A MULTIFACTORIAL ANALYSIS OF CHINESE LONG PASSIVE CONSTRUCTIONS MARKED BY BEI, GEI AND RANG IN CONTEMPORARY CHINESE

Sumin Guan Benedikt Szmrecsanyi Weiwei Zhang
KU Leuven, Belgium

ABSTRACT

Passive constructions marked by *bèi*/*gěi*/*ràng* are the predominant strategy to indicate passive voice in Mandarin Chinese. These passive variants are in principle well-studied, but the alternation has yet to be fully analyzed using variationist techniques to explore the probabilistic grammatical knowledge that language users have about the long passives. To explore the variation in contemporary Chinese, we tap into the *Chinese Web Corpus* (zhTenTen17). The objective is to identify what the language-internal factors are that predict the choice between Chinese *bèi*/*gěi*/*ràng* long passives and to establish if variant choice differs as a function of language variety. To determine the combined effect of multiple linguistic factors, mixed-effects logistic regression is used, following best practice in variationist (socio)linguistics. The results reveal that in atypical passive constructions, the marker *bèi* is preferred. Both *gěi* and *ràng* passives convey a more colloquial tone, while *ràng* passives inherit the attribute of avoidable events from *ràng* causatives, favoring an inanimate NP1 to avoid ambiguity. Compared to Mainland Chinese, we show that in Taiwan Chinese, the typical passive marker *bèi* aligns more closely with old-before-new information structure and is favored by adversative contexts; *ràng* serves as a substitute for *bèi* when the semantic prosody is positive, whereas the usage of *gěi* is influenced by the use of *hoo* in Taiwanese and also conditioned by the cognitive complexity principle. This analysis confirms the complexity of language-internal and external linguistic conditioning on Chinese long passive constructions and paves the way for further research in Chinese passive constructions.

KEYWORDS
Long Passives **R**àng Passives **G**ěi Passives
Variationist (socio)linguistics **M**ultifactorial analysis

1. INTRODUCTION

In many languages, the passive voice is a construction that offers speakers an alternative to the more commonly used active form (Thompson, Ferreira, and Scheepers 2018). In Mandarin Chinese, the most commonly used passives are passive constructions marked by *bèi* 'suffer', *gěi* 'give', *ràng* 'let, make', and *jiào* 'tell, make'[1]. Except for *bèi*, the other three serve additional functions beyond indicating passive voice. For example, *ràng* and *jiào* can function as both lexical verbs and causative markers (Jiang 2000; Jiang 2002), while *gěi* can act as a verb, a dative marker, a goal marker, a causative marker (Lin and Huang 2015), and even as an affectedness marker (Tang 2001), to reinforce the affectedness or disposal meaning of the predicate.

Passive constructions involve placing the nominal phrase bearing the prototypical patient role (NP1) in the subject position, while the prototypical agent (NP2) either appears in a position marked by the passive marker or is omitted. It is important to distinguish between long passives and short passives (Ting 1998; Feng 1997; Huang 1999; Tang 2004; Shi 2008; Huang, Li, and Li 2009; Feng 2012; Feng 2013). The passive construction with NP2 is referred to as a "long passive", while the one without NP2 is called a "short passive". Passive constructions are typically formed with *bèi* (e.g., Hashimoto 1988; Huang 1999; Huang et al. 2009), and in contemporary Chinese, a sizable portion of *bèi* passives is formed with an implicit NP2 (Xiao, McEnery, and Qian 2006; Guo and Han 2012; Guo and Chow 2014). The more colloquial variants with *jiào* and *ràng*, however, must introduce an NP2 to make the passive construction well-formed (Liu, Pan and Gu 2001, 760; Shi 2023, 175−202). Corpus-based analyses by Xiao et al. (2006), Song, Luo and Yu (2007), Yang (2018), and Wang (2022) demonstrate that all *jiào* and *ràng* passives are long passives, except for one instance of *ràng* found in fiction texts in Song et al. (2007). Nevertheless, when considering frequency patterns, the corpus used in this study shows that passives in *jiào* concordances account for only 0.114%.

To address issues related to interchangeability and data sparsity, we have opted to exclude the marker *jiào* and focus exclusively on long passives marked by *bèi*/*gěi*/*ràng* (1-3)[2].

(1) zhTenTen17 (Simplified) marker=*bèi*
chinadevelopment.com.cn

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

'This abandoned baby was picked up by the welfare home.'

(2) zhTenTen17 (Traditional) slime.com.tw marker=*gěi*

[狐狸精]$_{NP1}$	**给**	[鬼]$_{NP2}$	捉	去	了。
húlijīng	***gěi***	*guǐ*	*zhūo*	*qù*	*le*
fox	PASS	ghost	capture	RES	PFV

'The fox was captured by the ghost.'

