Papers of the Linguistic Society of Belgium

18 | 2024

A multifactorial analysis of Chinese analytic long passive constructions marked by *bèi, gěi* and *ràng* in contemporary Chinese

Sumin Guan, Benedikt Szmrecsanyi & Weiwei Zhang

KU Leuven

URL : https://sites.uclouvain.be/bkl-cbl/en/journals/papers-of-the-lsb/volume-18-2024/Guan-et-al

Electronic reference:

Guan, Sumin, Benedikt Szmrecsanyi & Weiwei Zhang (2024). A multifactorial analysis of Chinese analytic long passive constructions marked by *bèi, gěi* and *ràng* in contemporary Chinese. In Astrid De Wit, Tanja Mortelmans & Margot Vancauwenbergh (eds.), *Papers of the Linguistic Society of Belgium* [online] 18, 1-14.
DOI: https://doi.org/10.61430/DRHH1704

A multifactorial analysis of Chinese analytic long passive constructions marked by *bèi, gěi* and *ràng* in contemporary Chinese

Sumin Guan, Benedikt Szmrecsanyi, Weiwei Zhang

KU Leuven

Analytic passive constructions marked by *bèi*/*gěi*/*ràng* are the predominant strategy to indicate passive voice in Mandarin Chinese. While the Mandarin passive has been studied from various perspectives, it has not yet been analyzed using multifactorial methods to investigate its alternation. This study employs such methods to determine how language-internal factors predict the choice between *bèi*, *gěi*, and *ràng* long passives in Mandarin. Additionally, it examines whether the choice of variant differs between Mainland Chinese and Taiwan Chinese. To determine the combined effect of multiple linguistic factors, we use mixed-effects logistic regression based on a richly annotated dataset, following best practices in variationist (socio)linguistics. The results show that the marker *bèi* is preferred in atypical passive constructions. *Gěi* passives convey a more colloquial tone, while *ràng* passives inherit the attribute of avoidable events from *ràng* causatives, avoiding inanimate NP1s to prevent ambiguities. In comparison to Mainland Chinese, it is observed that in Taiwan Chinese, the typical passive marker *bèi* is more closely aligned with the old-before-new information structure paradigm and the traditional requirements of passive constructions in Mandarin Chinese. The marker *ràng* serves as a substitute for *bèi* when the context is positive. In contrast, Mainland Chinese is more innovative than Taiwan Chinese in the use of *gěi* passives.

1. Introduction

This paper focuses on Chinese long passive constructions marked by *bèi, gěi* and *ràng* (1-3), which are the three most used passive markers in contemporary Mandarin Chinese [1].

(1) marker=*bèi*

[这	名	弃	婴]$_{NP1}$	被	[福利院]$_{NP2}$	接	走	了
zhè	*míng*	*qì*	*yīng*	*bèi*	*fúlìyuàn*	*jiē*	*zǒu*	*le*
this	CLF	abandoned	baby	PASS	welfare	pick	RES	PST

'This abandoned baby was picked up by the welfare home.'
(zhTenTen (Simplified) chinadevelopment.com.cn)

(2) marker=*gěi*

[狐狸精]$_{NP1}$	给	[鬼]$_{NP2}$	捉	去	了
húlijīng	*gěi*	*gǔi*	*zūo*	*qù*	*le*
fox	PASS	ghost	capture	RES	PST

'The fox was captured by the ghost.'
(zhTenTen (Traditional) slime.com.tw)

(3) marker=*ràng*

[钱]$_{NP1}$	都	让	[老板]$_{NP2}$	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.'
(zhTenTen (Simplified) cye.com.cn)

