THE LONG-DISTANCE 'TAZIJI' REVISITED: EXPERIMENTAL EVIDENCE FOR INHERENT NON-LOCAL BINDINGS*

YUHANG XU
JEFFREY T. RUNNER
University of Rochester

1 Introduction

Chinese reflexives provide a unique typological test bed for evaluating theories of anaphora in theoretical linguistics and psycholinguistics (see recent reviews, e.g., Jäger, Engelmann, and Vasishth 2015; Charnavel et al. 2017). It is well documented that there are at least two types of reflexives in Mandarin Chinese (e.g., Y.-H. Huang 1984; L. Xu 1993; Pan 1998; Yu 2000; C.-T. J. Huang, Li, and Li 2009; Hu 2019), the compound/polymorphemic reflexive *taziji* (pronoun + "self", similar to *himself* or *herself* in English), and the bare/monomorphemic reflexive *ziji* ("self"). The pronoun *ta* and the compound reflexive *taziji* are often considered to be similar to their English counterparts *him/her* and *himself/herself* which strictly follow the binding principles (e.g., Chomsky 1981) while the bare reflexive *ziji* is thought to belong to the long-distance reflexive family which are "exempt" from Principle A and are subject to another set of constraints (e.g., Tang 1989; Pan 1998).

We present in (1) the original examples[1] and reported judgments from a widely used textbook, *The syntax of Chinese* (C.-T. J. Huang, Li, and Li 2009), to illustrate the syntactic behavior of Chinese anaphors as generally understood by the field. According to C.-T. J. Huang, Li, and Li (2009), the pronoun *ta* as in (1a) can only take, abiding by binding principles, the non-local subject *Zhangsan* as its antecedent . Whereas the reflexive *taziji* as in (1b), must only take the local subject *Lisi* as its antecedent. On the other hand, for the bare reflexive *ziji*, both local and long-distance binding options are available as shown in (1c).

*We would like to thank Prof. Ash Asudeh for his insightful comments and suggestions that improved this paper.

[1]The examples are presented in Latin transcriptions in the textbook. The Chinese scripts are based on a translation version of the same textbook (C.-T. J. Huang, Li, and Li 2013).

(1) C.-T.J.Huang et al. (2009, pp. 330,333).

a. 张三 知道 李四 老 批评 他
Zhangsan$_i$ zhidao Lisi$_j$ lao piping ta$_{i/*j}$
Zhangsan$_i$ know Lisi$_j$ incessantly criticize *ta*$_{i/*j}$
'Zhangsan knows that Lisi criticizes him all the time.'

b. 张三 知道 李四 老 批评 他自己
Zhangsan$_i$ zhidao Lisi$_j$ lao piping taziji$_{*i/j}$
Zhangsan$_i$ know Lisi$_j$ incessantly criticize *taziji*$_{*i/j}$
'Zhangsan knows that Lisi criticizes himself all the time.'

c. 张三 知道 李四 老 批评 自己
Zhangsan$_i$ zhidao Lisi$_j$ lao piping ziji$_{i/j}$
Zhangsan$_i$ know Lisi$_j$ incessantly criticize *ziji*$_{i/j}$
'Zhangsan knows that Lisi criticizes self all the time.'

Also similar to English, Mandarin Chinese has different forms for the masculine and feminine third person singular pronouns and reflexives. However, they are homophones that are distinguishable only in their orthography[2]. For example, besides the gender neutral reflexive *ziji* 自己, the masculine and feminine pronoun *ta* are homophones written as 他 and 她 respectively and the homophonic masculine and feminine reflexive *taziji* are written as 他自己 and 她自己 respectively[3]. This unique feature, although discussed in the literature, is often ignored by researchers because the Chinese examples in the linguistic literature are often transcribed using the Latin alphabet instead of the original script (Chinese characters). For example, Pan (1998) only mentions this orthographic distinction in a footnote without discussing any role this feature might play in the interpretations of Chinese anaphora. Some researchers have even proposed that Chinese speakers are not sensitive to this gender information at all and use the masculine form (i.e., pronoun 他 or reflexive 他自己) as the default form to refer both male and female referents (e.g., Su et al. 2016).

These two types of Chinese reflexives (*ziji* and *taziji*) have drawn much attention in the field as a critical test bed for evaluating both theoretical and psycholinguistic models of anaphora. For successful theoretical anaphor models, they should account for the existence of both types of reflexives in the same/different language(s) and their different behaviors as well: e.g., Reinhart and Reuland (1993) proposed two types of anaphoric elements cross-linguistically with SELF-anaphors as locally bound and SE-anaphors as non-local bound. Another similar approach is to argue that only the locally-bound reflexive is a true anaphor and the non-local ones are logophors which are subject to a different set of constraints (e.g., C.-T. J. Huang and Liu 2001). For the psycholinguistics field, these two types of reflexives have been widely used, beyond English (e.g., see a review in Jäger, Engelmann, and Vasishth 2015), to adjudicate competing claims (e.g., structure first vs. multiple constraint approaches), to evaluate predominant processing models (e.g.,

[2]We only focus on Mandarin Chinese in this study. For many variants/dialects of Chinese, they only have one form for pronouns: e.g., Cantonese only has one pronoun *keoi* and Hokkien only has one pronoun 伊*yi* for both masculine and feminine forms.

[3]We consider it a pure orthographic (rather than morphological) feature because the homophonic characters 他 and 她 only differ in their semantic radicals (亻vs.女 and both share the same phonetic radical 也) and these radicals are not morphemes (see a comprehensive review of the Chinese writing system in Shu 2003).

cue-based retrieval models), and to investigate how both structural (e.g., locality and c-command) and non-structural constraints (e.g., gender or animacy cues) affect anpahor resolutions (e.g., Chen, Jäger, and Vasishth 2012; Dillon et al. 2014; Dillon, Chow, and Xiang 2016; Chang et al. 2020). However, despite the research on these forms, some basic behaviors of these reflexives still remain controversial. In this study, we focus on one of them which is rarely discussed in the literature: non-local bindings for *taziji*.

1.1 Long-distance binding options for *taziji*

Although most of the literature, e.g., C.-T. J. Huang, Li, and Li (2009), assumes that Chinese *taziji* is similar to English himself/herself in strictly requiring a local antecedent, there are still some exceptional cases reported in the literature. Most of these cases can be categorized as due to the influence of other (non-syntactic) factors. For example, Pan and colleague argued that locality is not a strict condition and both prominence (e.g., *animacy hierarchy*) and locality regulate the interpretation of *taziji* (Pan 1998; Pan and Hu 2003).

Pan (1998) claimed that in (2), where the local subject is inanimate, *taziji* can only take *Zhangsan* but not the local subject *book* as the antecedent following a "Prominence Constraint" that the animate subject is more "prominent" than the inanimate one in the animacy hierarchy.

(2) Pan (1998): Example 4.

张三	说	那	本	书	害了	他自己
Zhangsan$_{i}$	shuo	na	ben	shu$_{j}$	hai-le	taziji$_{i/*j}$
Zhangsan$_{i}$	say	that	CL	book$_{j}$	hurt-Perf	himself$_{i/*j}$

'Zhangsan said that that book hurt himself.'

Yu (1992) also provided some data challenging the strict local binding requirement of *taziji*. For instance, if the local antecedent mismatches the person feature of *taziji*, as in (3), only the non-local subject John but not *wo* "I" can be the antecedent.

(3) Yu (1992): Example 9a.

约翰	要	我	坐	在	他自己	的	身边.
Yuehan$_{i}$	yao	wo$_{j}$	zuo	zai	taziji$_{i/*j}$	de	shenbian.
John$_{i}$	want	I$_{j}$	sit	at	himself$_{i/*j}$	's	side.

'John wants me to sit at his side.'

Similarly, he argued for a "gender match" requirement for *taziji* as well. In the following examples, *Zhangsan* and *Lisi* are typical male names and *Weiling* is a typical female name in Chinese. According to Yu (1992), although marked as ? in (4), both *Zhangsan* and *Lisi* can be the antecedent of masculine *taziji* (他自己) while in (5) only the non-local *Zhangsan* can be the antecedent because *Zhangsan* is a typical male name but *Weiling* is not. This is what Pan and Hu (2003) categorized as a "Feature Compatibility Constraint" which requires that two coindexed elements must have compatible features (e.g., person or gender feature).

