetorp et al., 2012; see Figure 1).

To evaluate the performance of the ACT-R model of pronoun resolution for the English and Chinese text, we ranked all the candidate antecedents for each pronoun according to their activation levels, such that the most activated antecedent has a rank of 1. We allow a certain level of ambiguity for cases where more than one antecedents can be linked to the pronoun, and decide that if the rank of the correct antecedent is

	English		Chinese		
1st	i	me	我(wo)		
	505	121	621		
	we	us	我们(women)		
	16	3	18		
2nd	you		你(ni)		
	302		261		
			你们(nimen)		
			37		
3rd	she	her	她(ta)		
	41	14	62		
	he	him	他(ta)		
	268	64	303		
	it		它(ta)		
	136		73		
	they	them	她们(tamen)	他们(tamen)	它们(tamen)
	94	58	2	74	15

Table 1: Attestations of each pronoun type in the English and Chinese texts. Note that Chinese 3rd person pronouns are homophones.

less or equal to 3 (top 3 on the ranking list), then the model is correct in predicting the antecedent for the pronoun.

Table 2 shows the accuracy of the model for each pronoun type in English and Chinese. As predicted, the model has higher overall accuracy for pronoun resolution in Chinese than in English. A further division of 1st, 2nd and 3rd person pro-

nouns reveals a 78% accuracy for 1st person pronouns in both English and Chinese, yet 2nd and 3rd person pronouns in Chinese has ∼ 8% higher accuracy than 2nd and 3rd person pronouns in English.

	Pronoun Type	Accuracy (%)
English	1st person	0.78
	2nd person	0.62
	3rd person	**0.54**
	all pronoun	0.66
Chinese	1st person	0.78
	2nd person	0.71
	3rd person	**0.62**
	all pronoun	0.71

Table 2: *Accuracy of ACT-R model on pronoun resolution in English and Chinese.*

Direct comparison of the activation levels of the antecedents in the two groups by a two-sample t-test confirmed that the activation level for Chinese antecedents (M=3.07, SD=1.52) is significantly higher than that for English antecedents (M=2.86, SD=1.64; $t(3097) = 3.4$, $p < 0.001$). Taken together, the model performance suggests that salience-based pronoun resolution predicts antecedents better in Chinese than in English, confirming our hypothesis about the strong reliance of Chinese on discourse salience compared to English speakers.

6 Correlating ACT-R Measures with Brain Activity

6.1 Participants

English participants were 49 healthy, right-handed, young adults (30 female, mean age = 21.3, range = 18-37). They self-identified as native English speakers, and had no history of psychiatric, neurological or other medical illness that could compromise cognitive functions. All participants were paid for, and gave written informed consent prior to participation, in accordance with the guidelines of the Human Research Participant Protection Program at Cornell University.

Chinese participants were 35 healthy, right-handed, young adults (15 female, mean age=19.3, range = 18-25). 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 for, and gave written informed consent prior to participation, in accordance with

the guidelines of the Ethics Committee at Jiangsu Normal University.

6.2 Stimuli

The English audio stimulus was an audiobook version of Antoine de Saint-Exupéry's *The Little Prince*, translated by David Wilkinson and read by Nadine Eckert-Boulet. This text contains 3127 non-pronominal mentions and 645 1st person pronouns, 302 2nd person pronouns and 675 3rd person pronouns (see Table 1).

The Chinese audio stimulus was a Chinese web version of *The Little Prince* [1], read by a professional female Chinese broadcaster. Within this text, 2947 non-pronominal mentions and 639 1st person pronouns, 298 2nd person pronouns and 529 3rd person pronouns are identified (see Table 1). The offset time of the pronouns and the non-pronominal mentions in the English and Chinese audiobook are marked as 1 and are entered as binary regressors into our GLM analysis.

6.3 Procedure

After giving their informed consent, participants were familiarized with the MRI facility and assumed a supine position on the scanner. The presentation script was written in PsychoPy (Peirce, 2007). Auditory stimuli were delivered through MRI-safe, high-fidelity headphones (English: Confon HP-VS01, MR Confon, Magdeburg, Germany; Chinese: Ear Bud Headset, Resonance Technology, Inc, California, USA) inside the head coil. The headphones were secured against the plastic frame of the coil using foam blocks. An experimenter increased the sound volume stepwise until the participants could hear clearly.