The marker *bèi* has evolved from the verb "to suffer" to a fully grammaticalized passive marker. In contrast, *gěi* and *ràng* retain additional functions: *ràng* means "to allow" and is also a causative marker, while *gěi* means "to give", which can function as a focus marker in dative constructions. Within the realm of Chinese linguistics, research efforts have focused largely on individual passive variants and are grounded in pure semantic analysis through introspection (e.g., Zhu 1982; Li & Thompson 1989; Chappell 1983; Shi 2023), native speakers' acceptability judgements (e.g., Zan & Xu 1999), and small-scale descriptive corpus-based research (e.g., He 1989; Xiao, McEnery & Qian 2006; Song, Luo & Yu 2007). Only Guo and Chow (2014) employed predictive modelling to test the linguistic predictors for *bèi* passives diachronically, and only Zhang and Wang (2017) explored the alternation between *ràng* and *gěi* passive constructions using

[1] Long passives have an NP2, while short passives do not. Chappell & Shi (2016) discuss this distinction.

Multiple Correspondence Analysis (MCA). Therefore, the Chinese analytic passive alternation has yet to be fully explored using variational linguistics approaches to explore the probabilistic grammatical knowledge that language users' choices are based on.

Against this backdrop, this study probes the choice of *bèi*/*gěi*/*ràng* Chinese analytic long passive constructions and their variation across language varieties (Mainland Chinese and Taiwan Chinese). According to the framework of Probabilistic Grammar (Bresnan 2007; Bresnan and Ford 2010; Szmrecsanyi 2013) and variationist sociolinguistics (e.g. Tagliamonte 2012), these passives can be considered "alternate ways of saying 'the same' thing" (Labov 1973:188), and are subject to both language-internal syntactic and semantic constraints, as well as language-external constraints like language variety. Our study investigates the following research questions: 1) What are the language-internal and external factors that predict the choice among Chinese long passives marked by *bèi* vs. *gěi* vs. *ràng* in contemporary Chinese? 2) To what extent does each factor contribute to this choice? 3) Do these factors differ across varieties of Mandarin Chinese?

To answer these questions, we use both large-scale corpus data and multifactorial analysis, specifically, mixed-effects binary logistic regression, to provide the first comprehensive set of probabilistic constraints on this grammatical variation. We compare the different usage preferences for *bèi*/*gěi*/*ràng* long passive constructions and compare these preferences across two varieties of Mandarin Chinese.

This paper is structured as follows. Section 2 provides details of the corpus, data extraction procedures, the annotation scheme, and the analysis method. Section 3 reports the statistical modeling results. Section 4 discusses the significant stable and fluid constraints on the choice of *bèi*/*gěi*/*ràng* long passive constructions across language varieties. Section 5 presents some concluding remarks.

2. Data and Methods

2.1. *The Corpus*

To conduct our study, we tap into the Chinese Web Corpus 2017 from the TenTen corpus family (Vít 2021), which is a set of web corpora built using the same method with a target size of 10+ billion words and is made available through the Sketch Engine corpus manager. The Chinese Web Corpus 2017 (zhTenTen17) is a Chinese corpus composed of texts collected from the Internet in 2017. It is divided into two language varieties: the Chinese Simplified corpus with simplified Chinese

characters and the Chinese Traditional corpus with traditional Chinese characters. Based on the regional sources of data, each variety is further divided into four sub-corpora. Data was extracted from the largest sub-corpora for Mainland Chinese and Taiwan Chinese: Chinese domain.cn (41.5%) and Taiwan domain.tw (41.8%), respectively.

Table 1 displays the basic information of these two sub-corpora. The corpus metadata, including the source, title, URL, and website, can also be extracted along with concordances, and used for subsequent analysis.

Table 1: Information of two sub-corpora.

Sub-corpus	Tokens	Corpus	Variety
Chinese domain.cn	6,882,294,281	Chinese Web 2017 (zhTenTen17) Simplified	Mainland Chinese
Taiwan domain.tw	1,245,642,357	Chinese Web 2017 (zhTenTen17) Traditional	Taiwan Chinese

2.2. *Data extraction*

Given that the passive marker *ràng* can only be used in long passives, we have restricted the passives marked by both *bèi* and *gěi* to long passives as well (see footnote 1), thereby ensuring their interchangeability. Furthermore, tokens marked by *ràng* and *gěi* are included if these markers can be paraphrased by the fully grammaticalized passive marker *bèi* without changing the meaning in context.