(3) zhTenTen17 (Simplified) cye.com.cn 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	PFV

'All the money was earned by the boss.'

The alternation between English *be*/*get* passives has been studied thoroughly (e.g., Collins 1996; Carter and McCarthy 1999; Thompson and Scheepers 2013; Fehringer 2022; Bohmann et al. 2023), and factors such as transitivity, affectedness, adversity, patient responsibility, subject and object animacy, and aspectuality have been identified to predict the use of *be*/*get*. However, within the realm of Chinese linguistics, research efforts have focused largely on the evolution of passive constructions (e.g., Bennett 1981; Peyraube 1989; Li 2007; Chen and Lien 2011) and specific variants of passive constructions (e.g., Li and Thompson 1989; Chappell 1983; Xiao et al. 2006; Song et al. 2007; Guo and Chow 2014). These studies are grounded in pure semantic analysis through introspection and small-scale descriptive corpus-based research. Only Zhang and Wang

(2017) explored the alternation between *ràng* and *gěi* passive constructions using multifactorial analysis. Therefore, the Chinese passive alternation has yet to be fully explored using variational linguistics approaches to explore the probabilistic grammatical knowledge that language users have.

Against this backdrop, this study probes the choice of *bèi*/*gěi*/*ràng* long passive constructions and their language variation. Working within the framework of Probabilistic Grammar (Bresnan 2007; Bresnan and Ford 2010; Szmrecsanyi 2013) and variationist (socio)linguistics (e.g., Tagliamonte 2012), we assume that passive variants as "alternate ways of saying 'the same' thing" (Labov 1972, 188) are subject to both language-internal syntactic and semantic constraints, as well as language-external constraints such as variety. Our study investigates the following research questions:

- 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?
- To what extent does each factor contribute to this choice?
- Are there differences between varieties of Mandarin Chinese, i.e., Mainland Chinese and Taiwan Chinese? If so, to what extent and in what direction?

To answer these questions, we use both large-scale corpus data and multifactorial analysis, specifically, mixed-effects binary logistic regression, to test a comprehensive set of probabilistic constraints on this grammatical variation. Regression modeling reveals that the alternation of long passives is influenced by a combination of semantic and syntactic factors. Six factors remain consistent across varieties of Mandarin Chinese, delineating the different usages of these three variants, while eight significant factors exhibit fluidity (i.e., behave differently) across varieties of Mandarin Chinese.

This paper is structured as follows. In Section 2, we provide details of the corpus we used and explain the definition of the variable context, data extraction procedures, annotation scheme of constraints, and the analysis methods. Section 3 reports the statistical results of three parallel binary models. In Section 4, we discuss and interpret our key findings. 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 (zhTenTen17) from the TenTen corpus family (Vít 2021) in the Sketch Engine corpus manager. The zhTenTen17 is a Chinese corpus composed of texts collected from the Internet and is divided into two language varieties: the Chinese Simplified corpus and the Chinese Traditional corpus. 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).

Table 1 Sub-corpora under analysis

Sub-corpus	Tokens	Corpus	Varieties
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 Definition of the Variable Context and Data Extraction

As observations marked by *bèi* are passive constructions without any doubt, we consider *gěi* and *ràng* observations as interchangeable passive constructions if these two markers can be paraphrased by *bèi* without any semantic or functional change, taking the context into consideration.

The concordance search tool of Sketch Engine 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 KWIC format. Concordances underwent manual scrutiny to eliminate instances that did not meet the criteria of long passive constructions or that were potentially controversial as non-passive occurrences. To mitigate subjectivity and to assess the objectivity of manual filtering, we conducted an inter-rater reliability analysis using a dataset of 100 *gěi* and *ràng* observations. A second coder – a Chinese

native speaker and trained linguist – was given these observations in random order and instructed to code them independently with the question "Can the marker be paraphrased by *bèi*?" After coding, using the irr package (Gamer, Lemon, and Singh 2019) in R (R Core Team 2022), we obtained a Cohen's kappa value (Levshina 2015, 201) of 0.92, indicating an almost perfect level of agreement.

