

# A multifactorial analysis of Chinese analytic passive constructions

Sumin Guan, Benedikt Szmrecsanyi & Weiwei Zhang  
(KU Leuven, KU Leuven & KU Leuven)

Keywords: Chinese; passive; probabilistic grammar; variation; regression

Analytic passive constructions marked by *bèi/ràng/gěi* (1-3) are the predominant strategy to indicate passive voice in Mandarin Chinese.

- (1) zhTenTen17 (Simplified) chinadevelopment.com.cn marker=*bèi*  
[这 名 弃 婴]<sub>NP1</sub> 被 [福利院]<sub>NP2</sub> 接 走 了。  
*zhè míng qì yīng bèi fúliyuàn jiē zǒu le*  
this CLF abandoned baby PASS welfare pick up PST  
“This abandoned baby was picked up by the welfare home.”
- (2) zhTenTen17 (Simplified) cye.com.cn marker=*ràng*  
[钱]<sub>NP1</sub> 都 让 [老板]<sub>NP2</sub> 赚 去 了。  
*qián dōu ràng lǎobǎn zhuàn qù le*  
money all PASS boss earn RES PST  
“All the money was earned by the boss.”
- (3) zhTenTen17 (Traditional) slime.com.tw marker=*gěi*  
[狐狸精]<sub>NP1</sub> 给 [鬼]<sub>NP2</sub> 捉 去 了。  
*húlijīng gěi guǐ zhuō qù le*  
fox PASS ghost capture RES PST  
“The fox was captured by the ghost.”

These analytic passive variants are in principle well-studied (Li & Thompson 1989, Zhu 1982, Guo & Chow 2014, and Ding 2020), but the alternation has yet to be fully analyzed using variationist techniques to explore the probabilistic grammatical knowledge that language users have about the Chinese long passive constructions. To explore the variation synchronically, we tap into the *Chinese Web Corpus* (zhTenTen17 Simplified and zhTenTen17 Traditional). The objective is to identify the language-internal factors that predict the choice between Chinese *bèi/ràng/gěi* long passives, and to establish if variant choice differs as a function of language variety (i.e., Mainland Chinese versus Taiwan Chinese). To this end, we extracted a rather balanced dataset, totaling 4,172 observations of Chinese analytic long passive constructions. Drawing inspiration from the literature on passive constructions in Chinese, we manually annotated all tokens for 17 syntactic and semantic language-internal factors.

To determine the combined effect of multiple linguistic factors, mixed-effects logistic regression is used, following best practice in variationist sociolinguistics (Tagliamonte 2012). Six factors remain consistent across varieties of Mandarin Chinese, delineating the different usage of these three variants, while eight factors exhibit fluidity across varieties of Mandarin Chinese, specifying different usage patterns of these three variants in two language varieties. Comparing to Mainland Chinese, we conclude that Taiwan Chinese aligns more closely with the archaic requirements of passive constructions in Mandarin Chinese.

## References

- Ding, Yingying (2020), A Corpus-Based Study of Passive Constructions in Chinese. In B. Ahrens, M. Beaton-Thome, M. Krein-Kühle, L. Link, and U. Wienen (eds), (2020), *Interdependence and Innovation in Translation, Interpreting and Specialised Communication*, Berlin: Frank & Timme GmbH, 173–186.
- Guo, Hongjie, and Daryl Chow (2014), A Corpus-Based Variationist Approach to Bei Passives in Mandarin Chinese. *Corpus Linguistics and Linguistic Theory* 10(1), 139–173.
- Li, Charles N., and Sandra A. Thompson (1989), *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: CA University of California Press.
- Tagliamonte, Sali A (2012), *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden & Oxford: Wiley-Blackwell.
- Zhu, Dexi (1982), *语法讲义 [Yufa Jiangyi]*. Beijing: The Commercial Press.