The English and Chinese audiobooks last for about 94 and 99 minutes, respectively. They were both divided into nine sections, each lasts for about ten minutes. Participants then listened passively to the nine sections and completed four quiz questions after each section (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. The entire session lasted around 2.5 hours.

---

[1] http://www.xiaowangzi.org

6.4 MRI Data Collection and Preprocessing

Both English and Chinese brain imaging data 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 (MP-RAGE) pulse sequence. Blood-oxygen-level-dependent (BOLD) functional scans were acquired using a multi-echo planar imaging (ME-EPI) sequence with online reconstruction (TR=2000 ms; TE's=12.8, 27.5, 43 ms; FA=77°; matrix size=72 x 72; FOV=240.0 mm x 240.0 mm; 2 x image acceleration; 33 axial slices, voxel size=3.75 x 3.75 x 3.8 mm). Cushions and clamps were used to minimize head movement during scanning.

All fMRI data is preprocessed using AFNI version 16 (Cox, 1996). The first 4 volumes in each run were excluded from analyses to allow for T1-equilibration effects. Multi-echo independent components analysis (ME-ICA; Kundu et al.,2012) were used to denoise data for motion, physiology and scanner artifacts. Images were then spatially normalized to the standard space of the Montreal Neurological Institute (MNI) atlas, yielding a volumetric time series resampled at 2 mm cubic voxels.

6.5 Statistical Analysis

At the single subject level, the observed BOLD time course in each voxel were modeled by the difficulty of pronoun resolution derived by the ACT-R model for each 1st, 2nd and 3rd person pronoun time-locked at the offset of each pronoun in the audiobook. We also include a binary regressor for non-pronominal mentions time-locked at the its offset.

To further examine the status of ACT-R activation as a cognitive model for pronoun resolution, We did a second GLM analysis where we correlated only the binary regressors time-locked at the offset of each 1st, 2nd, 3rd person pronoun and each non-pronominal mention in the audiobook.

These two analyses both included three control variables of non-theoretical interest: *RMS intensity* at every 10 ms of the audio; *word rate* at the offset of each spoken word in time; *frequency* of the individual words in Google Book unigrams [2]. These regressors were added to ensure that any conclusions about pronoun resolution would be

---

[2] http://books.google.com/ngrams

specific to those processes, as opposed to more general aspects of speech perception.

At the group level, for each group, the activation maps for the ACT-R activation regressors and the binary regressors for 3rd person pronouns were computed using one sample t-test. The voxelwise threshold was set at $p \leq 0.05$ FWE, with an extent threshold of 50 contiguous voxels ($k \geq 50$). Contrasts of the activation maps between the two groups were examined by a factorial design matrix, and statistical threshold was also set at $p \leq 0.05$ FWE. The GLM analysis was performed using SPM12 (Penny et al., 2011).

We only focused on the results of 3rd person pronouns because they provide gender information in English but not Chinese, which points to potentially different brain activation maps. In addition, 3rd person pronouns have been suggested to differ from 1st and 2nd person pronouns in that 1st and 2nd person pronouns mark proximity in space and 3rd person pronouns are further away (Ariel, 1990).

7 fMRI Results

For English speakers, the largest clusters for 3rd person pronouns, which represents brain activity associated with the presence 3rd person pronouns were observed in the right Temporal Pole, and the left Middle Temporal Gyrus (MTG). For Chinese speakers, the presence of 3rd person pronouns is associated with increased activity in the left Middle Temporal Gyrus (MTG; $p < 0.001$ FWE; see Figure 2a and Figure 2b). Direct comparison of the contrast maps between the English and Chinese group suggests stronger activity in the right Angular Gyrus, the right MTG and the right Precuneus for English speakers and stronger activity in the left Angular Fyrus for Chinese speakers ($p < 0.05$ *FWE*).

Brain regions showing an increased activation for pronouns with higher processing difficulty predicted by our ACT-R measure (i.e., the negative of ACT-R activation level) include the left MTG and the left IFG and the left Superior Frontal Gyrus (SFG) for English speakers ($p < 0.001$ FWE; see Figure 2c), whereas Chinese speakers has peak clusters in the left Angular Gyrus, the left SFG and the left MTG ($p < 0.05$ FWE; see Figure 2d)).

