MEMORY RETRIEVAL IN MANDARIN CHINESE REFLEXIVE PROCESSING: EVIDENCE AGAINST THE LOCAL SEARCH HYPOTHESIS

YUHANG XU
JEFFREY T. RUNNER
University of Rochester

1 Introduction

The Mandarin Chinese reflexive *ziji* is unique for its long-distance binding options (e.g., Tang, 1989, Huang, Li and Li, 2009). English reflexives must be bound within their immediate clause as in (1):

(1) John$_i$ thinks Tom$_j$ knows Bill$_k$ likes himself$_{*i/*j/k}$.

In (1), *himself* can only refer to Bill, not Tom or John. However, in a parallel example (adapted from Hsieh, 2015) in Mandarin Chinese as shown in (2), Bill, Tom and John can each be the antecedent of *ziji*:

(2) 约翰$_i$ 认为 汤姆$_j$ 知道 比利$_k$ 喜欢 自己$_{i/j/k}$
John$_i$ renwei Tom$_j$ zhidao Bill$_k$ xihuan *ziji*$_{i/j/k}$
John$_i$ think Tom$_j$ knows Bill$_k$ likes *ziji*$_{i/j/k}$
"John$_i$ thinks Tom$_j$ knows Bill$_k$ likes himself/him."

Also, the antecedent of *ziji* is required to be an animate and sentient subject (Tang, 1989). In (3), only John but not the football can be the antecedent of *ziji* because of the animacy requirement.

(3) 约翰$_i$ 说 足球$_j$ 打到-了 自己$_{i/*j}$
John$_i$ say football$_j$ hit-PRF *ziji*$_{i/*j}$
John$_i$ say football$_j$ hit-PRF *ziji*$_{i/*j}$
"John said that the football hit him/*itself."

Although the reflexive *ziji* has the long-distance binding option, a locality bias has been observed in the literature: speakers prefer the local binding option over the long-distance one in their final choices and expend less effort processing the local binding as well. This effect has been reported in many experimental studies on *ziji* using various techniques, for instance, event-related potentials/ERPs (e.g., Li and Zhou, 2010); self-paced reading (e.g., Chen, Jäger and Vasishth, 2012); the speed-accuracy tradeoff/SAT paradigm (e.g., Dillon et al., 2014) and eye-tracking during reading (e.g., Jäger, Engelmann and Vasishth, 2015).

The standard manipulation that researchers have used in these studies is to include one potential antecedent with one or more distractors varying in position in the sentence in order to detect the interference effect caused by inaccessible distractors during processing. While the interference effect shows some aspects of how people identify the antecedent in real time, studies on the precise time-course are needed in order to develop a psycholinguistic model on processing this long-distance reflexive.

Researchers have proposed several psycholinguistic models of the mechanisms people may use to identify and retrieve a linguistic constituent from working memory. For instance, a content-addressable (or cue-based) memory retrieval mechanism has been shown to be a reliable model to account for many aspects of human sentence processing (e.g., McElree, 2000). Recently, researchers have begun to test this model on the retrieval process of the long-distance reflexive *ziji* (Dillon et al., 2014, Jäger, Engelmann and Vasishth, 2015). In a multiple-response speed-accuracy tradeoff (MR-SAT) study examining the binding of *ziji* in a bi-clausal structure, varying the position of animate and inanimate subject NPs, Dillon et al. (2014) found increased latency when an inanimate NP was the subject of the local clause. For instance, in sentences like (4a) and (4b)[1], only *Coach Zhang* can be the antecedent of *ziji* (due to the animacy requirement). However, people will spend more time on (4a) than (4b) to determine the correct antecedent (*Coach Zhang*).

(4a) 张教练 表明 那篇 报导 低估-了 自己
Zhangjiaolian biaoming napian baodao digu-le ziji
Coach Zhang say that report underestimate-PRF *ziji*
"Coach Zhang says that that report underestimated him"

(4b) 那篇 报导 表明 张教练 低估-了 自己
napian baodao biaoming zhangjiaolian digu-le ziji
That report say Coach Zhang underestimate-PRF *ziji*
"That report says that Coach Zhang underestimated himself"

Dillon et al. (2014) interpreted this as a locality bias: antecedent search starts locally, only extending long distance when a local antecedent is not possible. They further proposed the Local Search Hypothesis suggesting that people do use positional (or syntactic) information as a cue to guide their retrieval process:

> The parser uses positional syntactic information during the retrieval of syntactic dependents, and positional cues serve to restrict retrieval to constituents in some local syntactic domain. (p.3)

[1](4a) and (4b) were adapted from the critical conditions used in Dillon et al. (2014)

This hypothesis implies a structural sensitivity in human parsers' memory access when interpreting long distance dependencies. Because Mandarin *ziji* lacks phi-features (e.g., gender, number), it seems natural that positional or syntactic cues would be relevant during the retrieval process. However, it is still an open question which features people use as retrieval cues under the cue-based retrieval framework of sentence processing (for a review, see Jäger, Engelmann and Vasishth, 2015). It is even less clear what role of syntactic structure plays during anaphor resolution within this framework. Although some researchers propose that syntactic constraints play a dominant role in processing anaphoric dependencies (e.g., Dillon et al., 2013, Parker, Shvartsman and Van Dyke, 2017), other researchers instead propose a multiple-constraints approach that both structural and non-structural (e.g., semantic) constraints influence the anaphora resolution starting from the earliest moment of processing (e.g., Badecker and Straub, 2002, Runner, Sussman and Tanenhaus, 2003, Kaiser et al., 2009). The long-distance reflexive *ziji* in Mandarin Chinese provides an excellent test bed for these questions. The requirements for *ziji*'s antecedent are simple, i.e., an animate subject in the same sentence as *ziji* according to the literature (e.g., Huang, Li and Li, 2009), making it possible to control other confounding factors. *Ziji* also provides a chance to investigate the role that purely positional retrieval cues (e.g., local c-command) play in reflexive processing, something which is difficult to test in other languages, like English.