(4) Yu (1992): Example 9e.

?张三 叫 李四 吹捧 他自己
Zhangsan$_i$ jiao Lisi$_j$ chuipeng taziji$_{i/j}$
Zhangsan$_i$ ask Lisi$_j$ promote himself$_{i/j}$
'?Zhangsan asked Lisi to promote him/himself.'

(5) Yu (1992): Example 10.

张三 叫 魏玲 吹捧 他自己.
Zhangsan$_i$ jiao Weiling$_j$ chuipeng taziji$_{i/*j}$.
Zhangsan$_i$ ask Weiling$_j$ promote himself$_{i/*j}$.
'Zhangsan asked Weiling to promote him.'

Another line of unexpected cases comes from experimental studies on Chinese reflexives processing. Although most of them focus on the bare reflexive *ziji* (e.g., Dillon et al. 2014; Chen, Jäger, and Vasishth 2012; also see a comprehensive review in Jäger, Engelmann, and Vasishth 2015), there are a few studies looking at "long-distance" options for *taziji* as well. For instance, Dillon, Chow, and Xiang (2016) found a less-than-expected locality effect for *taziji* with a subcommanding antecedent (i.e., subject antecedents properly contained within a c-commander of *taziji*). In an eye-tracking-during reading-experiment, Chang et al. (2020) found no (gender-matching) interference effect of a long-distance antecedent but the off-line judgment questionnaire did show a significant amount of selections of the long-distance antecedent (~20%). The authors interpreted this as an "error" due to the experimental design and processing pressures. This long-distance binding was also found using other off-line tasks as well, e.g., Lyu and Kaiser (2021). However, these studies did not directly focusing on *taziji* (i.e., using it as a baseline comparison to *ziji*) and only provide indirect evidence for long-distance *taziji* or even treated them as a non-linguistic effect (e.g., processing errors).

Taken together, these cases of the long-distance binding option show that the antecedent of *taziji*, similar to English reflexives, should agree with the animacy, person and gender features of the reflexive, though the gender feature, for instance, is marked through the use of different characters (homophone *taziji*: 他自己*himself* vs. 她自己*herself*). *Taziji* can take a non-local antecedent when no matching one available locally. Also, as proposed by those scholars, there is also a possibility that the long-distance binding option is merely a by-product of other constraints or that the long-distance binding option is only allowed in some special syntactic environments. For example, in (2) the local subject is inanimate, so the long distance option is the only possibility given the animacy requirement[4] of *taziji* (他自己). Examples (3)–(5) all involve "control" structures, and although the binding theory also predicts a local binding for *taziji* in such structures, the fact that these are control structures complicates the interpretation. If both local and long-distance antecedents are exactly the same except for their relative positions in a sentence similar to (1), are both local and long-distance options still available for the reflexive *taziji*? If so, then what is the alternative explanation for those long-distance cases?

[4]Here we consider animacy as a semantic feature. It is also possible to treat animacy as a syntactic feature: e.g., Asudeh(1998) used animacy as a syntactic requirement in his HPSG account of *ziji*.

1.2 Proposal: long-distance binding option for *taziji* is inherently available

We propose that the long-distance binding option reflects the non-obligatory locality requirement for the reflexives *taziji*: the long-distance binding option of *taziji* is inherently available. Here we operationalize the concept of the "inherent long-distance" as: a non-local binding option is still available even when a local legitimate antecedent exists. For instance, in the example (1) above, we would show that *taziji* can also take the matrix subject *Zhangsan* as its antecedent even though the local subject *Lisi* is also a legitimate one. Although this proposal is in apparent conflict with some previous accounts of *taziji*, we will demonstrate in our experiments, that the strict locality requirement claimed in the literature is too strong, i.e., that "locality" is a preference rather than a requirement for *taziji*.

A natural empirical test for this proposal is also straightforward: in a structure with both local and non-local antecedents, we will test if speakers have a long-distance interpretation of *taziji*, when a local interpretation is also available. To do so, we constructed a bi-clausal structure "*Name1-says-Name2-Verb-taziji*" in which two potential antecedents only differed in their relative positions (local vs. long-distance) in the sentence while other structural aspects were controlled, i.e., they are all proper names in subject positions. As discussed previously, the antecedent needs to match the gender of *taziji* as well. We manipulated this gender agreement feature using gender stereotyped names (for antecedents) and different gender *taziji* (through different orthographies). By manipulating both syntactic and gender factors, we were able to explore whether the long-distance binding option for *taziji* is available above and beyond the effect of other constraints (i.e., gender-matching constraints in this case).

The rest of the paper is organized as follows: we first demonstrate the validity of our method in Experiment 1 on the pronoun *ta*, showing that our method is able to reveal a robust pattern in which a clear structural constraint exists. Next in Experiment 2, we test our proposal of the inherent long-distance (LD) option for *taziji* that both local and non-local bindings are available at the same time. In Experiment 3 and 4, we refute a hypothesis of the logophoric interpretation of the long-distance *taziji*. Last, we also discuss the implications of this study.

2 Experiment 1 & 2: inherent LD options for *taziji*

2.1 Overview

Since Experiment 1 and 2 share a similar design, items and task, we will first describe the general method, then present results separately for each experiment. Adopting the same sentence structure across both experiments (viz. *Name1-V1-Name2-V2-anaphor*), Experiment 1 serves as a control study and sanity check to demonstrate the validity of our method. Then in Experiment 2, we tested the hypothesis of inherent long-distance binding options for the reflexive *taziji*. The materials, data and code for this study are available in the OSF project repository (https://shorturl.at/ikQV8).

2.2 Methods

2.2.1 Participants

84 native Chinese speakers (28 women; 28.64 +/- 5.99 years) from Mainland China were recruited through the online platform "Witmart" which is the largest China-based crowd-sourcing pool. This

platform has been used widely for social science research (see To and Lai 2015 for a review) and particularly linguistics-related studies on the Chinese population (e.g., Zhan, Levy, and Kehler 2020). Participants were randomly assigned to one of the experiments (42 participants for each experiment) and paid about 2 US dollars for participation.

2.2.2 Design and materials

Bi-clausal sentences of the structure "*Name1-V1-Name2-V2-pronoun/reflexive*" were designed for the critical items. For clarity and consistency purposes, two potential antecedents (i.e., Name 1 and 2) were labeled as the "Target" or "Competitor" with respect to the predictions of Binding principles: if the position of the antecedent was consistent with what Binding principles indicated, we then labeled it as the "Target", otherwise we labeled it as the "Competitor". For example, in the reflexive conditions, the local subject "Name2" was labeled as the "Target" and the subject of the matrix clause Name1 was labeled as "Competitor". This enabled us to evaluate the relevant strength of the structural constraints (i.e., binding principles) within and between pronouns and reflexives. The "V1" was one of a set of verbs like "say" that select for a sentential complement while the "V2" was taken from a set of transitive verbs that were normed in a previous study (Y. Xu and Runner 2019) to ensure that both local and long-distance binding options were plausible. We provide a detailed description of these norming processes in the section 2.2.3 .

For both experiments, three factors were manipulated: the first two manipulations were the gender type of the *Target* as well as the *Competitor* (either matched or mismatched the gender of the pro-form) using gender stereotyped names taken from a previous study (Qiu et al. 2012). By manipulating the "gender" feature for both antecedents, we were able to test the separate effects of two factors (locality vs. gender matching) as well as their relative weights. Crucially, it enabled us to evaluate whether or not speakers chose the long-distance antecedents even when a local legitimate antecedent was available. The third manipulation was the gender of the pro-form (masculine or feminine form using different characters: 他vs. 她; 他自己vs. 她自己). Although this manipulation was not relevant to our main research question[5], from a design perspective, the full factorial design has the advantage of achieving a balance of the stereotyped names in Name 1 and 2 positions.

A set of sample materials is shown in Table 1. A total of 16 sets of critical items and 40 filler sentences were created. We also included one of C.-T. J. Huang, Li, and Li (2009)'s original examples (as shown in 1a & b) at the beginning of each list as a direct replication to see if there is any effect of different lexical items used in our experimental stimuli. Recent studies demonstrated the importance of direct replications of original examples in evaluating the reliability of linguistic judgments (e.g., Chen, Xu, and Xie 2020). In total, each participant read 57 sentences for the experiment.