The difference between the ACT-R difficulty measure for 3rd person pronoun resolution in Chinese and English is confirmed by the direct com-

parison reported in Table 3c ($p < 0.05\ FWE$). Although the cluster size is relatively small at the corrected threshold, Chinese ACT-R effect significantly differs form English in the Angular Gyrus, and English ACT-R effect shows a more right lateralized activation of the language network involving the Precuneus cortex. Table 3b lists of all the significant clusters using region names from the Harvard-Oxford Cortical Structure Atlas.

8 Discussion

ACT-R activation levels for 3rd person pronouns in Chinese are significantly higher than that in English (see Section 5), demonstrating as predicted a stronger reliance of Chinese on salience compared to English. It also relates to the rich linguistic literature on Chinese as a discourse-oriented language (Xu, 2006). When used as a regressor against hemodynamic responses to naturalistic text, ACT-R activation level reveals different left-lateralized activation patterns in English and Chinese, supporting different model performance on pronoun resolution in the two languages (cf. Table 2)

Only English shows a significant increased activation in the left Broca's area pars Triangularis, which has been recurrently reported as correlating with syntactic processing cost linked to antecedent pronoun (Santi and Grodzinsky, 2012), and particularly to the distance between the antecedent and the pronoun (Matchin et al., 2014; Santi and Grodzinsky, 2007).

Chinese, on the other hand, shows greater activation in the Angular Gyrus for the ACT-R difficulty measure compared to English. Notably, previous literature on German gender agreement in anaphoric reference reported increased activation in the Angular Gyrus (BA 39) for incongruent biological gender matching (Hammer et al., 2007). One interpretation of the result, therefore, is that the ACT-R model predicts processing effort when the antecedent is not salient or when there are equally salient competitors. In these cases, only English speakers could use morpho-syntactic cues such as gender to identify the correct antecedent, whereas Chinese speakers might have to work extra hard to accomplish gender matching using the discourse information.

The difference between the brain areas highlighted by the Chinese and English ACT-R results, is confirmed by the simpler GLM analysis that correlated brain activity with only the presence of pronouns in the text (cf. GLM Analysis 2 in Section 6.5). Whole-brain pronouns effects were observed with a temporally distributed response patterns (for backward anaphora see Matchin et al.; for intra-sentential co-referential link see Fabre), with the additional involvement of the right Broca's area for English.

In sum, this study highlighted brain areas involved in the discourse and syntactic dimension of pronoun-antecedent linking as modeled in the ACT-R activation level of pronoun resolution.

9 Conclusion

ACT-R activation levels of antecedents suggest interesting differences between the two languages, English appears to have significantly lower ACT-R activation levels, hence higher difficulty for salience-based pronoun resolution. This difficulty, as suggested in the fMRI results of the present study, is associated with activity in the left Broca's area.

Although the ACT-R model accuracy was higher for Chinese, the difficulty measure of pronoun resolution yields a greater activation in an area where non-grammatical gender information is processed. This suggests that the salience-based pronoun resolution model is more effortful when no gender information is encoded in the language.

The current study only compares pronoun resolution in speech, thus the conclusion may not hold for reading comprehension where pronouns with different gender have different forms in Chinese as well. Future directions may include comparing brain activity during pronoun resolution in both reading and listening comprehension. Overall, these results show that crossing computational approach and naturalistic ecologically valid linguistic stimuli to tease apart strongly interwoven cognitive processes is a promising perspective in neuroimaging. As such, they pave the way for increasing cross-fertilization between computational linguistics and the cognitive neuroscience of language.

10 Acknowledgments

(a) T-score map for the binary 3rd person pronoun effect in English

(c) T-score map for the ACT-R difficulty measure for 3rd person pronouns in English