Dillon et al. (2014) first explored the time-course of how people identify *ziji*'s antecedent and found a structural sensitivity in people's search, which led them to propose the Local Search Hypothesis. However, the task they used, the speed-accuracy trade-off paradigm, can only infer the time-course of the antecedent search indirectly through the complexity-based latency difference. In addition, their materials only contained one possible antecedent in each sentence, so it is still unclear how antecedent search will proceed when multiple possible antecedents are available, a critical condition for testing the Local Search Hypothesis.

In this working paper, we present two experiments aimed at testing the Local Search Hypothesis directly. The first experiment is an off-line judgment task designed to determine people's preferences when facing multiple potential antecedents for *ziji*. The second is an eye-tracking study designed to observe the time-course of people's search during reflexive processing. In order to investigate the time course of people's search for the antecedent of *ziji* in real time, we used the visual world eye-tracking paradigm (Tanenhaus et al., 1995), which is well-known for its close time-locking to the linguistic input. The task for this paradigm is referential in nature, enabling the examination of the time course of anaphora resolution (Runner, Sussman and Tanenhaus, 2003, Kaiser et al., 2009). This paradigm, although it has seldom been used to study this aspect (see one attempt in Sekerina, Campanelli and Van Dyke, 2016), may be a better tool to explore memory retrieval during sentence processing. Unlike paradigms like ERPs, which requires a high proportion of anomalous stimuli, the visual world paradigm can give us a precise time course of people's sentence processing in relatively natural settings.

2 Experiment 1: offline judgment task

There were aims for conducting this offline judgment study. First, we wanted to understand people's preferences (final judgments) when facing multiple potential antecedents for *ziji*. In Experiment 1 (and also in Experiment 2), we used materials in a tri-clausal structure containing two animate NPs which are both appropriate antecedents for *ziji*. According to some scholars (e.g., Tang, 1989), the intervening animate subject NP cannot be the long-distance binder for *ziji* in a

tri-clausal sentence. For instance, in (5) below, it is claimed to be impossible for *ziji* to take *Lisi* (the intermediate subject) as its antecedent.

(5) 吴三$_i$ 知道 李四$_j$ 觉得 王五$_k$ 对 自己$_{i/*j/k}$ 没 信心
Wusan$_i$ zhidao Lisi$_j$ renwei Wangwu$_k$ dui *ziji*$_{i/*j/k}$ mei xinxing
Wusan$_i$ know Lisi$_j$ think Wangwu$_k$ to *ziji*$_{i/*j/k}$ no confidence
"Wusan knew that Lisi thought that Wangwu had no confidence in himself/him (Wusan)"

Experiment 1 is designed to test if native speakers consider the intermediate NP to be the antecedent of *ziji* in a tri-clausal structure. Previous studies using off-line judgment tasks, usually used a forced-choice task to detect people's preference for the antecedent of *ziji*. However, this approach does not allow us to determine if people also consider other potential antecedents at the same time. So, in this experiment, we used the "check-all-apply" task allowing participants to provide all the responses they judged as acceptable. The second aim of this experiment was to serve as a norming study to make sure that both long-distance and local binding options were possible in the materials that we could then use for the visual world eye-tracking study.

2.1 Participants

The experiment was performed using a web-based survey platform *sojump* (www.sojump.com). 125 native Chinese speakers from Mainland China participated in this experiment voluntarily without receiving any payment. Five of them were excluded because of low accuracy (< 85%) on the catch questions or unreasonably fast completion time. The results from the remaining 120 participants (51 females) were used for the final analysis.

2.2 Stimuli and Design

A survey consisting of 21 critical items and 30 fillers was conducted to investigate native Mandarin speakers' interpretations of *ziji* when multiple accessible antecedents were available. The critical conditions were created in a tri-clausal structure (NP_1-V_1-NP_2-V_2-NP_3-V_3-*ziji*) containing two animate subject NPs (ANP) and one inanimate subject NP (INP). Verbs that select for a sentential complement (e.g., 认为 *think*) were used in the matrix clause (V_1) and middle clause positions (V_2). Verbs in the local clause (V_3) were chosen from a separate norming study to ensure that both local and distance binding is possible (see section 6 for a description of the norming task)[2]. Participants were asked to indicate who *ziji* referred to by choosing all the responses they thought to be possible. The choices were the three NPs mentioned in that sentence. We manipulated the position of the inanimate NP in each critical condition (see

[2] It is well observed in the literature (e.g., Li and Zhou, 2010, Schumacher, Bisang and Sun, 2011, He, 2014) that verb semantics will also affect the interpretation of *ziji*. For instance, for the example below, only *wo* 'I' can be the antecedent of *ziji* because of the semantics of the verb *genzong* 'follow': you can only follow other people, not yourself.

我 说 比尔 在 公园 跟踪-了 自己
Wo$_i$ shuo Bill$_j$ zai gongyuan genzong-le ziji$_{i/*j}$.
I$_i$ say Bill$_j$ in the park follow-perf self$_{i/*j}$.
'I said that Bill followed me in the park.'

Table 1 for a set of sample materials):

- Condition A: Inanimate NP in the local clause;
- Condition B: Inanimate NP in the middle clause;
- Condition C: Inanimate NP in the matrix clause.

Table 1 Conditions and sample materials for Experiment 1.

Conditions:	**Questions and Choices**
A: ANP_1-V_1-ANP_2-V_2-INP-PROG-V_3-*ziji* B: ANP_1-V_1-INP-V_2-ANP_2-PROG-V_3-*ziji* C: INP-V_1-ANP_1-V_2-ANP_2-PROG-V_3-*ziji* ANP = animate NP INP = inanimate NP	Q: 句中的自己指的是？ Q: What does *ziji* refer to in this sentence? 不定项选择/[Check all that apply] CHOICES (three NPs mentioned in this sentence): A.王老师/ Teacher Wang B.张厨师/ Chef Zhang C.那篇报导/ The report

Sample materials:

	NP(Long)	V_1	NP(Middle)	V_2	NP(Local)	V_3	*ziji*
A	王老师	认为	张厨师	表示	那篇报导	在-低估	自己
	Teacher Wang	think	Chef Zhang	say	the report	PROG-underestimate	*ziji*
B	王老师	认为	那篇报导	表示	张厨师	在-低估	自己
	Teacher Wang	think	the report	say	Chef Zhang	PROG-underestimate	*ziji*
C	那篇报导	认为[3]	王老师	表示	张厨师	在-低估	自己
	The report	think	Teacher Wang	say	Chef Zhang	PROG-underestimate	*ziji*

2.3 Results

Statistical analyses were conducted in R (R Core Development Team, 2014).