The concordance search tool was used to display concordances of the target characters (*被 bèi*, *给/給 gěi* and *让/讓 ràng*) with their surrounding contexts (maximum of 100 preceding characters and 100 following characters) in the format of KWIC. Concordances underwent manual scrutiny to eliminate instances that did not meet the criteria for long passive constructions. The validated dataset for analysis is shown in Table 2.

Table 2: Distribution of Chinese *bèi*/*gěi*/*ràng* analytic long passive constructions across variety.

	bèi	*gěi*	*ràng*	Row total
Mainland Chinese	758	624	634	2,016
Taiwan Chinese	747	655	754	2,156
Column total	1,505	1,279	1,388	4,172

2.3. *Annotation for probabilistic constraints*

We manually annotated Chinese long passive constructions marked by *bèi, gěi* and *ràng* for various language-internal constraints that may influence the variant choice. For reasons of space, we provide one example (4) and present the corresponding coding results in Table 3. The complete annotation scheme with detailed exemplification is provided at https://osf.io/gjeu5.

(4) marker=*bèi*

[…] 被 复 建 起来。被 [层层 大 树]$_{NP2}$ 所 环绕
bèi fù jiàn qǐlái. bèi céngcéng dà shù suǒ huánrào
PASS again build RES PASS layers big tree AUX surround
的 [神宫]$_{NP1}$ 内 […]
de shéngōng nèi
REL the Sacred Palace inside
'…was built up again. Inside the Scared Palace surrounded by layers of trees…' (zhTenTen17 (Traditional) pt-travel.com.tw)

Table 3: Application of the annotation scheme to Chinese analytic passive constructions from example (4).

Variable	Value/level
ExpltNP1	explicit
DefNP1	definite
AnimacyNP1	inanimate
DefNP2	indefinite
AnimacyNP2	inanimate
RoleNP2	agent
PreSemantics	cause-motion
SynForm	single verb
Aspect	none
Object	no
ClauseType	attributive
Polarity	affirmative
SemProsody	neutral
SynParallel	yes
Auxiliary	*suo*
RLength	7
Variety	Taiwan
Verb	环绕 *huánrào* 'surround'
Website	pv-travel.com.tw

2.4. *Analysis methods*

The dataset was analyzed using multifactorial statistical methods. Initially, we used multinomial mixed-effects logistic regression, but there were two main drawbacks: firstly, its performance evaluation is limited to prediction accuracy and AUC (Area Under the ROC Curve) (Hand & Till (2001)[2]; Secondly, the interpretation and visualization of the results proved to be challenging. This is why we conducted mixed-effects binary logistic regression with the help of the `lme4` package (Bates et al. 2015) in R (R Core Team 2022).

Of the 19 probabilistic constraints, only "RLength" is a numerical predictor. It was therefore log-transformed and centered by dividing it by two standard deviations. This transformation was employed to ensure that its effect on the choice of variant is more comparable to that of binary categorical variables (Gelman 2008). To guarantee the dependability of all model parameters that include random effects, we pruned the "Verb" and "Website" variables by merging all levels with fewer than 5 observations (Clarke 2008) into a single level called *other*.

Following the top-down procedure outlined by Zuur et al. (2009), the model selection started with the maximum model structure containing language-internal and language-external constraints, the interaction of the "Variety" variable with each of the language-internal constraints, and two random effects, "Verb" and "Website". To reduce collinearity, we set the most frequent category as the reference level for all categorical constraints. To arrive at the final model, we first assessed the contribution of each random effect using likelihood ratio tests with the `anova()` function. Second, the `drop1()` function was employed to assess the structure of the fixed effects. Third, we identified the predictor with the highest *p*-value and removed it from the model. Fourth, we evaluated the final optimal model to assess its robustness. Finally, three optimal binary models (*bèi* vs. non-*bèi*, *gěi* vs. non-*gěi*, and *ràng* vs. non-*ràng*) were obtained for the choice of Chinese analytic passive constructions marked by *bèi*, *gěi* and *ràng*.