Given that *bèi* is the prototypical passive marker in Mandarin Chinese, its frequency far exceeds that of *gěi* and *ràng*. To ensure statistical validity without bias, we intentionally opted for a well-balanced dataset for analyzing passive variant choice. To maintain comparability, we first determined target counts based on the smaller Taiwan sub-corpus, extracting 655 long *gěi* and 754 long *ràng* passives. These targets guided the sampling in the Mainland sub-corpus, where 624 long *gěi* and 634 long *ràng* passives were obtained after removing duplicates and ambiguous tokens. For *bèi* passives, a pilot study indicated that long *bèi* passives constitute approximately 30% of occurrences in the zhTenTen17 corpus. Thus, we randomly selected 2,500 *bèi* concordances from both sub-corpora, yielding 758 and 747 valid tokens for Mainland Chinese and Taiwan Chinese, respectively. The final dataset is summarized in Table 2, and the dataset is available at https://osf.io/mc3h2/files/osfstorage.

Table 2 Tokens of Chinese *bèi/gěi/ràng* long passives across varieties

	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 for various language-internal constraints plus language variety (Mainland vs. Taiwan Chinese). This subsection outlines the included predictors. For details on predictor selection and examples, see the annotation scheme in the supplementary materials at https://osf.io/mc3h2/files/osfstorage.

- Explicitness of NP1 (label: ExpltNP1): *explicit*, *implicit*.
- Definiteness of NP1 (label: DefNP1): *definite*, *indefinite*, *unclear*.
- Definiteness of NP2 (label: DefNP2): *definite*, *indefinite*.

- Animacy of NP1 and NP2 (labels: AnimacyNP1 and AnimacyNP2): *animate* (humans, animals), *inanimate* (collective, temporal, locative, event/action, plants).
- Semantic roles of NP2 (label: RoleNP2): *agent*, *others* (experiencer, instrument, cause).
- Syntactic forms of predicate (label: SynForm): *single verb* (bare verbs, verb+preposition), *result verb* (resultative verb constructions, verb+*le/zhe/guo*)[3], *complex verb* (verb+*de*, coordination verbs and serial verb constructions).
- Semantics of Predicate (label: PreSemantics): Following Mandarin VerbNet (Liu and Chiang, 2008), the annotation distinguishes between three semantic categories as follows: *action* (cause-motion), *interaction* (social-interaction, communication, emotion), *cog-perc* (cognition and perception).
- Aspect marker (label: Aspect): *PFV* (perfective markers including *le*, *guo* and any combination of these two markers), *PROG* (the progressive/imperfective marker *zhe*), *none* (no aspect marker in the construction).
- Retained object (label: Object): *yes*, *no*.
- Clause types (label: ClauseType): *main*, *attributive*, *others*.
- Polarity (label: Polarity): *affirmative*, *negative*, *interrogative*.
- Semantic Prosody (label: SemProsody): *positive*, *neutral*, negative.
- Auxiliary (label: Auxiliary): *gei, suo*, *none*.
- Syntactic Parallelism (label: SynParallel): The occurrence of passive constructions before the target marker. The annotation distinguishes between *yes* (there is another occurrence of the same marker in the preceding context) and *no* (a different passive marker or no passive marker is present in the preceding context).
- Length of the right periphery (label: RLength): the number of characters following the passive marker until the end of the clause.
- Verb lemma (label: Verb): the verb lemma of the predicate.
- Website (label: Website): The source of the token.
- Language Variety (label: Variety): *Mainland Chinese*, *Taiwan Chinese*.

To illustrate this annotation scheme, consider the example (4). The coding results are depicted in Table 3.

(4) zhTenTen17(Traditional) game2.tw marker=*bèi*

[我]$_{NP1}$	**被**	[对手]$_{NP2}$	骗	了。
wǒ	***bèi***	*duìshǒu*	*piàn*	*le*
I	PASS	opponent	cheat	PFV

'I was cheated by the opponent.'

Table 3 Application of the annotation scheme to example (4)

Variable	Value
ExpltNP1	explicit
DefNP1	definite
AnimacyNP1	animate
DefNP2	definite
AnimacyNP2	animate
RoleNP2	agent
PreSemantics	cog-perc
SynForm	result verb
Aspect	PFV
Object	no
ClauseType	main
Polarity	affirmative
SemProsody	negative
SynParallel	no
Auxiliary	none
RLength	4
Variety	Taiwan Chinese
Verb	骗 *piàn* 'cheat'
Website	game2.tw

2.4 Analysis Methods

The dataset was analyzed using multifactorial statistical methods. In variationist linguistics, binary logistic regression serves as the workhorse analysis technique and quantifies the simultaneous effect of multiple individual probabilistic constraints on a binary response variable, such as the choice between *bèi* and non-*bèi*. The mixed-effects models also include random effects, which increase the generalizability of the results by accounting for dependent observations and irrelevant sources of variation (Gries 2015).