Table 2 shows the proportions of different antecedent choices for each condition. First, as predicted in the literature, *ziji* has a strong bias to select an animate NP as its antecedent and the choices of the inanimate NPs did not show significant differences across the three conditions ($p > .05$).

Second, because of people's strong bias for choosing the animate NPs as the antecedents, we compared two animate NPs in each condition to see people's preference in terms of the position (i.e., more local animate NP vs. less local animate NP. NP_2 was always the more local animate NP). Using McNemar's chi-square test, we found a strong preference for the NP in the middle clause position, except in Condition B where this was not possible since the middle NP was

[3] The expression "the report thinks…" sounds natural in Chinese, which means "the opinions expressed in the report…"

inanimate. (Comparing two animate NPs: Condition A: $NP_1 < NP_2$, $\chi^2 = 70.9$, $p < .001$; Condition B: $NP_1 < NP_2$, $\chi^2 = 21.1$, $p < .001$; Condition C: $NP_1 > NP_2$, $\chi^2 = 8.1$, $p < .005$).

Table 2 Proportions of choices under each condition in Experiment 1.

	Animate NP 1 (ANP1) %	Animate NP 2 (ANP2) %	Inanimate NP (INP) %
Condition A (AAI)	55.83	**78.69**	9.64
Condition B (AIA)	62.02	74.40	6.55
Condition C (IAA)	**72.62**	64.76	3.33

2.4 Discussion

The Local Search hypothesis proposed by Dillon et al. (2014) does not predict the participants' judgments shown in our study. Following their proposal that the antecedent search starts within the local clause first, we should observe at least these patterns: first, the choices of the animate NP which appears in a more local clause (NP_2) should be the preferred one across all conditions. However, we find an exception in Condition c, where the less local animate NP (NP_1) is the preferred among other choices. What's more, we might expect significantly more incorrect choices of the inanimate NP in condition a than in the other conditions, but the choices of the inanimate NP did not show a significant difference across conditions (though it is still numerically more than other conditions).

Also, our results refute the claim that it is impossible for *ziji* to take the intermediate subject as its antecedent in a tri-clausal structure (e.g., Tang, 1989): when given a chance to select all potential options, people will consider both animate NPs. Interestingly, participants showed a preference for the animate subject NP embedded right below the inanimate subject's verb, and when there was no lower animate subject NP, they preferred the NP just preceding it. Though it could be a task effect that people were allowed to choose all that they thought to be correct, more work will be required to understand the theoretical ramifications of this empirical result. In short, our materials are sound and ready to test in the eye-tracking study.

3 Experiment 2: visual world eye-tracking study

In Experiment 1, we showed that the pure "local bias" cannot explain the pattern observed from people's interpretations of *ziji*. However, an off-line judgment task itself is not sensitive to the time course of the antecedent retrieval process.

In order to detect people's interpretation process in real-time, in Experiment 2, we employed the visual world eye-tracking paradigm pioneer by Tanenhaus et al. (1995) to test people's search patterns during *ziji* resolution in real time. The visual world eye-tracking paradigm is a powerful tool to study anaphora resolution(e.g., Runner, Sussman and Tanenhaus, 2003), for its task is referential in nature and provides processing information closely time-locked to the input

(Salverda and Tanenhaus, 2017). We used materials similar to those in Experiment 1 but in an auditory version with visual displays representing the referents mentioned.

3.1 Participants

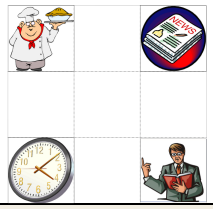
30 native Chinese speakers (25 females) with normal hearing and vision (or corrected-to-normal vision) were recruited from the University of Rochester community and were paid 10 dollars for participation.

3.2 Stimuli and Design

21 critical trials were interspersed among 43 filler trials to investigate how native Chinese speakers identify the antecedent of *ziji* in tri-clausal structures (NP_1-V_1-NP_2-V_2-NP_3-V_3-*ziji*) manipulating the positions of two animate and one inanimate subject NP. The stimuli were based on those used in Experiment 1 but modified so that the two animate NPs were all occupational nouns (e.g., *doctor*, *chef*, etc.); thus, the participants did not need a familiarization phase before the experiment to be acquainted with the characters they would hear about in the formal experiment (e.g., an image for a person called *Zhangsan*)[4].

Table 3 illustrates a set of sample materials. The embedded verbs (V_3) had been normed separately for Experiment 1 to ensure that both local and long-distance binding options were possible (again, see section 6 for a description of the norming study). A female native Mandarin Chinese speaker recorded all sentences at a sampling rate of 44.1 kHz. The visual stimuli were also normed to ensure that the images accurately represented the NPs used in the materials.

Table 3 Conditions and sample materials used in Experiment 2.

Conditions:	**Task and Display**
A: ANP_1-V_1-ANP_2-V_2-INP-PROG-V_3-*ziji* B: ANP_1-V_1-INP-V_2-ANP_2-PROG-V_3-*ziji* C: INP-V_1-ANP_1-V_2-ANP_2-PROG-V_3-*ziji* *ANP= animate NP* *INP= inanimate NP*	Task: Click on the picture that represents the last word you have heard. Display: Pictures represent three NPs mentioned in the sentence and one distractor. 
Sample spoken materials:	

[4] In Experiment 1, we used people's names (e.g., *Zhangsan*, *Lisi*) as well as occupational nouns (e.g., *Teacher Wang*) for animate NPs while in Experiment 2 we used occupational nouns only.