3. Results

For reasons of space, we summarize the model evaluation parameters of these three parallel binary mixed-effects logistic regressions in Table 4. A detailed report of statistical modeling results is available at https://osf.io/sv3ed. In summary, these

[2] We thank Prof. dr. Dirk Pijpops for bringing to our attention that AUC can serve as an additional evaluation metric for a multinomial model on the LSB Linguists' Day on 15 October 2023. We are grateful for his generous sharing of relevant papers and R codes.

three parallel models exhibit excellent fit, low multicollinearity (as evidenced in the low VIF scores), and high predictive accuracy. Of these models, the *bèi* vs. non-*bèi* model performs best with the highest concordance index *C*, $R^2_{mariginal}$ and $R^2_{conditional}$, and the classification accuracy.

Table 4: Evaluation parameters for three optimal binary mixed-effects logistic regressions.

	bèi vs. non-*bèi*	*gěi* vs. non-*gěi*	*ràng* vs. non-*ràng*
Concordance index C	0.937	0.859	0.867
$R2_{marginal}$	0.524	0.384	0.298
$R2_{conditional}$	0.700	0.507	0.516
Classification accuracy	87.58%	79.22%	80.25%
Condition index k	11.70	14.11	10.89
VIF scores	1.03–4.75	1.03–6.79	1.02–4.51

Regarding fixed effects, six factors remain consistent across language varieties and do not interact with "Variety" (referred to as "stable constraints"), including the animacy of NP1 and NP2, the definiteness of NP1, the semantics of the predicate, polarity, and syntactic parallelism. Meanwhile, eight factors show fluidity across varieties of Mandarin Chinese and interact with "Variety" (referred to as "fluid constraints"), including the definiteness of NP2, the syntactic form of the predicate, semantic prosody, clause type, the length of the right periphery of passive constructions, and the presence of aspect marker, auxiliary and object. Concerning random effects, "Verb" explains more variability in both the *bèi* vs. non-*bèi* and the *ràng* vs. non-*ràng* models, while "Website" explains more variability in both the *bèi* vs. non-*bèi* and the *gěi* vs. non-*gěi* models.

4. Discussion

This section compares the three parallel binary mixed-effects logistic regression models. Before providing a detailed interpretation of the results, it is important to acknowledge the number of significant main effects and interactions in each model. Each model identifies approximately 10 main effects (12 for *bèi* and non-*bèi*, 9 for *gěi* and non-*gěi*, and 11 for *ràng* and non-*ràng* exactly), and 5 interactions between language-internal factors and Variety, suggesting the complexity of linguistic conditioning in the context of Chinese long passive constructions. This is similar to the English passive alternation summarized by Bohmann et al. (2023), and provides further evidence for the critical role of multifactorial analysis in distinguishing the effects of competing conditioning factors. Section 4.1 discusses the stable constraints on the choice of the three passive constructions across language varieties. Based on these findings, we summarize the preferences for

different passive variants. Section 4.2 delves into fluid constraints and delineates the differing preferences for specific variants across different varieties of contemporary Chinese.

4.1. *Stable constraints across language varieties*

As demonstrated by Corrigan (1988), many verbs require an animate agent and an inanimate patient. Therefore, in passive constructions, an inanimate NP1 and an animate NP2 are typically required. Among the passive markers *bèi*, *gěi* and *ràng*, only *ràng* prefers an inanimate NP1, while *bèi* and *gěi* prefer an animate NP1 instead. For the animacy of NP2, *ràng* favors an animate NP2, while *bèi* is more likely when the NP2 is inanimate. Passive constructions typically assign the role of an agent (an animate actor of the event) to the NP2. However, the NP2 can also fulfill other thematic roles such as that of an instrument or a cause, both of which are inanimate and prefer *bèi* passives. To sum up, compared with *gěi* and *ràng*, the prototypical passive marker *bèi* is preferred when the passive construction is atypical, i.e., in passive constructions with an animate NP1 and an inanimate NP2.