(b) T-score map for the binary 3rd person pronoun effect in Chinese

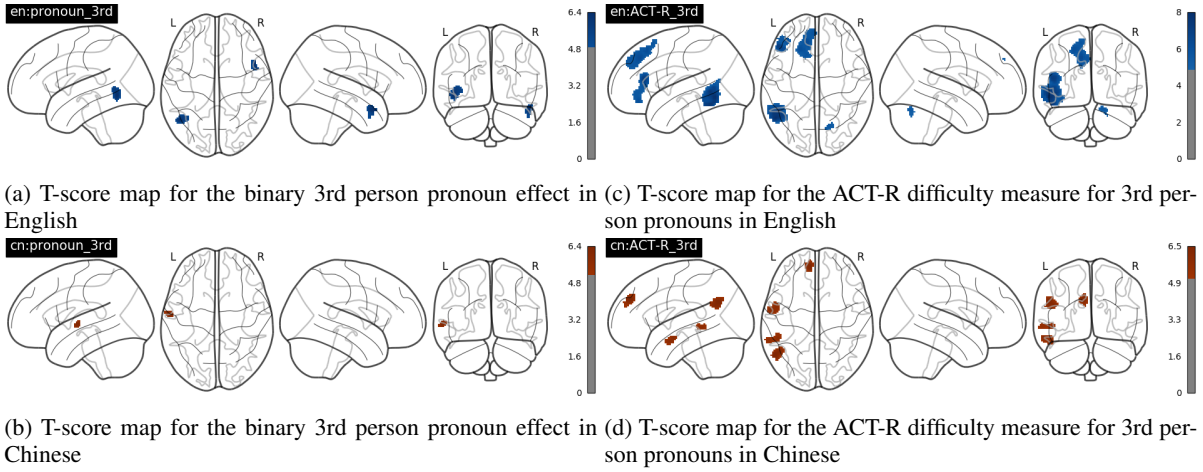(d) T-score map for the ACT-R difficulty measure for 3rd person pronouns in Chinese

Figure 2: Whole-brain effect with significant clusters for (a) binary 3rd person pronouns effect in English, (b) binary 3rd person pronouns effect in Chinese, (c) ACT-R difficulty measure for 3rd person pronoun resolution in English and (d) ACT-R difficulty measure for 3rd person pronoun resolution in Chinese. All images underwent *FWE* voxel correction for multiple comparisons with p < 0.05.

3rd person pronoun effect	MNI coordinates			Region	*p*-value *FWE-corr*	*k*-size *cluster*	*t*-score *peak*
	x	y	z				
English	50	12	-26	right Temporal Pole	0.001	101	6.4
3rd pronoun	-46	-60	-8	left Inferior/Middle Temporal Gyrus	0.001	194	6.34
Chinese	-60	-10	0	left Superior//Middle Temporal Gyrus	0.003	57	6.36
3rd pronoun							

(a) Significantly activated clusters by the binary 3rd person pronoun effect in English and Chinese ($p < 0.05$ *FWE*)

ACT-R difficulty measure	MNI coordinates			Region	*p*-value *FWE-corr*	*k*-size *cluster*	*t*-score *peak*
	x	y	z				
English	-52	-58	-10	left Middle Temporal Gyrus	< 0.001	1074	7.96
ACT-R 3rd pronoun	-48	34	14	left Inferior Frontal Gyrus	< 0.001	494	7.20
	-12	40	40	left Superior Frontal Gyrus	0.001	724	6.36
	18	-68	-26	right Cerebellum	0.009	61	5.53
Chinese	-54	-62	28	left Angular Gyrus	0.002	217	6.51
ACT-R 3rd pronoun	-8	52	30	left Superior Frontal Gyrus	0.002	150	6.50
	-58	-4	-18	left anterior Middle Temporal Gyrus	0.006	167	4.95
	-50	-42	-4	left posterior Middle Temporal Gyrus	0.007	117	4.90

(b) Significantly activated clusters by the ACT-R difficulty measure for 3rd person pronouns in English and Chinese ($p < 0.05$ *FWE*)

Comparison of ACT-R difficulty measure	MNI coordinates			Region	*p*-value *FWE-corr*	*k*-size *cluster*	*t*-score *peak*
	x	y	z				
English > Chinese	4	-40	44	right Precuneus Cortex	0.008	55	5.23
ACT-R 3rd	68	-46	10	right Middle Temproal Gyrus	0.01	27	5.19
	64	-48	24	right Angular Gyrus	0.015	70	5.07
Chinese > English	-52	-62	28	left Angular Gyrus	0.033	6	4.82
ACT-R 3rd pronoun							

(c) Contrast between the ACT-R difficulty measure for 3rd person pronouns in English versus Chinese ($p < 0.05$ *FWE*)

Table 3: Significant clusters of BOLD activation for (a) 3rd person pronouns, (b) ACT-R difficulty measure on 3rd person pronoun in English and Chinese and (c) their contrast after FWE voxel correction for multiple comparisons with $p < 0.05$. Peak activations are given in MNI Coordinates.

References

Hiyan Alshawi. 1987. *Memory and context for language interpretation.* Cambridge University Press.

John R Anderson. 2007. *How can the human mind occur in the physical universe?* Oxford University Press, Oxford.