	NP(Long)	V_1	NP(Middle)	V_2	NP(Local)	V_3	ziji
A	王老师 Teacher Wang	认为 think	张厨师 Chef Zhang	表示 say	那篇报导 the report	在-低估 PROG-underestimate	自己 ziji
B	王老师 Teacher Wang	认为 think	那篇报导 the report	表示 say	张厨师 Chef Zhang	在-低估 PROG-underestimate	自己 ziji
C	那篇报导 The report	认为 think	王老师 Teacher Wang	表示 say	张厨师 Chef Zhang	在-低估 PROG-underestimate	自己 ziji

3.3 Procedure

Participants clicked on a crosshair at the center of the computer screen to begin a trial. Then the visual display containing four visual referents immediately appeared for previewing (four pictures placed in a 3×3 grid, one in each corner cell of the grid. See
Table 3 *Task and Display*). The audio stimuli played through headphones 1000ms after the preview display appeared. Participants were instructed to click on the picture that represents the last word they had heard at the end of each trial. For critical conditions, the last word was the reflexive *ziji*. Additionally, for critical conditions, a flashing dot appeared for 250ms at the center of the screen at the onset of the aspect marker 在 (PROG) to distract fixations temporarily from one of the four pictures on the screen. For filler conditions, the position, duration and onset time of the flashing dot were all randomized across trials. Participants' eye movements were recorded using an Eyelink II head-mounted eye-tracker sampling at 500 Hz. A drift-correction was performed every four trials. A verbal working memory test , *Chinese Reading Span Task* taken from Chen, Deng and Jiang (2008), was also performed to detect the working memory capacity for each participant after the eye-tracking experiment. Figure 1 illustrates the procedure of a sample trial in this experiment.

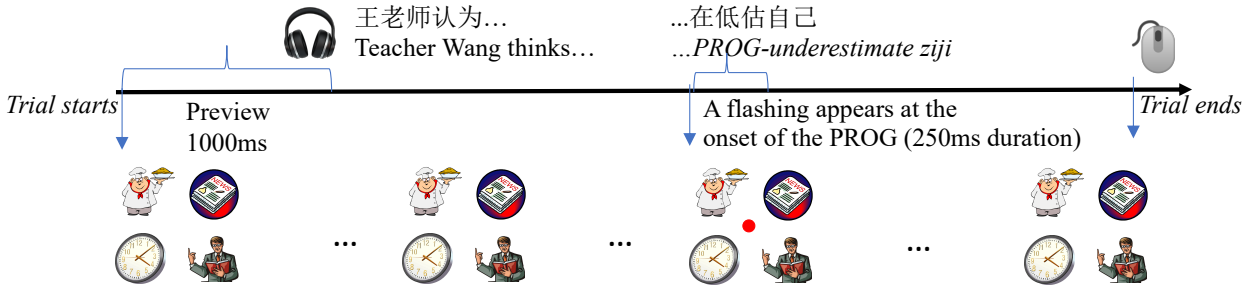


Figure 1. Procedure of a sample trial in Experiment 2.

3.4 Results

3.4.1 Click data

Similar to the pattern we observed in Experiment 1, antecedent choices showed that participants had a strong bias to choose the animate rather than inanimate NPs as the antecedents of *ziji* (see Figure 2): the inanimate NPs were only chosen 1.4% of the time in Condition A, 4.4% in Condition B and 4.3% in Condition C. The choices for the inanimate NPs did not differ significantly across three conditions (χ^2 =.019, p =.892). For the animate NPs, target choices indicated an overall preference to ultimately choose the more local NP: 53.8% (ANP_2) vs. 44.8% (ANP_1) in Condition

A (χ^2 = 9.01, p < .005), 64.2% vs. 31.4% in Condition B (χ^2 = 13.478, p < .001), and 57.1% vs. 38.6% in Condition C (χ^2 = 12.01, p < .001).

Unlike the pattern in Experiment 1, participants preferred the most local animate NP across all three conditions, which may be due to the different tasks we used in these two off-line experiments: In Experiment 1, people could choose all of the options they thought to be the correct antecedent. However, in Experiment 2, people were only allowed to choose the "best" option as the antecedent. Taken together, a local preference for the antecedent of *ziji* was observed in participants' final interpretations.

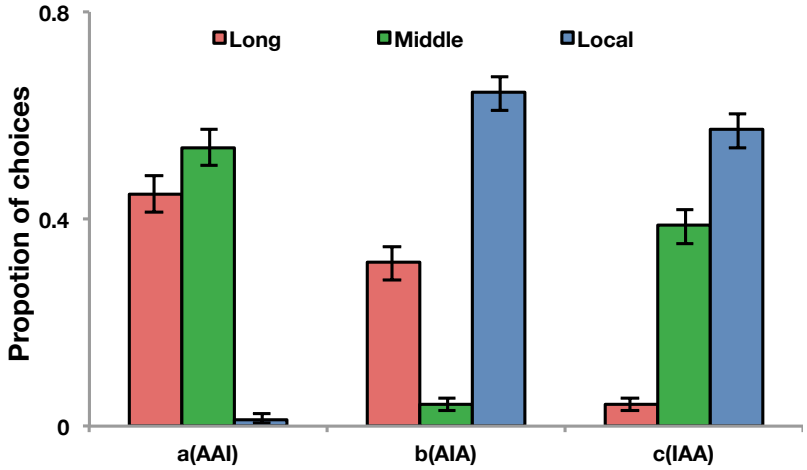


Figure 2. Proportion of antecedent choices for each condition in Experiment 2.

3.4.2 Eye-tracking data

The raw eye-tracking data were pre-processed and visualized using the *eyetrackingR* package (Dink and Ferguson, 2018). Statistical analyses were conducted using linear-mixed effects regressions implemented in the *lme4* and *lmerTest* packages (Kuznetsova, Brockhoff and Christensen, 2017). We set the 200ms prior to the onset time of the reflexive *ziji* as the starting point of the interest period for analysis. Displayed in Figure 3 below, we plotted people's fixations to the four pictures in the display for all three conditions for 1200ms starting 200ms prior to the onset of *ziji* and ending 1000ms after the onset of *ziji*. (top-left: Animate NP_1; top-right: Animate NP_2; bottom-left: Distractor; bottom-right: Inanimate NP).

