

Who's that? The binding of Mandarin reflexives

Ying Zhang

¹ Rutgers University

Who's that? The binding of Mandarin reflexives

1. Introduction

Various factors can affect the binding of reflexives. This study investigates how syntactic structure influences binding by examining the choice of binders for Mandarin reflexives in both full and elided structures. There are two reflexives— *ziji* and *taziji* in Mandarin which behave differently regarding their ability to take a non-local binder. *ziji* can take both a local and a non-local binder, whereas *taziji* is more restricted to a local binder.

(1) John₁ said Bill₂ liked *ziji*_(1/2).

(2) John₁ said Bill₂ liked *taziji*_(*1/2).

2. Methods

The experiment examines whether the choice of binders is influenced by syntactic structure. Two tasks are conducted: a free choice task, which explores the range of potential binders, and a binary choice task, which tests whether a specific binder can be selected.

We report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures in the study.

2.1 Participants

Native Mandarin speakers are recruited over the internet.

2.2 Materials and design

There are 80 target items. Each target item consists of a dialogue, with the answer being the critical sentence containing the tested reflexive. Each item includes two questions: Q1 (a free choice question) and Q2 (a binary choice question). The critical sentences follow a 2×2 design, crossing sentence type (full/elided) with reflexive type (*ziji*/*taziji*). An example of a target item is provided below.

(3)

A: Who did John say that Bill liked?

B: John said that Bill liked **reflexive** (full)./ **Reflexive** (elided).

Q1: Who do you think the reflexive refers to? John; Bill

Q2 Can the reflexive refer to Bill/John?

The 80 target items are distributed across four surveys, each defined by a combination of reflexive type and sentence structure. Within each survey, 20 target items are mixed with 20 fillers in a pseudo-randomized order.

2.3 Procedure

The experiment is conducted on Qualtrics. On each trial, participants first see a dialogue. Once they proceed, the dialogue disappears and is replaced by Q1. Participants select one of the two referents. After completing Q1, they are presented with Q2, in which they indicate whether a specific referent can be the binder of the reflexive by choosing “yes” or “no.”

2.4 Predictions

Taziji shows a higher rate of non-local binders in elided structures than in full structures. In contrast, *ziji* exhibits a similar distribution of local and non-local binders across both structures.

2.5 Data analysis

We used R (Version 4.4.2; R Core Team, 2024) and the R-packages *papaja* (Version 0.1.3; Aust & Barth, 2024) and *tinylabels* (Version 0.2.5; Barth, 2025) for all our analyses.

A mixed-effects logistic regression model is used to analyze the binary choice of binders, implemented using the lme4 package in R. The dependent variable is *response* (0 = non- local binder / 1 = local binder). The fixed effects are *survey* (ziji/taziji) and *sentence*, with random intercepts for each subject.

The significance of the main effects and possible interaction (reflexive type \times structure type) is assessed with hierarchical partitioning of the variance via nested model comparison.

3. Results

The code chunk below shows how the data is tidied, followed by a preview of the resulting dataset.

```
dat1 <- dat_raw |>
  pivot_longer(
    cols      = starts_with("item_"),
    names_to  = c("item", "blocking", "question"),
    names_prefix = "item_",
    names_sep  = "[.]",
    values_to = "response"
  ) |>
  mutate(
    item = as.integer(item),
    blocking = blocking
  )

knitr::kable(head(dat1), digits = 3 )
```

id	survey	sentence	item	blocking	question	response
P001	taziji	full	1	blocking	q1	1
P001	taziji	full	2	blocking	q1	1
P001	taziji	full	3	blocking	q1	1
P001	taziji	full	4	blocking	q1	1
P001	taziji	full	5	blocking	q1	1
P001	taziji	full	6	blocking	q1	1

57

Below is the code chunk for the plot. The plot shows the proportion of local and non-local binder responses. For *taziji*, the proportion of non-local binders is higher in the elided structure than in the full structure. In contrast, *ziji* exhibits a similar distribution of binders across both structure types.