Following the top-down procedure outlined by Zuur et al. (2009), the model selection started with the maximum model structure containing all annotated constraints, the interaction of language variety with each of the language-internal constraints, and the 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. 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, starting with non-significant interaction terms before moving on to significant interaction terms and main effects. Fourth, we evaluated the final optimal model to assess its robustness. The R script is available at https://osf.io/mc3h2/files/osfstorage.

3. RESULTS

In what follows, we fit three independent logistic regression models, each of which models choices between one of the variants and the other variants. Table 4 presents an evaluation of the three resulting minimal adequate regression models, each demonstrating excellent discrimination ($C \geq 0.8$; Levshina 2015, 259) and classification accuracy significantly exceeding baselines. All models are robust, with condition indices ($\kappa < 30$, Baayen 2008, 182) and variance inflation factors (VIFs < 10, Levshina 2015, 272) indicating no multicollinearity issues. The variance explained by fixed effects ($R^2_{marginal}$) and by both fixed and random effects ($R^2_{conditional}$) further reinforces the models' predictive strength (Baayen 2008, 204). Verb (N = 153) is the more important random effect in both the *bèi* vs. non-*bèi* and *ràng* vs. non-*ràng* models*, while* Website (N = 122) accounts for more variability in the *gěi* vs. non-*gěi* model.

Table 4 Evaluation of three parallel minimal adequate regression models

Regression models	*bèi* vs. non-*bèi*	*gěi* vs. non-*gěi*	*ràng* vs. non-*ràng*
concordance index *C*	0.937	0.859	0.867
classification accuracy (baseline)	87.6% (63.9%)	79.2% (69.3%)	80.2% (66.7%)
Index κ	14.67	17.52	12.29
VIFs	1.03–8.18	1.03–9.62	1.02–8.32
$R^2_{marginal}$	0.524	0.384	0.302
$R^2_{conditional}$	0.700	0.507	0.516
Verb Variance	1.461	0.318	1.307
Website Variance	0.463	0.502	0.174

3.1 *Bèi* vs. Non-*Bèi*

Table 5 presents the fixed effects in this model. Out of all the main effects, only eight probabilistic constraints do not interact with Variety. These constraints affect the choice of *bèi* and non-*bèi* in the same way for both Mainland Chinese and Taiwan Chinese. Compared to *gěi* and *ràng* passives, the *bèi* passive is preferred when there is an animate NP1, an inanimate NP2, a single verb as the predicate, another *bèi* marker in the preceding context, and a longer right periphery. On the other hand, the *bèi* passive is less likely to be used when the predicate indicates a state/change in cognition or perception, including a retained object, or comes in a non-affirmative sentence.

The model's remaining probabilistic constraints interact with Variety. Thus, the statistics relating to those significant interactions suggest that the corresponding change in predictor levels affects the choice of passive markers differently for Mainland Chinese and Taiwan Chinese. Overall, in comparison to Mainland Chinese, the *bèi* passive in Taiwan Chinese is favored with an indefinite NP2, a PFV aspect marker, an auxiliary *gěi* and when used as an attributive clause, while the *bèi* passive is disfavored when the semantic prosody is positive or neutral.

Table 5 Fixed effects of the individual predictors in the minimal adequate regression model of *bèi* vs. non-*bèi*. The reference level of predictors is given between brackets. Predictors shared across the three models are **bolded**. The predicted response level is the *bèi* passive. Positive coefficients suggest an increase in the likelihood of *bèi*, while negative coefficients suggest a decrease in the likelihood of *bèi*.

Predictor	Coefficient
(Intercept)	-1.511***
AnimacyNP1 (inanimate)	0.549***
AnimacyNP2 (animate)	0.877***
DefNP2 (definite)	-0.245
PreSemantics (action)	
cog_perc	-1.101***
interaction	0.151
SynForm (result verb)	
single verb	0.847***
complex verb	-0.022
Aspect (none)	
PFV	-2.255***
PROG	0.461
ClasueType (main)	
attributive	1.322***

others	-0.019
Object (no)	-0.519**
SynParallel (no)	1.706***
Polarity (affirmative)	
interrogative	-1.302***
negative	-1.507***
SemProsody (negative)	
neutral	0.840***
positive	0.923***
Auxiliary (none)	
gěi	-2.640***
suǒ	-0.159
RLength (logged)	2.124***
Variety (Mainland)	0.032
Interactions	
DefNP2:Variety	
definite + Taiwan	0.032
indefinite + Taiwan	0.838***
Aspect:Variety	
PFV + Taiwan	1.153***
PROG + Taiwan	-1.059
ClauseType:Variety	
attributive + Taiwan	1.053*
others + Taiwan	-0.007
SemProsody:Variety	
neutral + Taiwan	-0.696**
positive + Taiwan	-1.863***
Auxiliary:Variety	
gěi + Taiwan	1.961*
suǒ + Taiwan	0.327