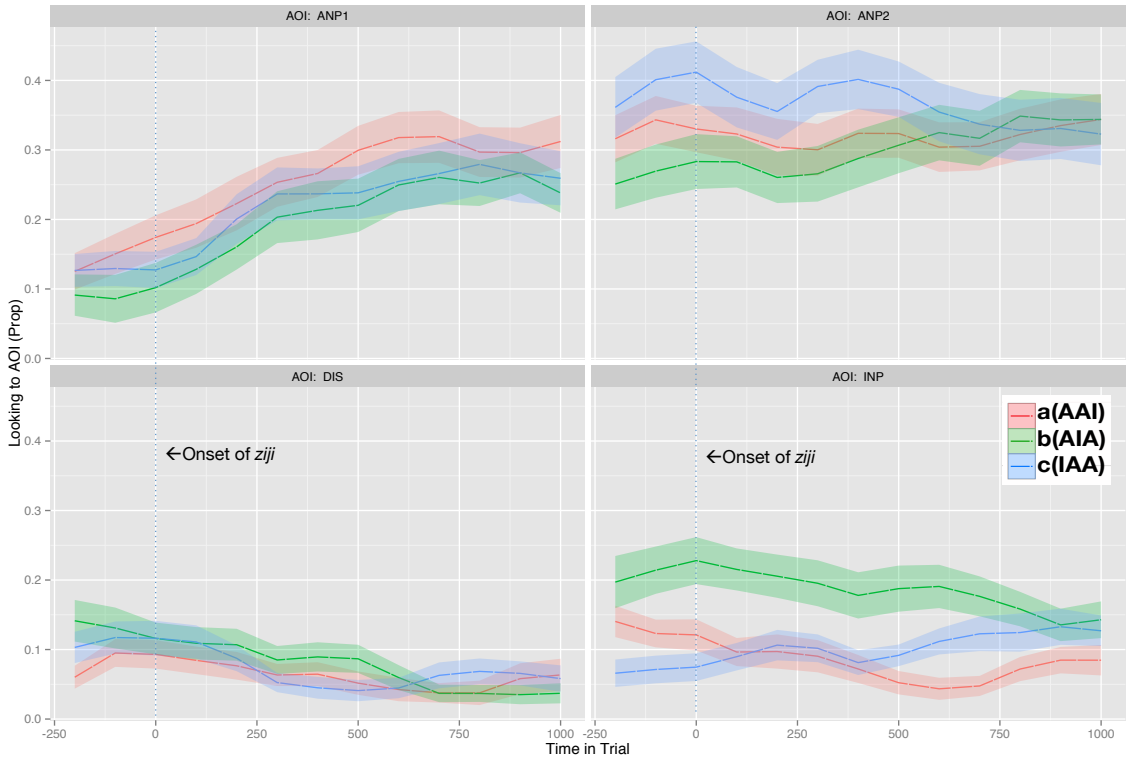


Figure 3. Proportion of fixations to the four objects (Animate NP_1, Animate NP_2, Inanimate NP and Distractor) over time for three conditions. Trials are aligned to the onset of the reflexive *ziji* at the 0 point.

First of all, we analyzed the pattern of the fixations to the inanimate NPs across the three conditions. Linear mixed effects models were used for analyzing the proportions (transformed using empirical logit) of the inanimate NP (INP) during the critical time window (from the onset of *ziji* to 500ms after the onset of *ziji*) with conditions and the working memory score as fixed factors and maximum random effect structures for both subjects and items, following recommendations in Barr et al. (2013). The results showed no main effect of working memory (WM) scores (β =.110, *SE* =.109, t =.743, p >.05) but did show a main effect for the conditions (β = -2.095, *SE* =1.087, t = -9.893, p <.0001). When we further explored the differences among the three conditions, we found a significant difference between conditions a and b (β = -0.693, *SE* = 0.186, t = -3.717, p < .001) as well as between conditions c and b (β = -0.475, *SE* = 0.186, t = -3.051, p < .001), but not between conditions a and c (β = 0.201, *SE* = 0.186, t = 1.107, p > .05).

Next, we focused on the eye-fixation patterns of the two animate NPs. As illustrated in Figure 3, for the animate NP1, the "slope" of the lines increased across the time window. However, the "slope" for animate NP2 began higher but stayed relatively flat. In order to see the competition between the two animate NPs during the retrieval process, we calculated a Locality Advantage Proportion (LAP) score for each sample point (ANP_2%-ANP_1%, ANP_2 was always the more local animate NP; see

Figure 4 below). This measure allowed us to compare across conditions the differential preferences for the more local and less local NPs during different windows. A positive score indicated a relative preference for NP_2 over NP_1 (i.e., *locality preference*), while a negative score reflected a relative preference for NP_1 over NP_2. The LAP was transformed through the empirical

logit function before analysis. Linear mixed effects models were used for analyzing the eye-tracking data, with conditions, time slices (400ms for each slice) and the working memory (WM) score as the fixed factors and maximum random structures for both items and subjects.

We found the LAP score decreased significantly over time in Window 1 (from the onset of *ziji* 0-400ms; β= -0.12, *SE* = 0.036, *t* = -4.89, *p* <.001), but not in Window 2 (400-800ms; β = -0.12, *SE* = 0.098, *t* = -1.17, *p* > .05) or Window 3 (800-1200ms; β = 0.095, *SE* = 0.087, *t* = 0.893, *p* > .05). The WM scores did not show a significant effect on LAP across the three windows (β=.072, *SE*=.031, *t*= .701, *p*>.05), indicating that the pattern we observed here was not specific to people's working memory capacity.