2.2.3 Norming of the materials

For gender stereotyped names (Name 1 & 2), we selected 32 male and 32 female Chinese names based on the pre-normed results from Qiu et al. (2012). Those names had been rated on a 5-point scale with 1 being extreme feminine and 5 being extreme masculine. The average gender

[5]However, as we mentioned in the introduction, a few researchers have claimed the masculine form could be used as the default form to refer to both male and female referents (e.g., Su et al. 2016). This manipulation gives us a chance to test this claim (which is not true according to our results).

stereotyped rating was 4.69 +/- 0.10 for male names and 1.16 +/- 0.07 for female names. The ratings differences for male and female names were significant ($t(62) = -165.15, p < .0001$). For the same gender name pairs, the ratings of Name 1 positions were not different from those of Name 2 positions (male groups, $t(30) = -0.001, p = .99$; female groups, $t(30) = -0.04, p = .97$).

For verbs in the matrix clause (V1), Y. Xu and Runner (2019) selected 9 verbs such as *renwei* "think" or *biaoshi* "say" which can naturally select a sentential complement. Those verbs were taken from previous experimental studies on Chinese anaphor resolution by Chen, Jäger, and Vasishth (2012) and Dillon et al. (2014) .

For verbs in the embedded clause (V2), they were taken from a previous studies by Y. Xu and Runner (2019). In order to ensure that both local and long distance binding interpretations were plausible, Y. Xu and Runner (2019) performed a separate norming task where 31 additional native Chinese speakers participated through a web-based questionnaire. Following the norming method introduced by Li and Zhou (2010) and Schumacher, Bisang, and Sun (2011), they created 45 sentence completion tasks in the form of "Name (e.g., *Zhangsan*) + *zai* (aspect marker) + Verb __". The choices were *A. ziji* (reflexive), *B. a proper name* (e.g., *Zhangsan*) and *C. Both*. If people chose the "*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 plausible for the current sentence. For the finally chosen 16 "neutral" verbs, at least 26 participants (out of 31) chose to be ambiguous (i.e., choice of the *C. Both* with an average score of 4.21 +/- 1.06 on a 7-point scale.

Table 1: Sample materials and tasks used in Experiment 1 & 2.

Sample martials					
					Experiment 1: pronoun ta
孙志/张艳	表示	国俊/陈凤	在	低估	他/她
Sun Zhi/Zhang Yan	biaoshi	Guo Jun/Chen Feng	zai	digu	*ta/ta*
male/female name	say	male/female name	PROG	underestimate	him/her
'Name 1 says that Name 2 is underestimating him/her.'					
					Experiment 2: reflexive taziji
孙志/张艳	表示	国俊/陈凤	在	低估	他自己/她自己
Sun Zhi/Zhang Yan	biaoshi	Guo Jun/Chen Feng	zai	digu	*taziji/taziji*
male/female name	say	male/female name	PROG	underestimate	himself/herself
'Name 1 says that Name 2 is underestimating himself/herself.'					
Tasks					
**Antecedent choice task:*			**Acceptability judgment task:*		
句子里的他/他自己指的是谁?			你觉得这句话读起来如何?		
Who does *ta/taziji* refer to?			How does this sentence sound to you?		
(Name 1 / Name 2)			(7-point scale: very bad : very good)		

2.2.4 Procedure

The experiments were performed using the online questionnaire platform *Qualtrics* with each participant randomly assigned to one of the 8 experimental lists for each experiment. Critical

items were distributed across the lists following a Latin square design and were interspersed pseudo-randomly among filler sentences such that no two critical items were adjacent. One sentence was presented to the participants each time with two questions. For critical items, the first was an antecedent choice task (e.g., "who does *taziji* refer to?") with two names mentioned in the sentence as choices (*Target* and *Competitor*). The second was an acceptability judgment task ("How does this sentence sound to you?") to solicit participants' ratings of each sentence using a 7-point Likert scale (1: very bad, 7: very good). For filler trials, similar comprehension and acceptability judgment questions were used. Before test trials, a practice session was included to help participants understand the task. The original example from C.-T. J. Huang, Li, and Li (2009) was always presented at the beginning of each experimental list. It took approximately 15 to 20 minutes to complete the task.

2.2.5 Analytical strategy

For statistical analysis, logistic (for antecedent choice data) or linear (for acceptability judgment data) mixed effects models were used with fixed effects of the *Target* gender type (match: 0.5; mismatch: -0.5), the *Competitor* gender type (match: 0.5; mismatch: -0.5) and the pro-form gender type (masculine: 0.5; feminine: -0.5). Following Barr et al. (2013), we also include the maximal random effects structure for both subjects and items in all models. When the maximal model did not converge, the random component with the least variance was removed and the model was refit until it converged. The dependent variable was choice of the Target (1: chosen; 0: not chosen) for antecedent choice data and the z-score transformed judgment score for acceptability judgment data. The z-score transformation was applied to eliminate potential scale bias which is a common strategy in the field (Schütze and Sprouse 2013). We also modeled the raw acceptability rating data as recommended by some researchers (e.g., Juzek 2015) which yielded similar results as z-score models. For clarity, we only present the z-scores analyses and plots here. Readers can refer to the R scripts available in the OSF repository for details.

2.3 Experiment 1 results: pronoun *ta*

2.3.1 Results summary

Figure 1 present data from Experiment 1 for both antecedent choice and acceptability judgment tasks.For antecedent choice data, we plotted the choice proportions for both Target (dark bars) and Competitor (light bars) across Target gender conditions (match/mismatch: left/right panels) and Competitor gender conditions (match/mismatch: left/right side for each panel). For acceptability judgment data, we plotted ratings across both Target gender conditions (match/mismatch: solid/dash lines) and Competitor gender conditions (match/mismatch: left and right sides).

The antecedent choice data patterns showed a clear structural constraint such that participants had overwhelming more Target choices across all conditions and their judgment ratings for Target match conditions were higher than mismatched ones: we observed a main effect of the Target gender type such that people had more Target choices when the Target matched the gender of the pronoun than when it mismatched ($\hat{\beta} = 5.83, SE = 2.59, z = 2.25, p = .02$). Also, the ratings for Target gender matched ones were significantly higher than mismatched ones ($\hat{\beta} = 0.52, SE = 0.07, t = 7.06, p < .001$). No other effects were observed.

Next, we focused on the Target gender match conditions (Figure 1: Left panel of the antecedent choice plot) to determine whether or not participants also considered the BT-incompatible choice (i.e., Competitor) when there was an acceptable BT-compatible antecedent available (i.e., gender match Target). First, following the standard proposed by He and Kaiser (2016), we compared the Competitor choices against 0% when both Target and Competitor matched the pronoun gender: the choice of the Competitor (average 3.1%) was not significantly higher than 0% (using log-transformed individual proportion means: $\hat{\beta} = -0.05, SE = 0.04, t = 1.37, p = .18$). Second, we tested the Competitor choices between the Competitor match and mismatch conditions. Comparing to the baseline Competitor gender mismatch condition, participants did not choose more Competitors even when they matched the gender of the pronoun *ta* ($\hat{\beta} = -0.61, SE = 5.23, z = -0.12, p = .91$). Both analyses indicate that participants did not consider the Competitor as an acceptable antecedent for the pronoun *ta*.

