

A multivariate corpus analysis of locative inversion in Mandarin Chinese

Yuting Li (Freie Universität Berlin)

Abstract. One challenge that figures prominently in the functionalist and psycholinguistic studies is identifying the factors that determine the variation in constituent ordering in a language. This study presents a corpus-based investigation of locatives in two alternating constructions in Mandarin Chinese, to evaluate the effect of potentially motivating factors in the choice of locative inversion construction and the canonical construction. Locative inversion, characterized by a preposed locative and a postposed NP that functions as the subject in the canonical sentence, has been documented in various languages including English, Chicheŵa, and Mandarin Chinese (Bresnan and Kanerva 1989; Bresnan 1994; Levin 1993; One-Soon 2003; Paul et al. 2020). Consider Chinese sentences:

- a. Tái-shàng zuò-zhe zhǔxítuán (locative inversion construction: Locatives+V+NP)
platform-on sit-DUR presidium
'On the platform was sitting the presidium.'
- b. Zhǔxítuán zuò zài tái-shàng (canonical construction: NP+V+Locatives)
presidium sit at platform-on
'The presidium was sitting on the platform.'

Early empirical studies on locative inversion have concerned solely one single factor, excluding the interaction with others (Levin and Rappaport Hovav 1994; Birner 1994; Holler and Hartmann 2012). Moreover, most claims on locative inversion have not yet been quantitatively assessed, we lack a good understanding of how factors influence the choice between locative inversion and its canonical variant. Using mixed-effects logistic regression, the current study conducted a quantitative analysis of 899 instances (562 locative inversion and 337 canonical constructions) collected from the Center for Chinese Linguistics (CCL) Corpus ¹, which were annotated for 6 factors.

Preliminary results showed that our model, which utilized 6 factors, achieved an average prediction accuracy of 81% for unseen data, suggesting a good fit (baseline: 63%). To further evaluate the effect of verb semantics, we included 10 verb classes as additional factors in our model. This new model showed a remarkable improvement, with the average prediction accuracy increasing from 81% to 87%. Our findings i) identified the importance of verb meaning in locative inversion, which had been previously overlooked, ii) provided evidence for the significance of 8 among factors considered, and iii) demonstrated that these factors are favored differently in alternating constructions: while locative inversion is favored by an indefinite, new, long theme and an animate, given, short location, the canonical variant is favored by a definite, given, short theme and an inanimate, new, long location. This aligns with the pattern of 'Harmonic Alignment' (Aissen 1999; Bresnan et al. 2007), which suggests that in sentences, referent objects that are given, definite, animate, and shorter in length tend to precede those that are new, indefinite, inanimate, and longer in length. Additionally, we found that the preference for verb classes also differs. Verbs of emission tend to occur more frequently in locative inversion, whereas verbs of activities and creation are more likely to occur in the canonical variant.

¹ http://ccl.pku.edu.cn:8080/ccl_corpus/index.jsp