A Multifactorial Corpus-based Analysis of Classifier Positioning in Mandarin Relative Clauses

Jingwen CAO[1,*], Lawrence Yam-Leung Cheung[1]

[1]The Chinese University of Hong Kong

*Corresponding author email: jingwencao@cuhk.edu.hk

Background: In Mandarin, the modifying relative clause (RC) can either precede the head noun (1) or be separated by the demonstrative (Dem), numeral (Num) and classifier (CL) sequence (2). The two constructions differ in that while (1) is ambiguous in specificity, (2) exclusively conveys specificity (Zhang, 2006). Language users can choose either variant depending on the specific contextual conditions.

Gap: Previous studies (Sheng & Wu, 2013; Wu & Sheng, 2014) have demonstrated a correlation between the grammatical position of the head noun and positioning variation. They proposed processing-driven principles, namely the *Early Occurrence Strategy* and the *Semantic Clash Avoidance Strategy*, to explain the observed distribution. However, these studies primarily focused on individual factors, failing to capture the simultaneous contribution of various syntactic, semantic and cognitive constraints in real communicative contexts. This study adopts a multifactorial corpus-based approach to investigate the probabilistic factors constraining classifier positioning and to explore the communicative principles underlying the choice between two near-synonymous constructions.

Method: A total of N=498 observations across 4 genres (as detailed in Table 1) were extracted from the Beijing Language and Culture University Corpus Center corpus (BCC corpus) and annotated based on 8 explanatory variables (see Table 2) proposed in previous studies. Logistic regression was then conducted to evaluate the effects of predictors and their interactions.

Results: **First**, our results reveal that the occurrence of two constructions is constrained by various factors, with *RelGPos* (the grammatical position of the head NP in RCs) being the most significant. Specifically, classifiers tend to precede subject RCs but follow object RCs, consistent with previous studies. **Second**, *RCLength*, though received less empirical evidence in the literature on classifier positioning, has also been shown to be a significant predictor. Our results suggest that longer RCs increase the probability of the [RC-Dem/Num-Cl] sequence regardless of *RelGPos* (as shown in *Fig. 1*). This tendency shows some cross-linguistic parallelism, i.e., a peripheral placement of the longer and more complex dependents relative to the head (Futrell et al. 2020, Gibson et al. 2019). Language users tend to prepose RCs to minimize the length of classifier-noun dependencies, ensuring more efficient processing of language information. Besides, the interaction between *RCLength* and *Genre* is also observed in the model, which indicates that the strength and direction of *RCLength* on *Construction* varies depending on *Genre* (as shown in *Fig. 2*). Since Chinese RCs primarily provide background information due to their structural features (i.e., RCs preceding the head noun), language users, especially in formal speech contexts (e.g., newspapers and academic writing), tend to prepose RCs to a more prominent position to enhance message saliency.

Examples:

(1) san-ge [$_{RC}$ t$_i$ <u>dai yanjing</u> de] xuesheng$_i$ (2) [$_{RC}$ t$_i$ <u>dai yanjing</u>   de] san-ge xuesheng$_i$

   3-CL   t$_i$ wear glass DE student$_i$ t$_i$ wear glass   DE 3-CL   student$_i$

   'three students who wear glasses'   'three students who wear glasses'

Register	Dem/Num-Cl-RC	RC-Dem/Num-Cl	Total
Newspaper	20 (31.7%)	43 (68.3%)	63
Weblogs	60 (58.3%)	43 (41.7%)	103
Literature	50 (33.3%)	100 (66.7%)	150
Academic writing	100 (54.9%)	82 (45.1%)	182

Table 1. Distribution of constructions in the dataset by genre

| Variable	Meaning and Levels (**Reference level**)
Genre	Source of the sentence (**newspapers**, weblogs, literature and academic writing)
DetType	Determiner type (**demonstrative,** numeral)
RelGPos	Grammatical position of head NP in relative clauses (**SRC**, ORC)
MatGPos	Grammatical position of head NP in the matrix clause (**subject**, object, others)
HeadNPAni	Animacy of head NP (**animate**, concrete inanimate, abstract inanimate)
EmbNPAni	Animacy of the argument in RC (**animate**, concrete inanimate, abstract inanimate, n/a)
HeadNPLength	Length of head NP in number of Chinese characters
RCLength	Length of RCin number of Chinese characters (incl. 'DE')
Construction	Construction of the sentence (**Dem/Num-Cl-RC**, RC-Dem/Num-Cl)

Table 2. Coding scheme
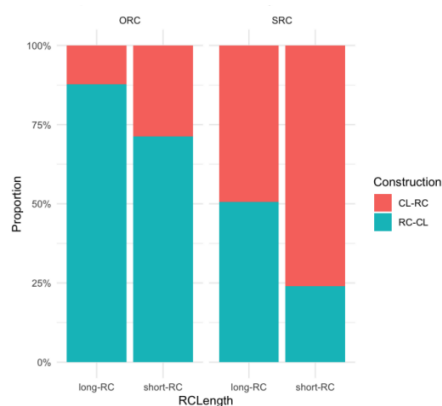
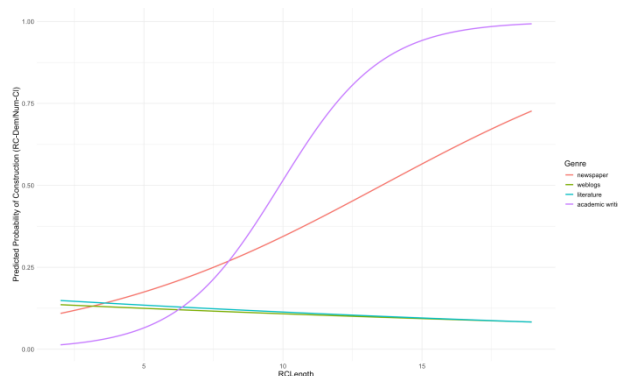**Figure 1.** Distribution of construction by *RelGPos* and *RCLength*

Figure 2. Interaction between *RCLength* and *Genre*

References:

Futrell, R., Levy, R., & Gibson, E. (2020). Dependency locality as an explanatory principle for word order. *Language, 96*(2), 371–412.

Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences, 23*(5), 389–407.

Sheng, Y., & Wu, F. (2013). Demonstrative-classifier positioning in Chinese relative clauses and its underlying reasons: A spoken corpus stud*y. Modern Foreign Languages, 2, 150*−*157.*

Wu, F., & Sheng, Y. (2014). Demonstrative-classifier positioning in Chinese relative clauses and its implication to language production models. *Journal of Foreign Languages, 37*(3), 49–58.

Zhang, N. (2006). Representing specificity by the internal order of indefinites. *Linguistics, 44*(1), 1–21.