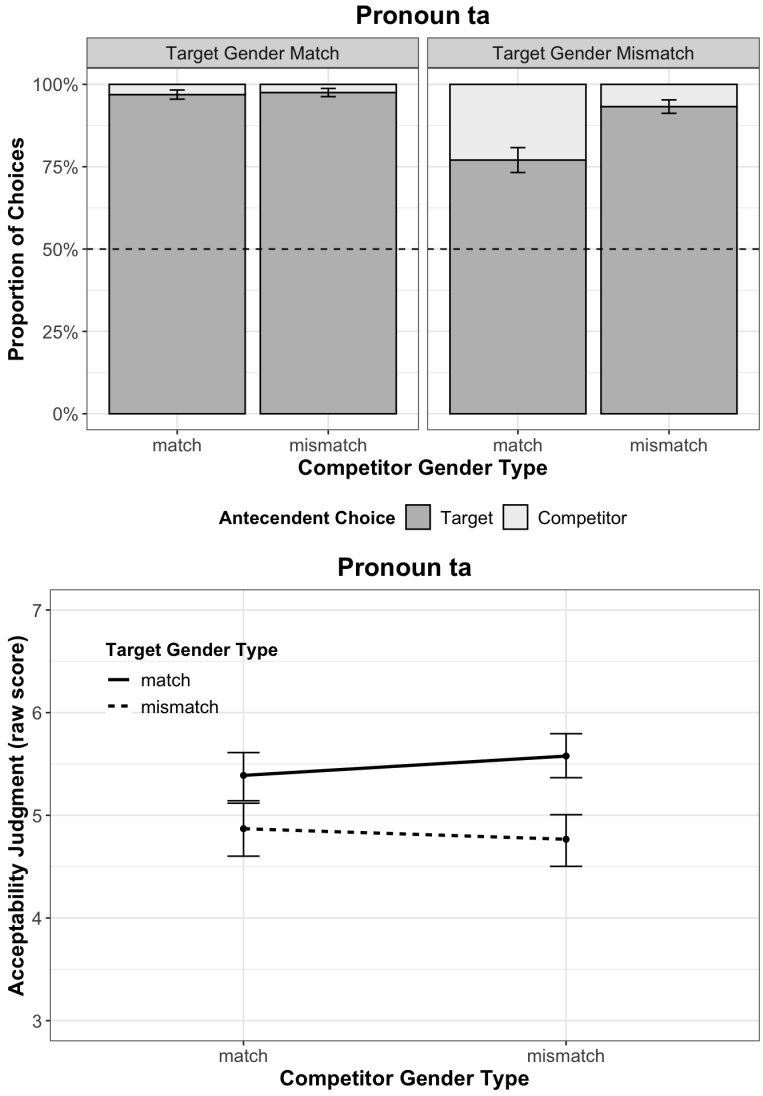

Figure 1: Antecedent choice and acceptability judgment (z-score transformed) results in Experiment 1. For choice results, dark bars represent the proportion of choosing the Target/Name1 as the antecedent of the pronoun *ta*, while the light bars represent the proportion of choosing the Competitor/Name2 as the antecedent. Error bars represent the standard error of the subject means. The judgment results figure shows the averaged z-score transformed judgment ratings across subjects and items in each condition. Solid lines represent conditions where the Target/Name1 matched the gender of the pronoun *ta* while dash lines are for the conditions where the Target mismatched the gender of the pronoun. Error bars represent bootstrapped 95% confidence intervals.

2.3.2 Predictions for Experiment 2/*taziji* results

Experiment 1 demonstrated the validity of our method. Participants were sensitive to both our gender manipulation as well as the structural constraints: while overwhelmingly they selected binding theory compatible antecedents across all conditions, they provided lower acceptability ratings for Target gender mismatch conditions. Thus, our methods are able to reveal a robust pattern when there is indeed a strong structural constraint: participants would only select BT-compatible antecedents as shown in their antecedent choice task and provided lower ratings when

the Target mismatched the reflexive gender. If the reflexive *taziji* also follows a strict syntactic constraint, we should observe a similar effect as that for the pronoun *ta* (i.e., overwhelmingly Target choices across all conditions). If instead *taziji* has an inherent long-distance binding option, we should see significant choices of the structural inaccessible but gender matching Competitor even when a gender match Target available. Next, we present results from Experiment 2 showing the long-distance binding option for the reflexive *taziji*.

2.4 Experiment 2 results: reflexive *taziji*

2.4.1 Results summary

Similar to Experiment 1, Figure 2 presents data from Experiment 2 for both antecedent choice and acceptability judgment tasks. For reflexive conditions, we observed a main effect of the target gender type ($\hat{\beta} = 2.49, SE = 0.71, z = 3.53, p < .001$) and of the competitor gender type ($\hat{\beta} = -2.36, SE = 0.71, z = -3.31, p < .001$) in people's antecedent choices: people had more Target choices when the Target matched and also the Competitor mismatched the gender[6]. For the acceptability judgment task, we observed a main effect of the Target gender type ($\hat{\beta} = 0.23, SE = 0.07, t = 3.03, p < .001$) as well as a trending interaction of the Target gender type and Competitor gender type ($\hat{\beta} = -0.28, SE = 0.15, t = -1.89, p = .06$). We further conducted a planned pairwise comparison showing that the ratings for the Target matched vs. mismatched conditions did not differ under the Competitor gender matched conditions ($\hat{\beta} = 0.17, SE = 0.20, t = 0.83, p = .41$) and only differed under Competitor gender mismatched conditions ($\hat{\beta} = 0.65, SE = 0.20, t = 3.22, p < .001$): people rated sentences with only the Competitor gender match as perfectly legitimate showing no significant difference form the ones in which the Target gender also matched ones.

Similar to Experiment 1, we also compared the Competitor choices under the Target gender match conditions (i.e., Figure 2: Left panel of the antecedent choice data) to test if participants also considered the binding theory incompatible but gender matched Competitor as the antecedent of the reflexive *taziji*. First, we compared the Competitor choices when both Target and Competitor matched the gender: the choice of the Competitor (average 19.4%) was significantly higher than 0% (using log-transformed individual proportion means: $\hat{\beta} = 0.09, SE = 0.04, t = 2.36, p < .03$). Second, we found people had significantly more Competitor choices when the Competitor also matched the condition ($\hat{\beta} = 2.79, SE = 1.28, z = 2.17, p < .03$) comparing to the Target-only matched conditions. Both results indicate that people consider both the Target and Competitor as the potential antecedents of the reflexive *taziji*.

[6]We also observed a main effect of the reflexive gender such that people had more Target choices under masculine conditions (他自己*himself*) than feminine conditions (她自己*herself*). One possibility is that people treated the masculine form as the default and used it to refer to a feminine antecedent as well. Another possibility is that the masculine form is much frequent (almost six times more) than the feminine one (estimation based on a Google Ngram search from 1988 to 2008) and the uncertainty was reduced (i.e., fewer Competitor choices) when a more frequent form was presented. The interpretation of this effect won't change the main conclusions of this study and we leave this as an open question and expect future work to address this phenomenon.