58

59

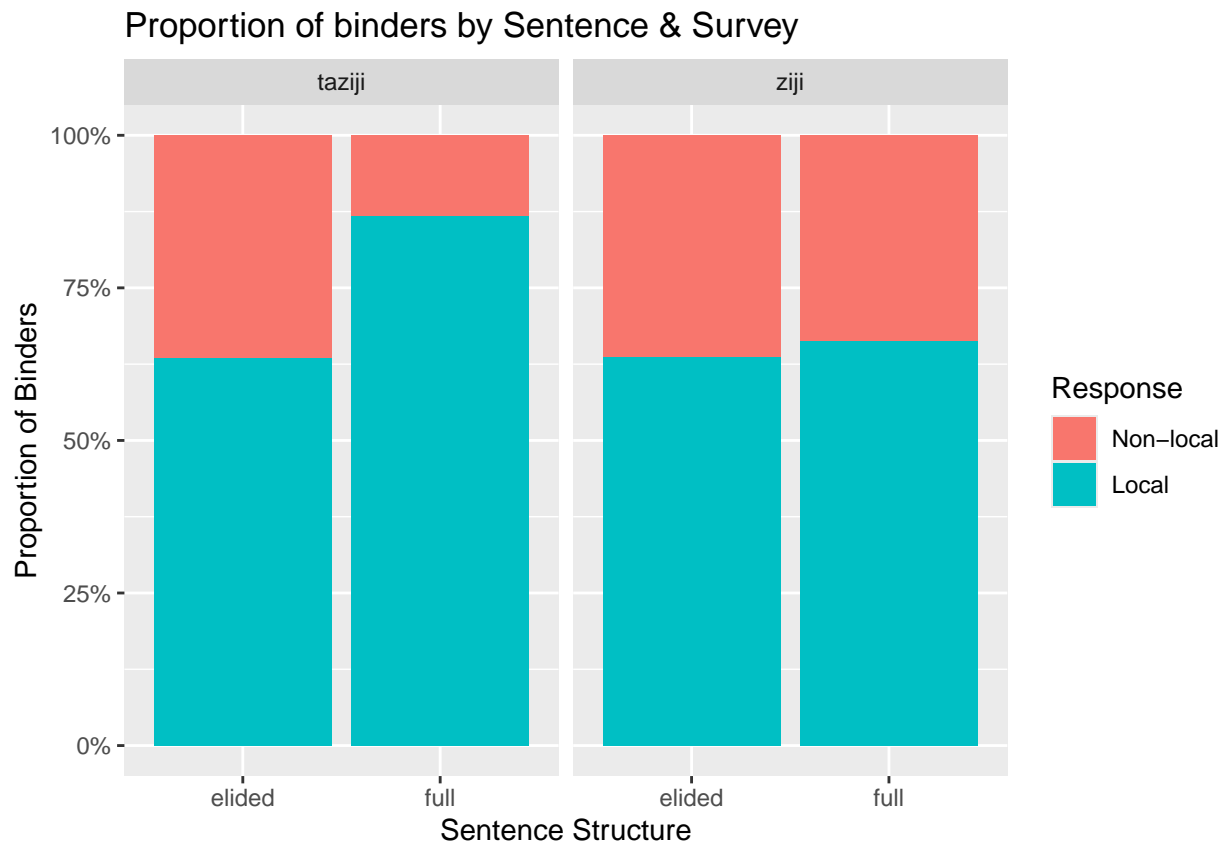
60

61

```

dat1 |>
  ggplot() +
  aes(x = sentence, fill = factor(response, labels = c("Non-local", "Local")))+
  geom_bar(position = "fill") +
  facet_wrap(~ survey) +
  scale_y_continuous(labels = percent_format())+
  labs(
    x = "Sentence Structure",
    y = "Proportion of Binders",
    fill = "Response",
    title = "Proportion of binders by Sentence & Survey"
  )

```



62

63 First, the data is analyzed with a null model.

```
mod_0 <- glmer(
  formula = response ~ 1 + (1 | id),
  family = binomial(link = "logit"),
  data = dat1
)

mod_summary_0 <- tidy(mod_0,
  effects = "fixed",
  conf.int = TRUE
)

knitr::kable(mod_summary_0, digits = 3 )
```

effect	term	estimate	std.error	statistic	p.value	conf.low	conf.high
fixed	(Intercept)	0.902	0.07	12.839	0	0.764	1.039

As shown in the table above, there is a significant baseline preference for local binders ($p < .001$), consistent with theoretical predictions. However, this result does not provide information about the non-local binding.