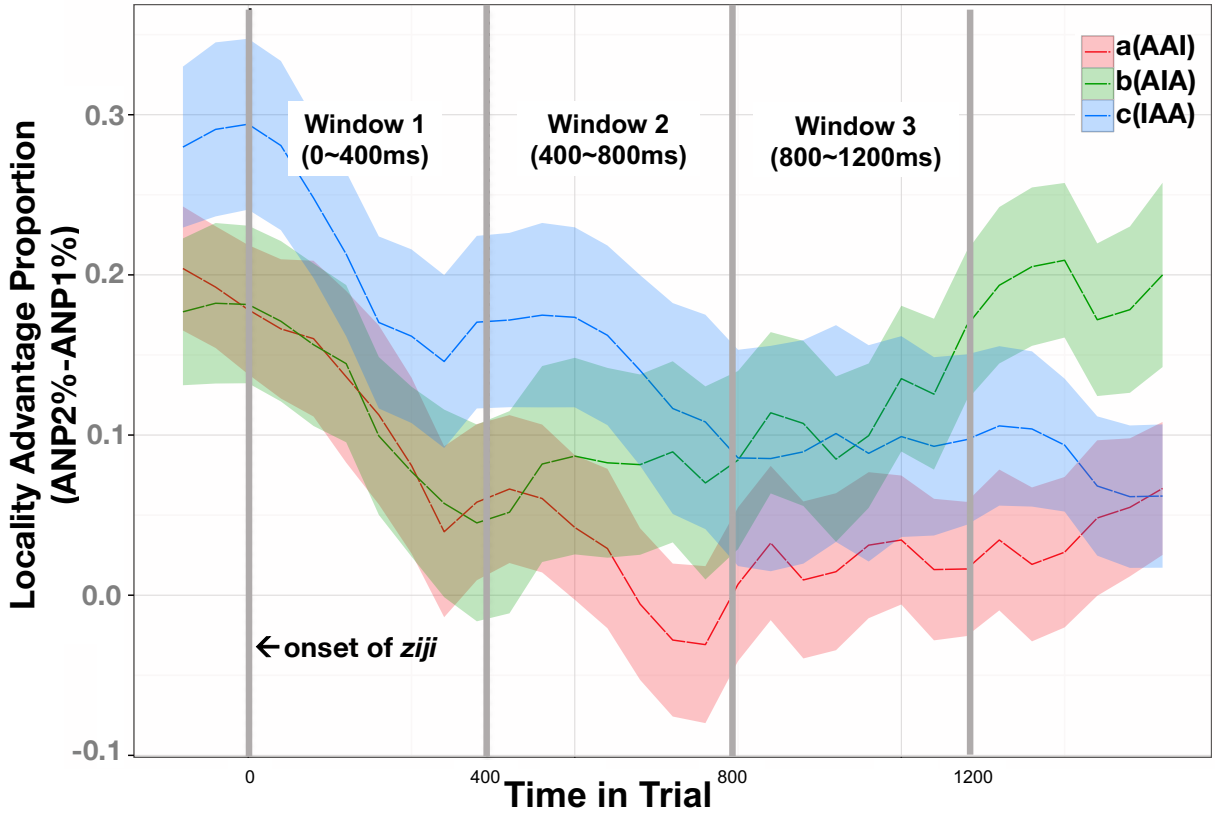


Figure 4. Locality Advantage Proportion (LAP) score (ANP_2%-ANP_1%) across time. The 0-time point represents the onset of *ziji.*

3.5 Discussion

Following the prediction by Dillon et al. (2014), we should observe, at least for the earliest window (0-400ms after the onset of the reflexive *ziji*), a significant increase for the LAP score, indicating that participants were searching within the local domain first (more fixation on the picture of ANP_2 in our case). What's more, for the inanimate NP, we should see more fixations on INP in the condition A (where INP is in the local clause) than other two conditions.

However, neither pattern was observed in our experiment. Instead, participants' search patterns showed consideration of both animate NPs initially regardless of condition. The significant

decrease in LAP during the early windows suggested that, when retrieving antecedents of *ziji* from working memory, participants tended to compare all the potential antecedents in parallel at first, updating their beliefs continuously before making their final choice, a search pattern at odds with the Local Search Hypothesis.

One possibility for the local bias observed in Dillon et al. (2014) may be a result of the "recency effect" of the subject NP in the local clause, which reflects the fact that the local subject NP may be less decayed in working memory. However, we cannot conclude, based on our experiments, that people are searching within the local clause first when interpreting *ziji*.

The Local Search Hypothesis fails to explain the patterns observed in both experiments: participants do use the animacy cue to guide their antecedent search and consider all the potential antecedents in parallel initially during the resolution of *ziji*.

4 Future directions

In this study, we only explore the effect that syntactic structure plays in the long-distance resolution of the reflexive *ziji*. It is still an open question what other possible factors that influence interpretation of *ziji* are and how these different constraints interact during the resolution. We plan to explore this question in the future using a different experimental manipulation. One phenomenon we found that we have not seen discussed in the literature before is a gender effect in *ziji* processing. It is well known that *ziji* is a gender-neutral reflexive (e.g., Tang, 1989). In (6), both Tom and Mary can be the antecedent of *ziji*:

(6) 汤姆 说 玛丽 低估-了 自己
Tom$_i$ shou Mary$_j$ digu-le *ziji*$_{i/j}$
Tom$_i$ say Mary$_j$ underestimate-PRF *ziji*$_{i/j}$
"Tom said that Mary underestimated him/herself"

In our eye-tracking experiment, the majority of the critical items contained two same gender images for the two animate NPs while the rest contained two differ gender images for the two animate NPs. Interestingly, when we divided the items into two groups, the same-gender group (two animate NPs were displayed using images with the same gender images) vs. the different-gender group (two animate NPs were displayed using images with different genders), we found a difference in participants' antecedent choices. As can be seen from Figure 5 below, for the same-gender group (top-right), the difference between the choices of the two animate NPs was similar to the pattern of the all-group (left, all 21 items). However, for the different-gender group (bottom-right), the choices between the two animate NPs are almost the same for the condition A and C, which contradicts the overall pattern (all-group) of ANP_2 (red) >> ANP_1 (blue).