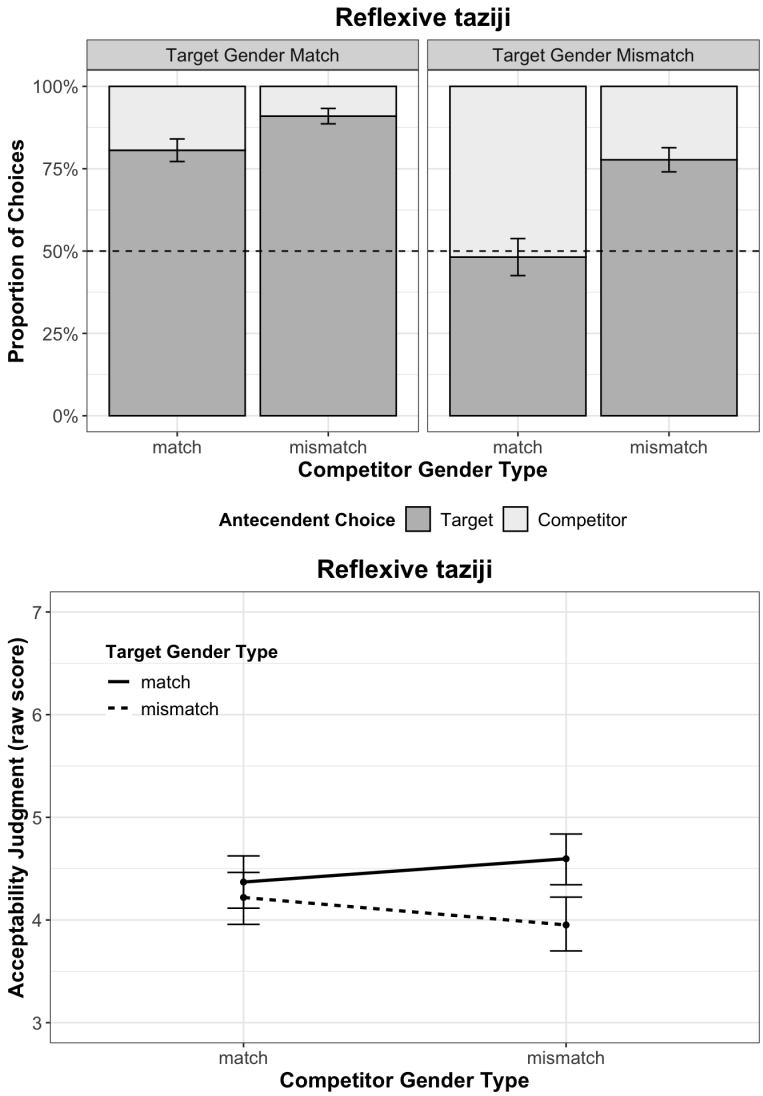

Figure 2: Antecedent choice and acceptability judgment (z-score transformed) results in Experiment 2. For choice results, dark bars represent the proportion of choosing the Target/Name2 as the antecedent of the reflexive *taziji*, while the light bars represent the proportion of choosing the Competitor/Name1 as the antecedent. Error bars represent the standard error of the subject means. The judgment results figure shows the averaged z-score transformed judgment ratings across subjects and items in each condition. Solid lines represent conditions where the Target/Name2 matched the gender of the reflexive *taziji* while dash lines are for the conditions where the Target mismatched the gender of the reflexive. Error bars represent bootstrapped 95% confidence intervals.

2.4.2 *Ta* vs. *taziji* analysis and textbook original examples

We conducted a "cross-experiment" analysis using similar mixed-effects models of the antecedent choices, as well as sentences acceptability ratings, comparing the pronoun *ta* and the reflexive *taziji* (treatment coded, *ta*: 0; *taziji*: 1) collapsed across all conditions. Results showed that participants had more Target choices (i.e., BT-compatible choices) for pronoun conditions than reflexive conditions ($\hat{\beta} = -1.57, SE = 0.33, z = -4.75, p < .0001$) and the acceptability ratings for pronoun conditions did not differ from reflexive conditions ($\hat{\beta} = 0.01, SE = 0.05, t = 0.01, p > .9$).

Compared to the pronoun *ta*, the reflexive *taziji* was less sensitive to the structural constraints such that both local and non-local subjects were considered as valid antecedents.

We also tested the original examples of the pronoun *ta* and the reflexive *taziji* (i.e., 1a and 1b) used in C.-T. J. Huang, Li, and Li (2009) textbook. Because they used two same-gender names (i.e., *Zhangsan* or *Lisi*) without any gender manipulation in their original examples, the corresponding conditions in our experiment should be Target gender-matched with Competitor gender-matched ones. Participants had 95.2% Target and 4.8% Competitor choices for the *ta* example and 69.8% Target and 30.2% Competitor choices for the *taziji* example. The Target choices for the *ta* example was significantly higher than those for the *taziji* example ($\hat{\beta} = -2.16, SE = 0.80, z = -2.71, p <$ .01). These patterns mirrored our experimental data shown in the *ta* and *taziji* experiment.

2.4.3 Intermediate summary

In Experiment 1, we demonstrated the validity of our method which revealed a robust pattern where the syntactic constraint served a major role in the interpretation of the pronoun *ta*. In Experiment 2, with the same structure and lexical items, the reflexive *taziji* showed a significant amount of non-local antecedent choices even when a local gender matching antecedent was available. The results of C.-T. J. Huang, Li, and Li (2009)'s original examples were consistent with the corresponding conditions in our experimental trials, which suggests that the patterns in our experimental data were not driven by the particular lexical items used in our experiments. Although the experiments themselves already provided sufficient evidence for an inherent long-distance binding option for *taziji*, there are still alternative explanations. For example, the use of the matrix verb like "*say*" in Experiment 2 could introduce a "source of information" or "logophoric center" and this logophoricity interpretation could arguably be an important factor responsible for the non-local bindings observed cross-linguistically (see a review: Charnavel et al. 2017). Thus, we shall explore this possibility in Experiments 3 & 4.

3 Experiment 3 & 4: LD *taziji* is not driven by logophoricity

3.1 Overview

In Experiment 3 & 4, we aim to assess the potential confounding factors of logophoricity when interpreting the reflexive *taziji*. As shown in Experiment 2, we used verbs like "*say*" or "*think*" in the matrix clause in order to introduce a clausal complement. Such verbs could potentially introduce a "source of information" (i.e., the long-distance antecedent/Name1 in our case) which is arguably one of the dominant discourse factors constraining logophors (Charnavel et al. 2017), for instance, in the English "picture noun phrase" structure (e.g., Pollard and Sag 1992) and the Chinese logophor *ziji* (e.g., C.-T. J. Huang and Liu 2001). Following this view, the long-distance binding we observed for *taziji* could be driven by an independent effect of the logophoric interpretations.

Unfortunately, there is no uncontroversial diagnostic tool available to verify logophoricity directly (Charnavel et al. 2017);instead our approach here is to exclude this possibility by showing that the long-distance *taziji* observed previously was ***not*** driven by any logophoric effects. If the "logophor hypothesis" is correct, we should expect the choices of the long-distance *taziji* would be modulated by the logophoric conditions: more long-distance antecedents should be selected when

they are the logophoric center (e.g., source of information, attitude holder, etc.) than when they are not. Alternatively, if the long-distance *taziji* is not driven by logophoricity, we should not see any difference in long-distance interpretations for *taziji* whether it is a logophoric center or not.

In these experiments, adopting the same structure in Experiments 1 & 2, we manipulated the matrix verbs to create a "source vs. perceiver" contrast for long-distance antecedents (see Table 2 for an example) and investigate how people interpreted the reflexive *taziji* (Experiment 3) and *ziji* (Experiment 4). We also include the *ziji* as a baseline comparison and to see how different these two types of reflexives are with respect to the sensitivity to the logophoric manipulations.

3.2 Methods

3.2.1 Participants and procedure

Another 94 native Mandarin speakers (64 women; 24.13 +/- 6.74 years) were recruited from the university communities in Nanjing, China. Participants were randomly assigned to one of the experiments (50 for Experiment 3 and 44 for Experiment 4) and paid 10 Chinese Yuan (~1.6 US dollars) for participation. The procedure was similar to Experiment 1 and 2 and implemented on the questionnaire platform *Qualtrics*.

3.2.2 Design and materials

A similar bi-clausal structure (*Name1-think/hear-Name2-V2-taziji/ziji*) with two same-gender names and gender-matching reflexives were used to ask participants to select the antecedents and rate the sentences. We used the same set of names and embedded verbs as in Experiments 1 and 2. We manipulated the logophocity center using the "*renwei/think*" vs. "*tingshuo/hear*" contrast in the matrix verb position. We chose this contrast pair because the verb "think" creates a stronger logophoric center because it indicates both a "source of information" and an "attitude holder" (Charnavel et al. 2017) , and the verb "hear" clearly indicates the opposite of a source as the "perceiver of information" [Kaiser et al. (2009)][7]. In total, 16 sets of critical items and 40 filler sentences were used. A set of sample materials are shown in Table 2.

[7]In a pilot study, we also used the "*biaoshi/say*" vs. "*tingshuo/hear*" pair as the logophocity manipulation. The results also showed a similar pattern as what we found in this experiment.

Table 2: Sample materials and tasks used in Experiment 3 & 4.