Significance codes: "***": $p < 0.001$; "**": $p < 0.01$; "*": $p < 0.05$

Regarding the intercept adjustments for Verb extracted from the model, we filtered the top ten verbs that strongly prefer the *bèi* passive and strongly disfavor the *bèi* passive, respectively, as shown in Figure 1. The top ten verbs that prefer *bèi* can be classified into 3 types. The first type is the favor-accepting verb summarized by Xing (2004), including 当 *dāng* 'be taken as', 称 *chēng* 'be named as', 封 *fēng* 'be recognized as' and 视 *shì* 'be treated as'; the second type is the adversative strong action verb, including 杀害 *shāhài* 'kill', 拦 *lán* 'stop' and 击 *jī* 'hit'; the third type is the most frequently used perception or cognition verb, including 盯 *dīng* 'stare', 发现 *fāxiàn* 'discover' and 知 *zhī* 'know'.

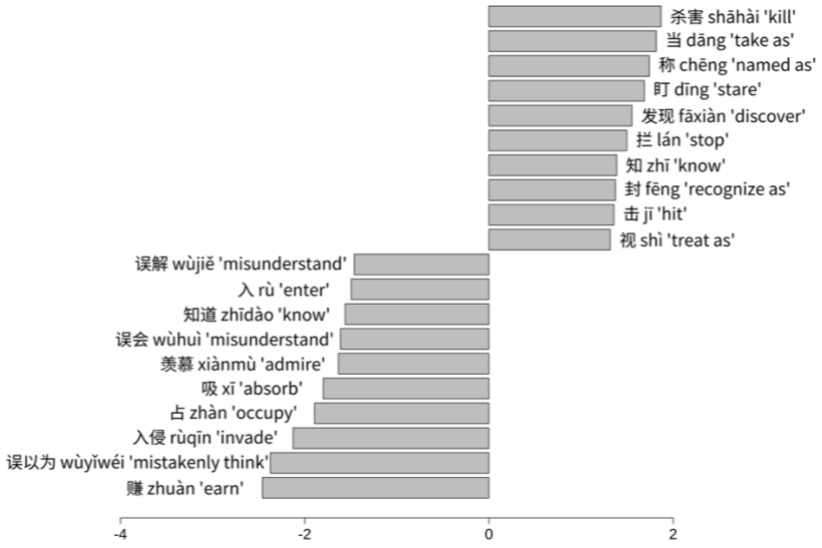

Figure 1 The top ten verbs strongly preferring the *bèi* passive (on the right side of the axis with values above 0) and the top ten verbs strongly disfavoring the *bèi* passive (on the left side of the axis with values below 0).

3.2 *Gěi* vs. Non-*Gěi*

The main effects in Table 6 show that, compared with *bèi* and *ràng* passives, the *gěi* passive is favored when used with an animate NP1 whose definiteness is unclear (when referring to a generic noun or the speaker). However, it is disfavored in contexts that have a non-action predicate, include a retained object, feature another passive marker *gěi* in the preceding context, possess non-negative semantic prosody, or have a longer right periphery. The interactions reveal that compared to Mainland Chinese, in Taiwan Chinese, the *gěi* passive is favored with a definite NP2, while the *gěi* passive is less likely to be used with a PFV aspect marker, a complex verb as the predicate, an auxiliary *gěi* or when the passive construction is not used in a main clause.

Table 6 Fixed effects of the individual predictors in the minimal adequate regression model of *gěi* vs. non-*gěi*. The reference level of predictors is given between brackets. Predictors shared across the three models are

bolded. The predicted response level is the *gěi* passive. Positive coefficients suggest an increase in the likelihood of *gěi*, while negative coefficients suggest a decrease in the likelihood of *gěi.*