M. Ariel. 1990. *Accessing noun-phrase antecedents.* Routledge, London, UK.

Jennifer E. Arnold. 2010. How speakers refer: The role of accessibility. *Language and Linguistics Compass*, 4(4):187–203.

Susan E. Brennan, Marilyn W. Friedman, and Carl J. Pollard. 1987. A centering approach to pronouns. In *Proceedings of the 25th annual meeting on Association for Computational Linguistics*, pages 155–162. Association for Computational Linguistics.

Noam Chomsky. 1982. *Some concepts and consequences of the theory of government and binding.* MIT Press, Cambridge, Mass.

R. W. Cox. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, an International Journal*, 29(3):162–173.

Murielle Fabre. 2017. *The sentence as cognitive object - The neural underpinnings of syntactic complexity in Chinese and French.* Ph.D. thesis, INALCO Paris.

P. C. Gordon, B. J. Grosz, and L. A. Gilliom. 1993. Pronouns, names, and the centering of attention in discourse. *Cognitive Science*, 17:311–47.

S. B. Greene, G. McKoon, and R. Ratcliff. 1992. Pronoun resolution and discourse models. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 18:266–83.

Barbara J. Grosz, Scott Weinstein, and Aravind K. Joshi. 1995. Centering: A framework for modeling the local coherence of discourse. *Computational linguistics*, 21(2):203–225.

Anke Hammer, Rainer Goebel, Jens Schwarzbach, Thomas F. Münte, and Bernadette M. Jansma. 2007. When sex meets syntactic gender on a neural basis during pronoun processing. *Brain Research*, 1146:185–198.

C.-T. James Huang. 1989. Pro-drop in Chinese: A generalized control theory. In Osvaldo Jaeggli and Kenneth Safir, editors, *The null subject parameter*, pages 185–214. Springer.

Robert Kantor. 1977. *The management and comprehension of discourse connection by pronouns in English.* Ph.D. thesis, Ohio State University.

Prantik Kundu, Souheil J. Inati, Jennifer W. Evans, Wen-Ming Luh, and Peter A. Bandettini. 2012. Differentiating BOLD and non-BOLD signals in fMRI time series using multi-echo EPI. *NeuroImage*, 60(3):1759–1770.

Shalom Lappin and Herbert J. Leass. 1994. An algorithm for pronominal anaphora resolution. *Computational linguistics*, 20(4):535–561.

Charles N. Li and Sandra A. Thompson. 1976. Subject and topic: A new typology. In Charles N. Li, editor, *Subject and topic*, pages 457–89. Academic Press, New York.

William Matchin, Jon Sprouse, and Gregory Hickok. 2014. A structural distance effect for backward anaphora in Broca's area: An fMRI study. *Brain and Language*, 138:1–11.

Brian McElree. 2001. Working memory and focal attention. *Journal of experimental psychology. Learning, memory, and cognition*, 27(3):817–835.

Jonathan W. Peirce. 2007. PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1-2):8–13.

William Penny, Karl Friston, John Ashburner, Stefan Kiebel, and Thomas Nichols. 2011. *Statistical parametric mapping: The analysis of functional brain images.* Academic Press.

Jacolien van Rij, Hedderik van Rijn, and Petra Hendriks. 2013. How WM load influences linguistic processing in adults: A computational model of pronoun interpretation in discourse. *Topics in Cognitive Science*, 5(3):564–580.

Andrea Santi and Yosef Grodzinsky. 2007. Taxing working memory with syntax: Bihemispheric modulations. *Human Brain Mapping*, 28(11):1089–1097.

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

Pontus Stenetorp, Sampo Pyysalo, Goran Topić, Tomoko Ohta, Sophia Ananiadou, and Jun'ichi Tsujii. 2012. BRAT: a web-based tool for NLP-assisted text annotation. In *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, pages 102–107. Association for Computational Linguistics.

Leila Wehbe, Ashish Vaswani, Kevin Knight, and Tom M. Mitchell. 2014. Aligning context-based statistical models of language with brain activity during reading. In *EMNLP*, pages 233–243.

Liejion Xu. 2006. Topicalization in asian languages. In Marting Everaert and Henk van Riemsdijk, editors, *The Wiley Blackwell Companion to Syntax, 2nd Edition*, pages 137–174. Hoboken, NJ: Blackwell Publishing, Hoboken, NJ.