Sample martials					
				Experiment 3: reflexive taziji	
孙志	认为/听说	国俊	在	低估	他自己
Sun Zhi	renwei/tingshuo	Guo Jun	zai	digu	*taziji*
male name	think/hear	male name	PROG	underestimate	himself
'Name 1 thinks/hears that Name 2 is underestimating himself.'					
				Experiment 4: reflexive ziji	
孙志	认为/听说	国俊	在	低估	自己
Sun Zhi	renwei/tingshuo	Guo Jun	zai	digu	*ziji*
male/female name	think/hear	male name	PROG	underestimate	self
'Name 1 thinks/hears that Name 2 is underestimating self.'					
Tasks					
**Antecedent choice task:*				**Acceptability judgment task:*	
句子里的他自己/自己指的是谁?				你觉得这句话读起来如何?	
Who does *taziji/ziji* refer to?				How does this sentence sound to you?	
(Name 1 / Name 2)				(7-point scale: very bad : very good)	

3.3 Results and discussions

Figure 3 presents the antecedent choice and acceptability judgment results of Experiment 3 and 4. For *taziji*, consistent with the pattern in Experiment 2, we found about 17.1% long-distance antecedent choices across conditions. However, no effect of the logophoricity manipulation was observed: there was no difference in long-distance choices between the "think" and "hear" conditions ($\hat{\beta} = 0.07, SE = 0.21, z = 0.31, p = .76$). For *ziji*, the antecedent choices results were consistent with the claims in the literature for a typical long-distance reflexive that can take both local and non-local antecedents (~64% vs. 36%). Also as expected, it showed significant more long-distance choices than *taziji* (64% vs. 17%, $\hat{\beta} = 2.19, SE = 0.12, z = 17.87, p < .0001$), but there was no difference between "think" and "hear" conditions under *ziji* either ($\hat{\beta} = 0.14, SE = 0.18, z = 0.79, p = .43$). Lastly, the acceptability ratings did not differ across conditions and anaphors ($ps > .05$).

The results supported our claim that the long-distance *taziji* is not a logophor or any by-product of logophoric effects (e,g., *source of information effect*): no difference in choices of the long-distance antecedent between the source/"think" and the perceiver/"hear" condition. Also, the lack of logophoric effects for *ziji* provided additional evidence against the "logophoricity hypothesis" in such environments (e.g., Pollard and Xue 2000 also argued that "source of information" does not license long distance *ziji*). *Taziji* is still different from *ziji* though in terms of the binding behavior: there was a clear long-distance preference for *ziji* and local preference for *taziji*. However, this difference is due to the different sensitivity to locality rather than logophicity.

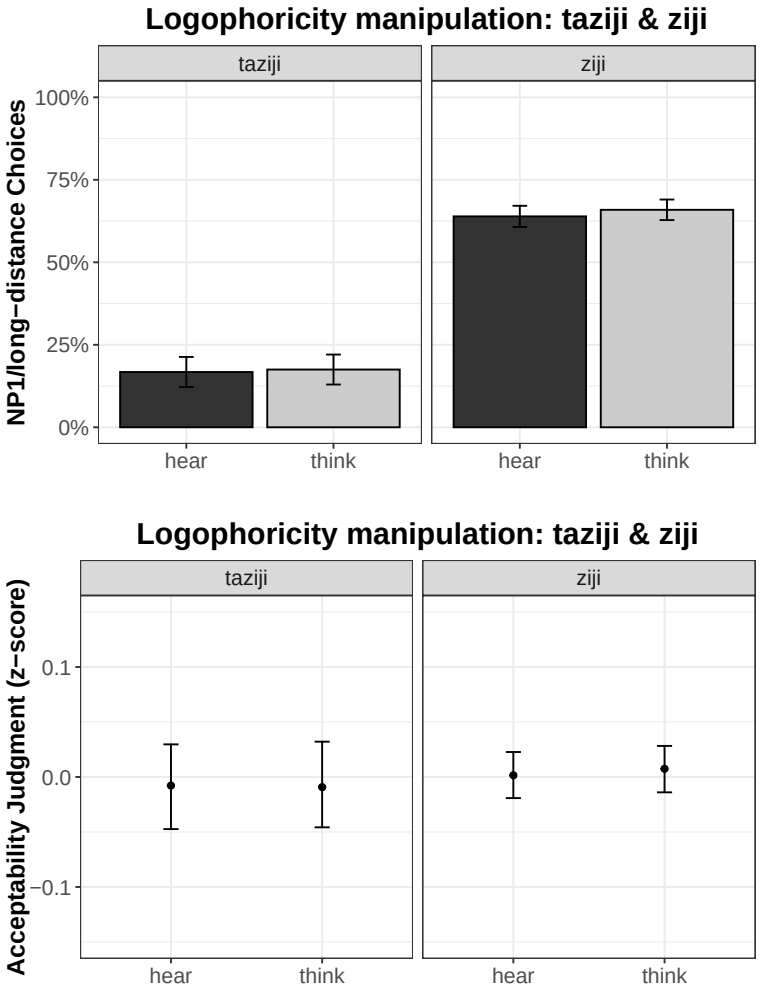

Figure 3: Long-distance Antecedent choice and acceptability judgment (z-score transformed) results for *taziji* (Experiment 3) and *ziji* (Experiment 4). For the choice results plot, dark and light bars represent the long-distance choices/Name1 under "hear"and "think"conditions respectively. The judgment results plot shows the averaged z-score transformed judgment ratings in each condition. Error bars represent the standard error of the subject means.

3.4 Summary of the experiments

Together, these experiments show that the reflexive *taziji* clearly exhibits a long-distance binding option even when a local legitimate antecedent is available. First, the pronoun *ta* conditions demonstrated the validity of our method and revealed the pattern when a strong syntactic constraint does have an effect on the interpretation of the anaphor. Using the pronoun *ta* as a baseline, we found a stronger structural constraint on pronoun conditions than reflexive conditions. For pronouns, people predominantly chose non-local subjects across all conditions. This illustrates a strong structural bias for a non-local antecedent, so strong that it can even override the gender matching requirements of the pronoun. Although people made fewer Target choices when the Target mismatched the gender of the pronoun *ta*, especially when the Competitor gender matched, their low rating indicated their sensitivity to the gender mismatch required in order to make that selection. On the other hand, the reflexive *taziji* showed a long-distance binding option even

when the Target also matched the gender of the reflexive: participants still had almost 20% of the non-local choice and the ratings did not differ from the Target only matched ones. The results of *The syntax of Chinese* textbook examples were consistent with the corresponding conditions in our experimental trials, which suggests that the patterns in our experimental data were not driven by the particular lexical items used in our experiments. Next, we refuted the "logophoricity hypothesis" arguing that the long-distance *taziji* is not a logophor or a by-product of the independent influence of a logophoric interpretation: the selections of the long-distance antecedent did not vary under the different logophoricity conditions. In short, the long-distance binding option is inherent for *taziji*.

4 General discussion

In this study, we explored the degree to which long-distance binding was available for the Mandarin Chinese reflexive *taziji*. We hypothesized that the non-local binding behavior hinted at in the previous literature was due to the fact that the reflexive *taziji* is inherently capable of long-distance binding. To the best of our knowledge, this is the first experimental study to demonstrate the inherent long-distance binding option for *taziji* together with a direct replication of textbook examples. Unlike phenomena where reasoning about syntactic analysis is based on sentence pairs illustrating a sharp contrast (i.e., acceptable vs. unacceptable), we believe that the long-distance binding option for *taziji* may be harder to detect without doing formal experiments: the local-binding option is still much preferred (about 80% of choices) for native speakers. Intuition alone may not be sensitive enough to distinguish whether the local-binding option is a requirement or preference. We do not aim to call into question the value of informal acceptability judgment methods in syntax research as there have been a number of studies demonstrating the reliability of syntax data collected using such methods (e.g., Sprouse and Almeida 2012; Sprouse, Schütze, and Almeida 2013). Instead, this case study aims to illustrate that formal experimental investigation can provide important insight not easily gained in other ways (see also Runner, Sussman, and Tanenhaus 2006). Next, we discuss some implications of our results from both theoretical and methodological perspectives.

4.1 Theoretical consequences

The long-distance binding option for the reflexive *taziji* immediately challenges current syntactic theories on the distribution of Chinese anaphora. Solutions of modifying binding theories like extending binding domains in Chinese may not work because the pronoun *ta* behaves well with respect to the defined binding domain: the antecedent of the pronoun ta must be outside of the binding domain (i.e., the local clause in our case). Alternatively, we could treat *taziji* as a kind of long-distance reflexive similar to *ziji*, but this in turn challenges some of the current views on long-distance reflexives too.