Predictor	Coefficient
(Intercept)	-0.530*
AnimacyNP1 (inanimate)	0.202*
DefNP1 (definite)	
indefinite	0.127
unclear	0.384**
DefNP2 (definite)	0.076
PreSemantics (action)	
cog_perc	-0.744***
interaction	-0.379**
SynForm (result verb)	
single verb	-0.721***
complex verb	0.649*
Aspect (none)	
PFV	1.545***
PROG	-0.051
ClasueType (main)	
attributive	-0.497
others	0.332
Object (no)	-0.551***
SynParallel (no)	-0.655***
SemProsody (negative)	
neutral	-0.332**
positive	-0.920***
Auxiliary (none)	
gěi	-0.156
suǒ	0.091
RLength (logged)	-1.553***
Variety (Mainland)	0.571**
Interactions	
DefNP2:Variety	
definite + Taiwan	0.571**
indefinite + Taiwan	-0.100
Aspect:Variety	
PFV + Taiwan	-0.803***
PROG + Taiwan	0.251
SynForm:Variety	
single verb + Taiwan	-0.127
complex verb + Taiwan	-1.259***
ClauseType:Variety	
attributive + Taiwan	-1.181*
others + Taiwan	-0.491*
Auxiliary:Variety	
gěi + Taiwan	-1.202**
suǒ + Taiwan	0.382

Significance codes: "***": $p < 0.001$; "**": $p < 0.01$; "*": $p < 0.05$

When considering the intercept adjustments for Verb and Website, it is notable that the by-item adjustment for Verb displays lower variability. Out of all 153 verbs, only the verb 赚 *zhuàn* 'earn' exhibits a strong preference for *gěi*. Conversely, the by-subject adjustment for Website, which accounts for higher variability, reveals four websites with a strong preference for *gěi*, and all of these are from Taiwan. The following four websites are ranked in descending order of intercept adjustments: *lama.com.tw*, *nch.com.tw*, *adj.idv.tw*, and *coolloud.org.tw*. The texts extracted from these websites include sermons, fiction, blogs, and instant reports on social movements, respectively.

3.3 *Ràng* vs. Non-*Ràng*

Table 7 presents the fixed effects. Eight probabilistic constraints show no interaction with Variety. Compared to *bèi* and *gěi* passives, the *ràng* passive is favored with a predicate denoting a state/change of cognition or perception, an auxiliary *gěi*, and in non-affirmative sentences, while it is disfavored with an animate NP1 whose definiteness is unclear, an inanimate NP2, the construction being used as an attributive clause, another passive marker *ràng* in the preceding context, and a longer right periphery. The interactions show that compared to Mainland Chinese, the *ràng* passive is favored in Taiwan Chinese with a complex verb as predicate, a positive semantic prosody, and a longer right periphery, while it is less likely with an implicit NP2 and a retained object.

Table 7 Fixed effects of the individual predictors in the minimal adequate regression model of *ràng* vs. non-*ràng*. The reference level of predictors is given between brackets. Predictors shared across the three models are **bolded**. The predicted response level is the *ràng* passive. Positive coefficients suggest an increase in the likelihood of *ràng*, while negative coefficients suggest a decrease in the likelihood of *ràng*.

Predictor	Coefficient
(Intercept)	-1.426***
ExpltNP1 (explicit)	0.549***
AnimacyNP1 (inanimate)	-0.639***
AnimacyNP2 (animate)	-0.623***
DefNP1 (definite)	
indefinite	0.086
unclear	-0.584***
PreSemantics (action)	

cog_perc	1.506***
interaction	0.116
SynForm (result verb)	
single verb	-0.304*
complex verb	-0.358
ClasueType (main)	
attributive	-1.056***
others	0.025
Object (no)	1.298***
SynParallel (no)	-1.628***
Polarity (affirmative)	
interrogative	0.648***
negative	1.126***
SemProsody (negative)	
neutral	-0.053
positive	-0.053
Auxiliary (none)	
gěi	1.669***
suǒ	-0.319
RLength (logged)	-0.926***
Variety (Mainland)	-0.104
Interactions	
ExpltNP1:Variety	
explicit + Taiwan	-0.151
implicit + Taiwan	-0.539**
SynForm:Variety	
single verb + Taiwan	0.169
complex verb + Taiwan	0.769*
SemProsody:Variety	
neutral + Taiwan	-0.026
positive + Taiwan	1.339***
Object:Variety	-0.637*
RLength (logged):Variety	0.591**

Significance codes: "***": $p < 0.001$; "**": $p < 0.01$; "*": $p < 0.05$

Regarding the intercept adjustments for Verb, the top ten verbs that strongly favor and disfavor the *ràng* passive are shown in Figure 2. Among the top ten verbs that prefer *ràng,* five are cognition or perception verbs, including 羡慕 *xiànmù* 'admire', 误以为 *wùyǐwéi* 'mistakenly think', 记 *jì* 'remember', 知晓 *zhīxiǎo* 'know', and 误会 *wùhuì* 'misunderstand', aligning with *ràng*'s semantic preference for predicates. Two are communication verbs, including 诟病 *gòubìng* 'denounce' and 赞 *zàn* 'praise', and the other three are cause-motion verbs including 入侵 *rùqīn* 'invade', 入 *rù* 'enter' and 吸收 *xīshōu* 'absorb'.