Next, an inclusive model is used to analyze the data. A comparison with the null model shows that the inclusive model is a better fit, as indicated by its lower AIC value (3774.3).

```
mod_1 <- glmer(
  formula = response ~ sentence*survey + (1 | id),
  family = binomial(link = "logit"),
  data = dat1
)

anova_tidy_1 <- tidy(anova(mod_1, mod_0,
                           test="Chisq"))

knitr::kable(anova_tidy_1, digits = 3)
```

term	npar	AIC	BIC	logLik	minus2logL	statistic	df	p.value
mod_0	2	3848.687	3860.828	-1922.343	3844.687	NA	NA	NA
mod_1	5	3757.618	3787.972	-1873.809	3747.618	97.069	3	0

The log-odds of a local-binder response is 0.5537 when both sentencefull and surveyziji are coded as 0, indicating that approximately 63% of responses are local binders in the elided-taziji condition. The significant main effect of sentence ($p < .001$) suggests that local binders are more likely in full structures than in elided ones. The significant sentence \times survey interaction shows that full structures substantially increase local-binder choice for

76 *taziji*, whereas *ziji*'s binder preferences remain largely unchanged across structure types.

	effect	term	estimate	std.error	statistic	p.value	conf.low	conf.high
	fixed	(Intercept)	0.554	0.073	7.540	0.000	0.410	0.698
77	fixed	sentencefull	1.325	0.128	10.391	0.000	1.075	1.575
	fixed	surveyziji	0.005	0.104	0.052	0.959	-0.198	0.209
	fixed	sentencefull:surveyziji	-1.210	0.165	-7.329	0.000	-1.534	-0.886

78 ## [1] 0.6349936

79 A restricted model is compared with the inclusive model. The AIC values show that
80 the inclusive model yields a better fit. The interaction effect is significant.

```
mod_2 <- glmer(
  formula = response ~ sentence + survey + (1 | id),
  family = binomial(link = "logit"),
  data = dat1
)
anova_tidy_2 <- tidy(anova(mod_2, mod_1,
                           test="Chisq")
)
knitr::kable(anova_tidy_2, digits = 3)
```

	term	npar	AIC	BIC	logLik	minus2logL	statistic	df	p.value
81	mod_2	4	3800.260	3824.543	-1896.130	3792.260	NA	NA	NA
	mod_1	5	3757.618	3787.972	-1873.809	3747.618	44.642	1	0

82 Significant main effects of sentence type ($\beta = 0.668$, $p < .001$) and reflexive type
83 ($\beta = -0.540$, $p < .001$) on participants' binder choices are observed.

84 The log-odds of a local binder response is 0.839 when both sentencefull and surveyziji

are 0. This corresponds to a predicted probability of approximately 70% for a local binder response in elided sentences within the *taziji* survey.

Holding *surveyziji* at 0, a one unit increase in *sentencefull* is associated with a 0.668 increase in log-odds. When holding *sentencefull* at 0, one unit increase in *surveyziji* is associated with a 0.540 decrease in log-odds. This means that the odds of a local binder response decrease by about 42% in the *ziji* survey compared to the *taziji* survey.

effect	term	estimate	std.error	statistic	p.value	conf.low	conf.high
fixed	(Intercept)	0.839	0.087	9.635	0	0.668	1.009
fixed	<i>sentencefull</i>	0.668	0.103	6.486	0	0.466	0.870
fixed	<i>surveyziji</i>	-0.540	0.103	-5.240	0	-0.741	-0.338

4. Discussion

The inclusive model is selected for the analysis. Sentence type and reflexive type both show significant main effects on binder choice. While the result primarily reflects the choice of local binders, it also indirectly indicates a difference in non-local binder selection. Specifically, non-local binders are more likely to be chosen in the elided structure within the *taziji* survey, supporting the prediction.

References

98

- 99 Aust, F., & Barth, M. (2024). *papaja: Prepare reproducible APA journal articles with R*
100 *Markdown*. <https://doi.org/10.32614/CRAN.package.papaja>
- 101 Barth, M. (2025). *tinylabels: Lightweight variable labels*.
102 <https://doi.org/10.32614/CRAN.package.tinylabels>
- 103 R Core Team. (2024). *R: A language and environment for statistical computing*. Vienna,
104 Austria: R Foundation for Statistical Computing. Retrieved from
105 <https://www.R-project.org/>