One of the features claimed to be relevant for long-distance reflexives is "monomorphemicity" (Pica 1987). For languages with both monomorphemic and polymorphemic reflexives (e.g., *ziji* and *taziji* in Mandarin Chinese), the claim has been that only the monomorphemic one can be long-distance bound while the polymorphemic is claimed to be strictly local (e.g., C.-T. J. Huang and Liu 2001). This is also argued to be a universal feature of long-distance reflexives cross-linguistically (e.g., Italian: Giorgi 1984; Norwegian: Hellan 1991; see a review in Charnavel et al. 2017). Also, the poly/monomorphemicity distinction in anaphor binding plays a key role in

some of the classic syntactic analyses of long-distance reflexives (e.g., Cole, Hermon, and Sung 1990; C.-T. J. Huang and Liu 2001) which proposed that the long-distance reflexive *ziji* undergoes successive-cyclic local movement in the LF component. Since only monomorphemic *ziji* is an X^0 category which can undergo head-movement, as opposed to polymorphemic/phrasal *taziji*, we should only see the long-distance binding option for *ziji*, and not for *taziji*. Our data clearly call this claim into question.

Then what factors constrain the binding behavior of *taziji*? Does *taziji* follow the same constraints as those for *ziji*? (e.g., "subject orientation," "blocking effects"; see a review in C.-T. J. Huang and Liu 2001). If so, why would a single language have two different types of long-distance reflexive? How do they differ? Cross-linguistically many languages have multiple reflexives, but are claimed to require a unique binding domain for each one: for example, Norwegian has three different reflexives *seg selv* , *seg* and *ham selv* and they differ with respect to their binding domains (see Dalrymple 1993 on LFG Binding Theories). This is clearly not the case for *taziji* and *ziji*: they share the same binding domain as shown in our results.

We suggest a promising first step is to conduct a series of studies directly comparing *ziji* and *taziji*. This may be necessary to tease apart their differences. Only after that can we know which empirical claims can be used to support our theoretical reasoning. One promising direction is to extend some non-uniform/mixed accounts on *ziji* to account for *taziji* as well (e.g., Pan 1998; Pollard and Sag 1992; C.-T. J. Huang and Liu 2001) or to account for both local and long-distance bindings using a different set of principles: see a recent attempt by Liu (2020) who proposed a unified logophoricity theory for both *ziji* and *taziji*.

4.2 Using formal experiments in developing linguistic theories

Here we do not aim to contribute to the recent debate on the reliability of data in the linguistic literature and whether or not linguistics—or syntax in particular—should employ more quantitative methods in developing theories (e.g., Sprouse and Almeida 2012; Gibson and Fedorenko 2013). Instead we want to suggest some advantages of using formal experiments in linguistic theorizing and offer a few tips in collecting judgment data and making the linguistic data more transparent and easier to share with peers.

First, using formal judgment experiments may help linguists attain reliable interpretations of subtle distinctions or controversial phenomena. Taking the current study as an example, although the long-distance binding option was clearly available for the reflexive *taziji*, the local binding option was preferred. This kind of distinction (preferred vs. obligatory), unlike other robust phenomena (e.g., blatant violation of word order), may not be easy to detect using informal means. We are not recommending doing formal experiments for every linguistic phenomenon, but we do believe they can be critical in revealing subtle effects not evident in informal acceptability judgments.

Also, performing formal experiments can help researchers tease apart different sources of constraints that influence a particular linguistic phenomenon. This can be useful for "interface" types of questions, for example, testing syntactic and semantic effects on anaphor resolution (e.g., Kaiser et al. 2009). Constructing sets of examples that vary on particular features in a controlled way can make it possible to tease apart the separate effects of each feature. In our case, we wanted to test the influence of both gender and structural constraints on people's interpretations of the reflexive taziji, so we created four types of sentences manipulating the position of the antecedents

and their gender type. By doing so, we can see the influence of the syntactic and gender constraints (main effects) as well as their interaction (interaction effects) from the statistical analysis (see also Sprouse et al. 2016 for a similar approach on "island effects").

Lastly, doing formal experiments can provide more transparency of the linguistic data used in the field. It is very common that for non-native linguists, the judgments of linguistic phenomenon heavily rely on the data provided by other linguists who are experts or native speakers of that language. Experimental results can inform researchers of the difference between two conditions (i.e., theory-driven contrast pairs used in linguistic theorizing) more than binary distinctions. Presenting raw data with statistical analysis help other researchers know more about a phenomena and evaluate the arguments proposed to analyze it[8].

To recapitulate, contradicting the predominant claim in the field, we provide strong evidence from both formal experiments and replications of the textbook examples showing that the Chinese reflexive *taziji* has a long-distance binding option. Although further work is needed to fully characterize the binding behaviors of Mandarin Chinese reflexives in order to have a unified theory, we believe this study is a promising start and hope the experimental approach presented here will benefit the future studies.

Open practices statement

The materials, data and code are available in the OSF project repository: https://shorturl.at/ikQV8

References

Asudeh, Arshia. 1998. "Anaphora and argument structure: Topics in the syntax and semantics of reflexives and reciprocals." MPhil Dissertation, University of Edinburgh.

Barr, Dale J., Roger Levy, Christoph Scheepers, and Harry J. Tily. 2013. "Random Effects Structure for Confirmatory Hypothesis Testing: Keep It Maximal." *Journal of Memory and Language* 68 (3): 255–78. https://doi.org/10.1016/j.jml.2012.11.001.

Chang, Wenshuo, Yunyan Duan, Jingjing Qian, Fuyun Wu, Xiaoming Jiang, and Xiaolin Zhou. 2020. "Gender Interference in Processing Chinese Compound Reflexive: Evidence from Reading Eye-Tracking." *Language, Cognition and Neuroscience* 35 (10): 1355–70. https://doi.org/10.1080/23273798.2020.1781213.

Charnavel, Isabelle, C.-T. James Huang, Peter Cole, and Gabriella Hermon. 2017. "Long-Distance Anaphora: Syntax and Discourse." In *The Wiley Blackwell Companion to Syntax, Second Edition*, edited by Martin Everaert and Henk C. van Riemsdijk, 1–82. Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/9781118358733.wbsyncom074.

Chen, Zhong, Lena Jäger, and Shravan Vasishth. 2012. "How Structure-Sensitive Is the Parser? Evidence from Mandarin Chinese." In *Empirical Approaches to Linguistic Theory*, edited by Britta Stolterfoht and Sam Featherston, 43–62. DE GRUYTER. https://doi.org/10.1515/9781614510888.43.

[8]We also realize that there are some cases in which running experiments is extremely difficult, or not practical; for example, studies on endangered languages or contexts where native speakers are hard to reach. However, approaches have been proposed to address these limitations as well, for instance, the small N acceptability paradigm (Mahowald et al. 2016) to deal with the problem of limited participants.

Chen, Zhong, Yuhang Xu, and Zhiguo Xie. 2020. "Assessing Introspective Linguistic Judgments Quantitatively: The Case of The Syntax of Chinese." *Journal of East Asian Linguistics* 29 (3): 311–36. https://doi.org/10.1007/s10831-020-09210-y.

Chomsky, Noam. 1981. *Lectures on Government and Binding: The Pisa Lectures*. Studies in Generative Grammar [SGG]. De Gruyter. https://books.google.com/books?id=XiJTlengwloC.

Cole, Peter, Gabriella Hermon, and Li-May Sung. 1990. "Principles and Parameters of Long-Distance Reflexives." *Linguistic Inquiry* 21 (1): 1–22. http://www.jstor.org/stable/4178657.

Dalrymple, Mary. 1993. *The Syntax of Anaphoric Binding*. CSLI Publications.

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 (September). https://doi.org/10.3389/fpsyg.2014.01025.

Dillon, Brian, Wing-Yee Chow, and Ming Xiang. 2016. "The Relationship Between Anaphor Features and Antecedent Retrieval: Comparing Mandarin Ziji and Ta-Ziji." *Frontiers in Psychology* 6 (January). https://doi.org/10.3389/fpsyg.2015.01966.

Gibson, Edward, and Evelina Fedorenko. 2013. "The Need for Quantitative Methods in Syntax and Semantics Research." *Language and Cognitive Processes* 28 (1-2): 88–124. https://doi.org/10.1080/01690965.2010.515080.

Giorgi, Alessandra. 1984. "Towards a Theory of Long Distance Anaphors: A GB Approach." *The Linguistic Review* 3 (4): 307–61. https://doi.org/10.1515/tlir.1984.3.4.307.