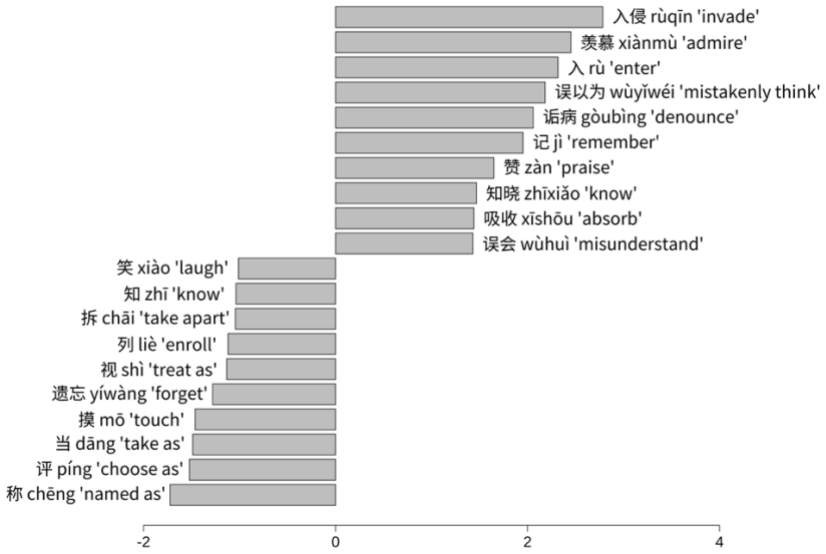

Figure 2 The top ten verbs strongly preferring the *ràng* passive (on the right side of the axis with values above 0) and the top ten verbs strongly disfavoring the *ràng* passive (on the left side of the axis with values below 0).

In summary, the parallel models exhibit excellent fit, low multicollinearity, and high predictive accuracy. In the big picture, then, which constraints are overall most important? The variable importance plot in Figure 3 provides an answer. Each subplot depicts the decrease in model goodness-of-fit when individual factors are removed. Despite their insignificant main effects, factors shown at the bottom are retained due to their significant interactions with Variety. Of these models, the *bèi* vs. non-*bèi* model performs best with the highest concordance index C, $R^2_{\text{mariginal}}$ and $R^2_{\text{conditional}}$, and classification accuracy.

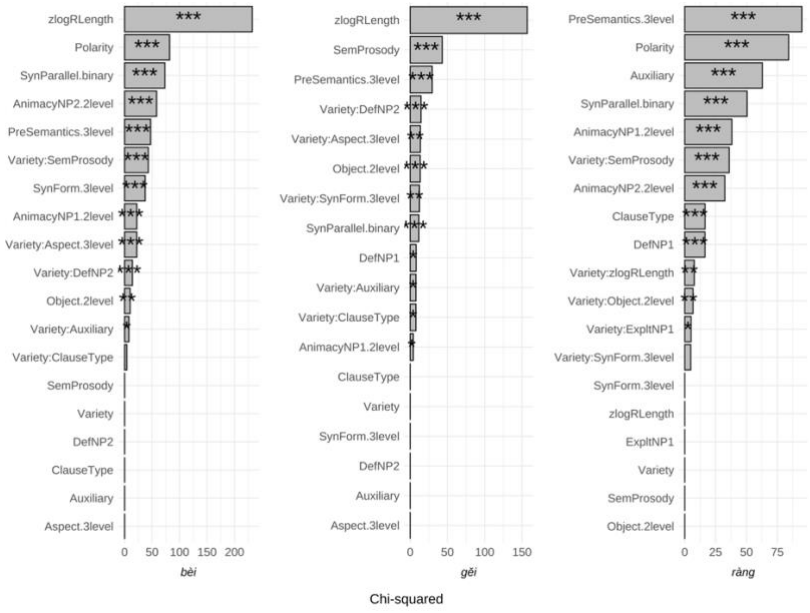

Figure 3 Variable importance (decrease in model goodness-of-fit if factor removed) of all factors in the three models. Left plot: *bèi* (see the model in Table 5); middle plot: *gěi* (see the model in Table 6); right plot: *ràng* (see the model in Table 7). Significance codes: "***": $p < 0.001$; "**": $p < 0.01$; "*": $p < 0.05$.