Our results suggest that only the passive marker *ràng* follows the general tendency of the animacy distribution of two noun phrase slots in passive constructions, i.e., an inanimate NP1 and an animate NP2, as in (5). This distinguishes passive *ràng* from its function as a causative marker from which the passive usage has grammaticalized (Hashimoto 1988). The causative construction shares the same syntactic structure as the passive construction, with 'causer' and 'causee' occupying the NP1 and NP2 slots, respectively. Providing usage-based, quantitative evidence, Liesenfeld, Liu and Huang (2022) observed that causative *ràng* attracts human/animate noun phrases in the NP1 slot. Replacing the inanimate NP1 in (5) with an animate NP1, *主编 zhǔbiān*, 'editor-in-chief', for example, as in (6), can cause ambiguity of *ràng,* which can be interpreted as both a causative and a passive marker. Thus, animate NP1s may be avoided in *ràng* passives to prevent ambiguities.

(5) marker=*ràng*

[主编	的	用户名	和	密码]$_{NP1}$	不	能
zhǔbiān	*de*	*yònghùmíng*	*hé*	*mìmǎ.*	*bù*	*néng*
editor-in-chief	REL	username	and	password	NEG	can

让	[普通	编辑]$_{NP2}$	查	到
ràng	*pǔtōng*	*biānji*	*chá*	*dào*
PASS	general	editor	find	RES

'The username and password of editor-in-chief cannot be found by general editors.' (zhTenTen (Simplified) magtech.com.cn)

(6) [主编]$_{NP1}$ 不 能 让 [普通 编辑]$_{NP2}$ 查 到
zhǔbiān bù néng ràng pǔtōng biānji chá dào
editor-in-chief NEG can PASS/CAUS general editor find RES
'The editor-in-chief cannot let general editors find (sth).'

Chappell (1983) argued that the *ràng* passive construction and the *ràng* causative construction share the same characteristic of denoting "avoidable events." This implies that the NP1 has the ability to prevent the current event but did not take any action with that intention. Based on our findings, the preferences exhibited by *ràng* passives regarding the variables, "Semantics of Predicate" and "Polarity", also highlights this characteristic. The *ràng* passive construction is more likely when the predicate indicates cognition or perception. For polarity, the *ràng* passive construction is more likely when preceded by a negator and in interrogative sentences. For instance, in (7), the predicate *看扁* (*kànbiǎn*), meaning 'be looked down upon', signifies people's negative perception. This can be readily altered by NP1 to project a positive image in front of others. The negator *不 bù* underscores the proactive action taken by the NP1 to avoid being looked down upon.

(7) marker=*ràng*
[我]$_{NP1}$ 不 想 让 [人]$_{NP2}$ 看 扁
wǒ bù xiǎng ràng rén kàn biǎn
I NEG want PASS people look down
'I don't want to be looked down upon by people.' (zhTenTen (Traditional) tabf.org.tw)

The final significant stable constraint is the definiteness of NP1 in the *gěi* vs. non-*gěi* model, which reflects a more colloquial usage of the passive marker *gěi* to some extent. According to the result of the regression analysis, when NP1 is unclear, the *gěi* passive is more likely to be used, in cases where NP1 refers to a generic noun (or noun phrase), as in (8), or refers to the speaker, as in (9). The former always appears as a warning to the public, while the latter is usually found in dialogues, internal monologues, or direct comments on social media platforms. These texts have more characteristics of spoken rather than written language. To sum up, the preference for *gěi* passives for an unclear NP1 indicates a more colloquial use of *gěi* when compared to *bèi* and *ràng*.