He, Xiao, and Elsi Kaiser. 2016. "Processing the Chinese Reflexive "Ziji": Effects of Featural Constraints on Anaphor Resolution." *Frontiers in Psychology* 7 (April). https://doi.org/10.3389/fpsyg.2016.00284.

Hellan, Lars. 1991. "Containment and Connectedness Anaphors." In *Long-Distance Anaphora*, edited by Jan Koster and Eric Reuland, 27–48. Cambridge, UK: Cambridge University Press.

Hu, J. 2019. *Prominence and Locality in Grammar: The Syntax and Semantics of Wh-Questions and Reflexives*. Routledge Studies in Chinese Linguistics. Taylor & Francis.

Huang, Cheng-Teh James, Yen-hui Audrey Li, and Yafei Li. 2009. *The Syntax of Chinese*. Cambridge, UK ; New York: Cambridge University Press. https://doi.org/10.1017/CBO9781139166935.

Huang, Cheng-Teh James, and Chen-Sheng Liu. 2001. "Logophoricity, Attitudes, and Ziji at the Interface." In *Between Syntax and Semantics*, 300–343. Routledge Taylor & Francis Group.

Huang, C.-T. James, Yen-hui Audrey Li, and Yafei Li. 2013. 汉语句法学/*The Syntax of Chinese*. Translated by Zhang Heyou and Gu Yang. Beijing,China: 世界图书出版公司.

Huang, Yun-Hua. 1984. "Reflexives in Chinese." *Studies in English Literature and Linguistics* 10: 163–88.

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 (May). https://doi.org/10.3389/fpsyg.2015.00617.

Juzek, Tom. 2015. "Acceptability Judgement Tasks and Grammatical Theory." Doctoral dissertation, University of Oxford.

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 (1): 55–80. https://doi.org/10.1016/j.cognition.2009.03.010.

Li, Xiaoqian, and Xiaolin Zhou. 2010. "Who Is Ziji? ERP Responses to the Chinese Reflexive Pronoun During Sentence Comprehension." *Brain Research* 1331 (May): 96–104. https://doi.org/10.1016/j.brainres.2010.03.050.

Liu, Yingtong. 2020. "Logophoricity and Mandarin Exempt Reflexives." *University of Pennsylvania Working Papers in Linguistics* 26 (1): 18.

Lyu, Jun, and Elsi Kaiser. 2021. "Antecedent Prominence and the Chinese Reflexive Ziji." Poster presented at the *34th Annual CUNY Conference on Human Sentence Processing*. https://www.cuny2021.io/2021/02/24/279/.

Mahowald, Kyle, Peter Graff, Jeremy Hartman, and Edward Gibson. 2016. "SNAP Judgments: A Small N Acceptability Paradigm (SNAP) for Linguistic Acceptability Judgments." *Language* 92 (3): 619–35. https://doi.org/10.1353/lan.2016.0052.

Pan, Haihua. 1998. "Closeness, Prominence, and Binding Theory." *Natural Language and Linguistic Theory* 16 (4): 771–815. https://doi.org/10.1023/A:1006056414208.

Pan, Haihua, and Jianhua Hu. 2003. "Prominence and Locality in the Binding of Mandarin Complex Reflexive "Ta-Ziji (s/He-Self)." *Journal of Chinese Linguistics Monograph Series*, no. 19: 152–70. http://www.jstor.org/stable/23825496.

Pica, Pierre. 1987. "On the Nature of the Reflexivization Cycle." In *Proceedings of the North East Linguistic Society*, 17:483–500. 2.

Pollard, Carl, and Ivan A. Sag. 1992. "Anaphors in English and the Scope of Binding Theory." *Linguistic Inquiry* 23 (2): 261–303. https://www.jstor.org/stable/4178768.

Pollard, Carl, and Ping Xue. 2000. "Syntactic And Nonsyntactic Constraints On Long-Distance Reflexives." In *Long Distance Reflexives*, edited by Peter Cole, Gabriella Hermon, and C.-T. James Huang, 317–42. BRILL. https://doi.org/10.1163/9781849508742_011.

Qiu, Lijing, Tamara Y. Swaab, Hsuan-Chih Chen, and Suiping Wang. 2012. "The Role of Gender Information in Pronoun Resolution: Evidence from Chinese." Edited by Kevin Paterson. *PLoS ONE* 7 (5): e36156. https://doi.org/10.1371/journal.pone.0036156.

Reinhart, Tanya, and Eric Reuland. 1993. "Reflexivity." *Linguistic Inquiry* 24 (4): 657–720.

Runner, Jeffrey T., Rachel S. Sussman, and Michael K. Tanenhaus. 2006. "Processing Reflexives and Pronouns in Picture Noun Phrase." *Cognitive Science* 30 (2): 193–241. https://doi.org/10.1207/s15516709cog0000_58.

Schumacher, Petra B., Walter Bisang, and Linlin Sun. 2011. "Perspective in the Processing of the Chinese Reflexive Ziji: ERP Evidence." In *Anaphora Processing and Applications*, edited by Iris Hendrickx, Sobha Lalitha Devi, António Branco, and Ruslan Mitkov, 119–31. Berlin, Heidelberg: Springer Berlin Heidelberg.

Schütze, Carson T, and Jon Sprouse. 2013. "Judgment Data." In *Research Methods in Linguistics*, edited by Robert J. Podesva and Devyani Sharma, 27–50.

Shu, Hua. 2003. "Chinese Writing System and Learning to Read." *International Journal of Psychology* 38 (5): 274–85. https://doi.org/10.1080/00207590344000060.

Sprouse, Jon, and Diogo Almeida. 2012. "Assessing the Reliability of Textbook Data in Syntax: Adger's *Core Syntax*." *Journal of Linguistics* 48 (3): 609–52. https://doi.org/10.1017/S0022226712000011.

Sprouse, Jon, Ivano Caponigro, Ciro Greco, and Carlo Cecchetto. 2016. "Experimental Syntax and the Variation of Island Effects in English and Italian." *Natural Language & Linguistic Theory* 34 (1): 307–44. https://doi.org/10.1007/s11049-015-9286-8.

Sprouse, Jon, Carson T. Schütze, and Diogo Almeida. 2013. "A Comparison of Informal and Formal Acceptability Judgments Using a Random Sample from Linguistic Inquiry 2001–2010." *Lingua* 134 (September): 219–48. https://doi.org/10.1016/j.lingua.2013.07.002.

Su, Jui-Ju, Nicola Molinaro, Margaret Gillon-Dowens, Pei-Shu Tsai, Denise H. Wu, and Manuel Carreiras. 2016. "When "He" Can Also Be "She": An ERP Study of Reflexive Pronoun Resolution in Written Mandarin Chinese." *Frontiers in Psychology* 7 (February). https://doi.org/10.3389/fpsyg.2016.00151.

Tang, Chih-Chen Jane. 1989. "Chinese Reflexives." *Natural Language and Linguistic Theory* 7 (1): 93–121. https://doi.org/10.1007/BF00141348.

To, Wai-Ming, and Linda S.L. Lai. 2015. "Crowdsourcing in China: Opportunities and Concerns." *IT Professional* 17 (3): 53–59. https://doi.org/10.1109/MITP.2015.47.

Xu, Liejiong. 1993. "The Long-Distance Binding of Ziji." *Journal of Chinese Linguistics* 21 (1): 123–42. http://www.jstor.org/stable/23756128.

Xu, Yuhang, and Jeffrey T. Runner. 2019. "Memory Retrieval in Reflexives Processing: Evidence Against the Local Search Hypothesis." *Working Papers in the Language Sciences at the University of Rochester* vol. 7 (1).

Yu, William Xian-fu. 1992. "Challenging Chinese Reflexive Data." *The Linguistic Review* 9 (3). https://doi.org/10.1515/tlir.1992.9.3.285.

———. 2000. *Chinese Reflexives*. Vol. 15. Orbis. Supplementa 15. Levven: Peeters.

Zhan, Meilin, Roger Levy, and Andrew Kehler. 2020. "Pronoun Interpretation in Mandarin Chinese Follows Principles of Bayesian Inference." Edited by Yang Xu. *PLOS ONE* 15 (8): e0237012. https://doi.org/10.1371/journal.pone.0237012.