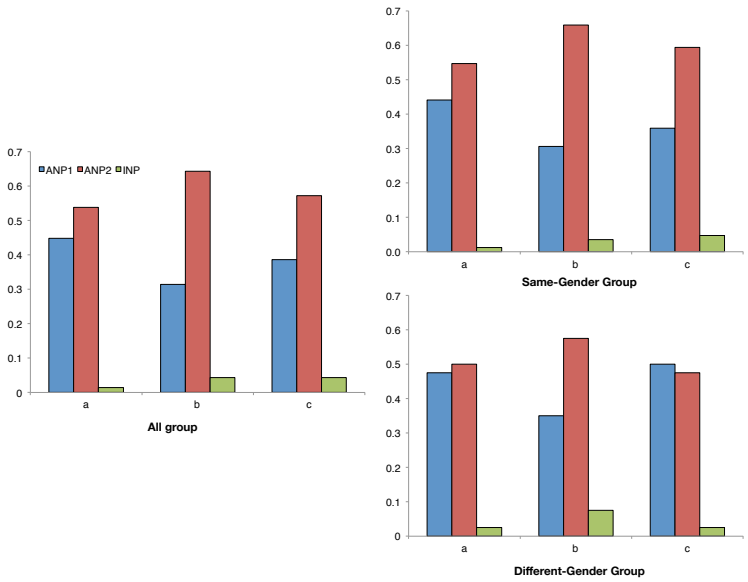


Figure 5. Antecedent choices by different groups (Left: all group; Top-right: the same-gender group; Bottom-right: the different-gender group).

We consider two possibilities for understanding these patterns. One possibility is that when the listener encodes the two NPs into memory, their gender plays a role in the encoding. In particular, two NPs of the same gender may cause similarity-based interference during the encoding stage. The other possibility is that gender is being used to guide the search during the retrieval stage (Jäger, Engelmann and Vasishth, 2017). The first possibility might predict a measurable effect immediately as the two NPs are encountered and being processed. The second possibility might predict an effect after encountering *ziji*. To test these two possibilities, we chose two critical time windows to analyze. The first window is from 50ms prior to the onset of the aspect marker *zai* to the onset of *ziji* (see Figure 6). From this window, we could test if there was an *encoding difference* for the two animate NPs in the different groups, because at this point, participants had already encountered all the NPs in the sentence before hearing *ziji.* The second window is from 50ms prior to the onset of *ziji* to 500ms after the onset of *ziji* (see Figure 7). From the second window, we could test if there was a *retrieval difference* for the two animate NPs in different groups, because during this time window, participants were searching for antecedents.

From the two plots (Figure 6 & 7), we can see clearly that the gender difference is due to a *retrieval difference* effect, not to an *encoding difference* effect: the pattern for the three groups is similar in the first window, but not in the second time window. The pattern in the second time window (Figure 7) showed that, in the Different-gender group, participants were more likely to consider the long-distance animate NP_1 as the antecedent of *ziji* than other groups.

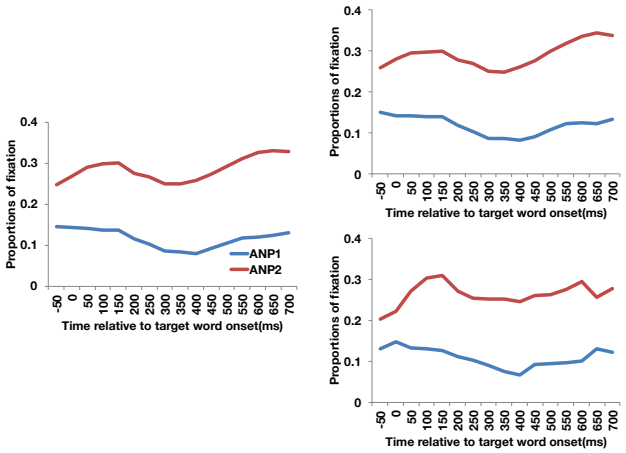


Figure 6. Proportion of fixations to the two animate NPs during the critical time window (50 ms prior to the onset of *zai* to the onset of *ziji*). Left: all group; Top-right: the same-gender group; Bottom-right: the different-gender group.

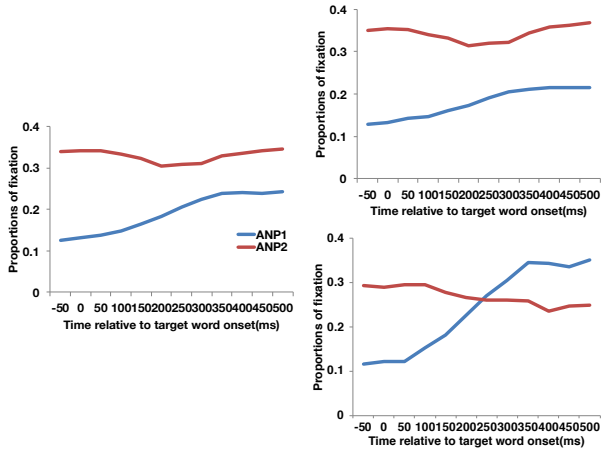


Figure 7. Proportion of fixations to the two animate NPs during the critical time window (50 ms prior to the onset of *ziji* to 500ms after the onset of *ziji*). Left: all group; Top-right: the same-gender group; Bottom-right: the different-gender group).

Currently, we do not have a clear explanation for the *retrieval difference* in the different-gender items. One possibility is that when processing *ziji*, people do consider gender as a cue in the antecedent retrieval even for *ziji*. Another possibility is that the difference is caused by the visually based item differences in our eye-tracking experiment. No matter which explanation is correct, this pattern reflects the interplay between the visual context, linguistic processing and memory retrieval during anaphor resolution. In future research we plan to manipulate the gender of the potential antecedents directly to better understand the role gender is playing in *ziji* resolution.

5 Conclusion

In this study, using the Chinese long-distance reflexive *ziji* as a test bed, we challenged the Local Search Hypothesis claim that the search for an antecedent for *ziji* begins in the local clause, only extending to higher clauses after exhausting the local context. Combining results from both off-line judgment tasks and visual world eye-tracking, we demonstrate that memory retrieval itself is probabilistic in nature during reflexive processing and that people do consider all potential antecedents in parallel initially during resolution.

6 Appendix: Norming task for V3

In order to ensure that both local and distance binding is possible for our experimental stimuli, we performed a separate norming task before Experiment 1 & 2. 31 native Chinese speakers participated in this task through a web-based questionnaire.