(8) marker=*gěi*
小心 给 [人]$_{NP2}$ 骗 钱 了
xiǎoxīn gěi rén piàn qián le
cautious PASS people cheat money PST

'Watch out, (don't) get ripped off by others.' (zhTenTen (Simplified) qingdao315.com.cn)

(9) marker=*gěi*

"真	没	用!	快	给	[你]$_{NP2}$	吓	死	了
zhēn	*méi*	*yòng*	*kuài*	*gěi*	*nǐ*	*xià*	*sǐ*	*le*
really	NEG	useful	almost	PASS	you	frighten	death	PST

'"Useless! (I) was almost frightened to death by you."' (zhTenTen (Traditional) ezla.com.tw)

4.2. *Fluid constraints across language varieties*

Definiteness overlaps with the information status of constituents. Chao (1968) asserts that Chinese preverbal NPs (NP1) typically represent given information and are definite, while postverbal NPs (NP2) represent new information and are indefinite, which has also been suggested in other languages, such as English (Bresnan & Ford, 2010). According to Song, Luo and Yu (2007), the primary feature of information delivery in Mandarin *bèi* passives aligns with a common linguistic tendency, namely old-before-new information structure paradigm. However, the modelling results indicate that the definiteness of NP2 does not have a significant main effect in these three parallel models. Nevertheless, as an interaction term with variety, indefinite NP2 is revealed as a notable feature preferred by the long *bèi* passive in Taiwan Chinese, implying that the NP2 in the long *bèi* passive in Taiwan Chinese tends to prominently convey new information. Thus, compared to Mainland Chinese, Taiwan Chinese may adhere more closely to the old-before-new information structure paradigm with respect to the prototypical passive marker *bèi*.

According to Chao (1968), traditionally, the predicative verbs used in *bèi* passives are primarily limited to dispositional verbs. Wang (1957) explains that this dispositional notion is based on perceiving the action as complete. Consequently, an essential syntactic constraint commonly associated with *bèi* passives is their frequent co-occurrence with resultative verbal constructions (RVC) or perfective aspect markers such as *le* or *guo.* In contrast, our findings show that *bèi* is preferred with a single verb when compared with *gěi* and *ràng*, and it is *gěi* that significantly prefers a complex verb or a verb with a perfective aspect marker. Nevertheless, regarding interaction between aspect markers and variety, the perfective aspect markers *le*/*guo* are more likely to co-occur with *bèi* passive constructions in Taiwan Chinese than in Mainland Chinese. Moving to semantic prosody, Xing (2004) interprets the loss of source meaning and semantic broadening of *bèi* passives as manifestations of grammaticalization. Our findings also suggest that the adversity effect of *bèi* passives is less pronounced than before,

as *bèi* is now more commonly used in non-negative context. However, in Taiwan Chinese, the *bèi* passive is almost seven times less likely in positive context than in Mainland Chinese. Conversely, it seems that the *ràng* passives are 2.5 times more likely to be used in positive contexts in Taiwan Chinese than in Mainland Chinese. It appears from these interactions that Taiwan Chinese also conforms more closely to the traditional semantic requirement of Mandarin Chinese.

The rise of *bèi* passives functioning as attributives has been noted by both Guo and Chow (2014) and Xiao et al. (2006). While both *gěi* and *ràng* passives tend to discourage attributive usage, the *bèi* passive is more likely to be chosen when the construction functions as an attributive clause instead of a main clause or another clause type. The tendencies for both *bèi* passives and *gěi* passives are further pronounced in Taiwan Chinese where the likelihood of choosing *bèi* passives as attributive clauses is significantly higher, while the likelihood of choosing *gěi* passives as attributive clauses is significantly lower. Regarding another predictor, "Auxiliary", both *gěi* and *suǒ* are found to be used in all three passive constructions. There is a notable difference between varieties in the use of the auxiliary *gěi*: it is more likely to appear in *bèi* passives in Taiwan Chinese than in Mainland Chinese. Conversely, in the context of *gěi* passives, the use of the auxiliary *gěi* is significantly less frequent in Taiwan Chinese compared to Mainland Chinese. Thus, regarding the use of the passive marker *gěi*, Mainland Chinese appears to be more innovative than Taiwan Chinese. This is evident in its significant preference for using *gěi* passives as attributive clauses and with the auxiliary *gěi* in the *gěi* passives compared to Taiwan Chinese.

5. Conclusion

The study reveals that *bèi* is predominantly used in atypical passives, while *gěi* passives convey a more colloquial tone. The marker *ràng*, linked to avoidable events, is less frequently used with inanimate NP1s to avoid ambiguity. When comparing between language varieties, Taiwan Chinese aligns more closely with the old-before-new information structure paradigm and adheres to the traditional conventions of passive constructions in Mandarin Chinese. In Taiwan Chinese, *ràng* is more likely than *bèi* in positive contexts. Conversely, Mainland Chinese demonstrates greater innovation with *gěi* passives. Our findings support prior research on the colloquial nature of *gěi* and the avoidable events indicated by *ràng*. This study uniquely captures passive marker choice in atypical constructions, focusing on NP1 and NP2 animacy. As the first multifactorial analysis of regional passive construction variation, our insights suggest significant regional influences on linguistic choice, an area less explored in earlier research.

Future research should conduct a large-scale, multifactorial diachronic study of Chinese long passive constructions to validate our conclusions. Additionally, since *bèi* and *gěi* are also used in short passives, a multifactorial analysis of both long and short constructions would clarify their overall usage as passive markers.

6. References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). 'Fitting Linear Mixed-Effects Models Using lme4'. *Journal of Statistical Software*, *67*(1), 1–48.

Bohmann, A., Müller, J., Honkanen, M., & Neuhausen, M. (2023). 'A Large-scale Diachronic Analysis of the English Passive Alternation'. In B. Busse, N. Dumrukcic, & I. Kleiber, **eds.,** *Language and Linguistics in a Complex World* (Vol. 32). Berlin, Boston: De Gruyter. 31–56.

Bresnan, J. (2007). 'Is syntactic knowledge probabilistic? Experiments with the English dative alternation'. In S. Featherston & W. Sternefeld **eds.,** *Roots: Linguistics in Search of its Evidential Base,* Berlin, New York: Mouton de Gruyter. 75–96.

Bresnan, J., & Ford, M. (2010). 'Predicting syntax: Processing dative constructions in American and Australian varieties of English'. *Language, 86*(1), 168–213.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, Calif: University of California Press.

Chappell, H. (1983). *A semantic analysis of passive, causative and dative constructions in standard Chinese*. Canberra: Australian National University.

Chappell, H., & Shi, D. (2016). 'Major non-canonical clause types: Ba and bei'. In C.-R. Huang & D. Shi, **eds.,** *A Reference Grammar of Chinese*. Cambridge University Press. 451–483.

Chen, L. (2002). '*"被"和"让"表被动的句法语义对立及其功能解释 [Syntactic distribution of Bei and Rang in Chinese passive construction and its Functional Interpretation]*' [Nankai University].

Clarke, P. (2008). 'When can group level clustering be ignored? Multilevel models versus single-level models with sparse data" *Journal of Epidemiology & Community Health*, *62*(8), 752–758.

Corrigan, R. (1988). 'Children's identification of actors and patients in prototypical and nonprototypical sentence types'. *Cognitive Development*, *3*(3), 285–297.

Gelman, A. (2008). 'Scaling regression inputs by dividing by two standard deviations'. *Statistics in Medicine*, *27*(15), 2865–2873.

Guo, H., & Chow, D. (2014). 'A corpus-based variationist approach to bei passives in Mandarin Chinese'. *Corpus Linguistics and Linguistic Theory*, *10*(1), 139–173.

Hand, D. J., & Till, R. J. (2001). 'A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems'. *Machine Learning*, *45*(2), 171–186.