Following the method Li & Zhou (2010) and Schumacher et al. (2011) used in their norming tasks, we created 45 sentence completion tasks in the form of "Name (e.g., "*Lisi*") +*zai* (aspect marker) + Verb_____". The choices were (A) *ziji* (reflexive), (B) Name (e.g., "*Zhangsan*") and (C) Both. If people chose (C) *Both*, then they were asked to judge from a 7-point scale to show

which end ("*ziji*" or "*Zhangsan*") they thought to be more reasonable for this sentence. For example:

李四	在	低估_____。
Lisi	zai	digu
Lisi	PROG	underestimate

"Lisi is underestimating___."

A. 自己/*ziji* B.张三/*Zhangsan*(name) C.都可以 (Both)

*如果您选择 C, 你觉得哪一个选项更合适呢，请在下面的量表里做出判断：

*(If you choose C, please indicate which end you think is better here on the following scale)

自己/*ziji*	。	。	。	。	。	张三/*Zhangsan*
1	2	3	4	5	6	7

For the finally chosen 21 "neutral verbs" (see below), at least 25 participants (out of 31) chose to be ambiguous (choice of *C*) with the average score of 4.21± 1.06.

低估 digu	*underestimate*	谈论 tanlun	*discuss*
耽误 danwu	*jeopardize*	维护 weihu	*defend*
责怪 zebei	*blame*	推荐 tuijian	*recommend*
高估 gaogu	*overestimate*	保护 tuijian	*protect*
激励 jili	*encourage*	折腾 zheteng	*frustrate*
夸奖 kuajiang	*praise*	介绍 jieshao	*introduce*
肯定 kending	*affirm*	拯救 zhengjiu	*heal*
伤害 shanghai	*harm*	挑战 tiaochan	*challenge*
吓唬 xiahu	*frighten*	训练 xunlian	*train*
否定 fouding	*deny*	批评 piping	*criticize*
欣赏 xinshang	*appreciate*		

References

Badecker, William and Kathleen Straub. 2002. The processing role of structural constraints on interpretation of pronouns and anaphors. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 28:748.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. 2013. Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* 68:255-278.

Chen, Qingrong, Zhu Deng, and Bo Jiang. 2008. The effects of working memory and syntactic complexity on Chinese sentence processing. *Psychological Science 心理科学*3:633–637.

Chen, Zhong, Lena Jäger, and Shravan Vasishth. 2012. How structure-sensitive is the parser? Evidence from Mandarin Chinese. *Empirical approaches to linguistic theory: Studies of meaning and structure*. 43-62.

Dillon, Brian, Alan Mishler, Shayne Sloggett, and Colin Phillips. 2013. Contrasting intrusion profiles for agreement and anaphora: Experimental and modeling evidence. *Journal of Memory and Language* 69:85-103.

Dillon, Brian, Wing-Yee Chow, Matthew Wagers, Taomei Guo, Fengqin Liu, and Colin Phillips. 2014. The structure-sensitivity of memory access: evidence from Mandarin Chinese. *Frontiers in Psychology* 5:1025.

Dink, Jacob, and Brock Ferguson. 2018. eyetrackingR: An R Library for Eye-tracking Data Analysis: Retrieved from http://www.eyetrackingr.com.

He, Xiao. 2014. *What 'you' and 'I' Can Say about Reference Resolution and Non-structural Constraints*. PhD Dissertation, University of Southern California.

Hsieh, I-Ta Chris. 2015. Long-distance reflexives, blocking effects, and the structure of Mandarin comparatives. *Syntax* 18:78-102.

Huang, C. T. James, Audrey Y.H. Li, and Yafei Li. 2009. *The Syntax of Chinese*: Cambridge Syntax Guides. Cambridge: Cambridge University Press.

Jäger, Lena A., Felix Engelmann, and Shravan Vasishth. 2015. Retrieval interference in reflexive processing: experimental evidence from Mandarin, and computational modeling. *Frontiers in Psychology* 6:617.

Jäger, Lena A., Felix Engelmann, and Shravan Vasishth. 2017. Similarity-based interference in sentence comprehension: Literature review and Bayesian meta-analysis. *Journal of Memory and Language* 94:316-339.

Kaiser, Elsi, Jeffrey T. Runner, Rachel S. Sussman, and Michael K. Tanenhaus. 2009. Structural and semantic constraints on the resolution of pronouns and reflexives. *Cognition* 112:55-80.

Kuznetsova, Alexandra, Per B. Brockhoff, and Rune Haubo Bojesen Christensen. 2017. lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* 82:1-26.

Li, Xiaoqian and Xiaolin Zhou. 2010. Who is ziji? ERP responses to the Chinese reflexive pronoun during sentence comprehension. *Brain Research* 1331:96-104.

McElree, Brian. 2000. Sentence comprehension is mediated by content-addressable memory structures. *Journal of Psycholinguistic Research* 29:111-123.

Parker, Dan, Michael Shvartsman, and Julie A. Van Dyke. 2017. The cue-based retrieval theory of sentence comprehension: New findings and new challenges. *Language Processing and Disorders*:121-144.

R Core Development Team. 2014. R: A language and environment for statistical computing: R foundation for statistical computing. Vienna, Austria.

Runner, Jeffrey T, Rachel S. Sussman, and Michael K. Tanenhaus, Michael K. 2003. Assignment of reference to reflexives and pronouns in picture noun phrases: evidence from eye movements. *Cognition* 89:B1-B13.

Salverda, Anne Pier and Michael K. Tanenhaus. 2017. The visual world paradigm. *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide* 9:89.

Schumacher, Petra B., Walter Bisang, and Linlin Sun. 2011. Perspective in the processing of the Chinese reflexive ziji: ERP evidence. In *Discourse Anaphora and Anaphor Resolution Colloquium*, 119-131: Springer.

Sekerina, Irina A., Luca Campanelli, and Julie A. Van Dyke. 2016. Using the visual world paradigm to study retrieval interference in spoken language comprehension. *Frontiers in Psychology* 7:873.

Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268:1632-1634.

Tang, Chih-Chen Jane. 1989. Chinese reflexives. *Natural Language & Linguistic Theory* 7:93-121.