Hashimoto, M. J. (1988). 'The structure and typology of the Chinese passive construction'. In M. Shibatani **ed.,** *Passive and Voice* 16, Amsterdam: John Benjamins Publishing Company. 329–354.

He, X. (1989). '*论曹禺《雷雨》《日出》《北京人》中的"被"字句* [Bei passives in Caoyu's Leiyu, Richu, Beijing Ren]'. *Journal of Sichuan Normal University (Philosophy and Social Sciences)*, *6*, 69–79.

Joan Bresnan & Marilyn Ford. (2010). 'Predicting syntax: Processing dative constructions in American and Australian varieties of English'. *Language*, *86*(1), 168–213.

Labov, W. (1973). *Sociolinguistic Patterns* (Vol. 10). Philadelphia: University of Pennsylvania Press.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: CA University of California Press.

Liesenfeld, A., Liu, M., & Huang, C.-R. (2022). 'Profiling the Chinese causative construction with rang (*讓*), shi (*使*) and ling (*令*) using frame semantic features'. *Corpus Linguistics and Linguistic Theory*, *18*(2), 263–306.

R Core Team. (2022). 'R: A Language and Environment for Statistical Computing'. *R Foundation for Statistical Computing, Vienna, Austria*. https://www.R-project.org/

Shi, Y. (2023). *The Evolution of Chinese Grammar* (1st ed.). Cambridge: Cambridge University Press.

Song, W., Luo, Z., & Yu, J. (2007). '*现代汉语被动句施事隐现的计量分析* [A quantitative analysis of the occurrence ratio of agent in contemporary Chinese passive constructions]'. *Zhongguo yu wen*, *2*, 113-124+191.

Szmrecsanyi, B. (2013). 'Diachronic Probabilistic Grammar'. *English Language and Linguistics*, *19*, 41–68.

Tagliamonte, S. A. (2012). *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden & Oxford: Wiley-Blackwell.

Thompson, D., & Scheepers, C. (2013). 'Harmonizing the passive: A new proposal for passive constructions in generative grammar'. *Newcastle Working Papers in Linguistics*, *19*, 74–96.

Vít, S. (2021). 'Genre Annotation of Web Corpora: Scheme and Issues'. In K. Arai, S. Kapoor, & R. Bhatia, **eds.**, *Proceedings of the Future Technologies Conference (FTC) 2020. Volume 1* (Vol. 1). Cham: Springer. 738–754.

Wang, L. (1957). '*汉语被动式的发展* [The development of Chinese passive constructions]'. *Yuyanxue Luncong*, *1*, 1–16.

Xiao, R., McEnery, T., & Qian, Y. (2006). 'Passive constructions in English and Chinese: A corpus-based contrastive study'. *Languages in Contrast*, *6*(1), 109–149.

Xing, F. (2004). 'On the favor-accepting Type of Passive Sentences with the Marker "bei."' *Studies in Language and Linguistics*, *1*, 1–11.

Zan, J., & Xu, Y. (1999). '*关于汉语被动句"被、叫、让、给" 的互换性* [About the interchangeability of Chinese passive sentences marked by "bei, jiao, rang, gei"]'. In *Selected papers of the third International Symposium on Chinese Language Teaching,* 427–437.

Zhang, W. (2007). '"*被，给，叫，让"的意愿性与其宾语从缺的关系* [The volition of "bei, gei, jiao, rang" and its relation with the absence of its object]'. *Hanzi Wenhua*, *01*, 45–48.

Zhang, W., & Wang, F. (2017). '*从基于样例的概念空间看构式交替——以"让"和"给"*

的被动用法为例 [An exemplar-based conceptual space of Chinese passives with "rang" & "gei"]'. *Foreign Language Teaching and Research*, *6*, 22-23+145.

Zhu, D. (1982). *语法讲义 [Yufa Jiangyi]*. Beijing: The Commercial Press.

Zuur, A., Ieno, E. N., Walker, N., Saveliev, A. A., & Smith, G. M. (2009). *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer Science & Business Media.