4. DISCUSSION

Section 4.1 discusses stable constraints across language varieties. Section 4.2 turns to fluid constraints and highlights differences between 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. Thus, in passive constructions, an inanimate NP1 and an animate NP2 are typically required. This pattern is corroborated by Guo and Chow's (2014) quantitative study, which examined both long and short *bèi* passives. Our results show that long *ràng* passives adhere to this expected animacy pattern, with 74.3% of NP1s

being inanimate and 82.8% of NP2s being animate, as in (5). In contrast, long *bèi* passives deviate from this pattern (e.g., (6)), displaying a higher likelihood of pairing with an animate NP1 and an inanimate NP2 (44.5% and 59.5%, respectively) compared to long *ràng* passives, where these figures are significantly lower (23.5% and 16.6%, respectively). Similarly, long *gěi* passives also strongly favor an animate NP1; however, unlike *bèi*, *gěi* does not exhibit a clear preference for NP2 animacy.

(5) zhTenTen17 (Simplified) hsw.cn marker=*ràng*

[存折	上	的	1.3 万	元]$_{NP1}$	**让**
cúnzhé	*shàng*	*de*	*13,000*	*yuán*	***ràng***
bankbook	on	REL	13,000	yuan	PASS
[人]$_{NP2}$	从	眉县	通过	银行	
rén	*cóng*	*méixiàn*	*tōnggùo*	*yínháng*	
people	from	Meixian	through	bank	
取	走	了。			
qǔ	*zǒu*	*le*			
withdraw	RES	PFV			

'The 13,000 yuan in the passbook was withdrawn from the bank in Meixian County.'

(6) zhTenTen17 (Traditional) books.com.tw marker=*bèi*

[我]$_{NP1}$	**被**	[电话]$_{NP2}$	骚扰	了。
wǒ	***bèi***	*diànhuà*	*sāorǎo*	*le*
I	PASS	phone-call	harass	PFV

'I was harassed by phone calls.'

Why do both long *bèi* and *gěi* passives deviate from the animacy distribution observed in the long *ràng* passive? Unlike *ràng*, which requires an explicit NP2 to form a well-formed passive construction (Shi 2023, 175–202), *bèi* and *gěi* can also be used in short passives. Short passives fall beyond the scope of this study, but were examined together with long passives in Guo and Chow (2014). While the expected animacy distribution holds across all *bèi* and *gěi* passives, including both long and short ones, it weakens when only long passives are considered.

Regarding NP2 animacy, passive constructions typically assign NP2 the role of an agent (an animate actor). However, as a more frequently used passive marker, *bèi* is also more heterogeneous and exhibits greater variability in usage. This pattern is similar to that of the Dutch causative

marker *laten* as discussed by Levshina, Geeraerts, and Speelman (2014). Thus, *bèi* is preferred when NP2 is inanimate, fulfilling atypical thematic roles such as an instrument or a cause. To sum up, compared to *ràng* and *gěi*, the prototypical passive marker *bèi* is preferred in the most atypical long passive constructions, specifically those with an animate NP1 and an inanimate NP2.

It is evident 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 (7). This pattern distinguishes *ràng* from its use as a causative marker from which the passive usage was 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 (7) with an animate NP1, 主编 *zhǔbiān* 'editor-in-chief', for example, as in (8), can cause ambiguity with the marker *ràng,* which can be interpreted as both a causative and a passive marker. Therefore, an inanimate NP1 plays a role in ambiguity avoidance. The *ràng* passive's preference for an animate NP2 is also consistent with Zhang and Wang's (2017) Multiple Correspondence Analysis of *gěi* and *ràng* passive constructions.

(7) zhTenTen17 (Simplified) magtech.com.cn marker=*ràng*

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

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

'The username and password of the editor-in-chief cannot be found by general editors.'

(8)

[主编]$_{NP1}$	不	能	**让**	[普通
zhǔbiān	*bù*	*néng*	***ràng***	*pǔtōng*
editor-in-chief	NEG	can	CAUS/PASS	general

编辑]$_{NP2}$	查	到…	
biānji	*chá*	*dào*	
editor	find	RES	

'The editor-in-chief cannot let general editors find (sth.).'

Nevertheless, the *ràng* passive construction shares some characteristics with the *ràng* causative construction, such as the avoidance of events, as argued by Chappell (1983). The "avoidable event" implies that NP1 (or the "owner" of NP1 if NP1 is inanimate) has the ability to prevent the current event but does not take any action with that intention. Considering (9), Chappell suggests that with the marker *ràng*, the focus of the sentence extends beyond the adverse impact of the event (as is the case with the marker *bèi*). Instead, it attributes responsibility to the implicit owner of NP1, who is deemed to have had the ability to prevent the knocking down of the ink bottle but chose not to act. When NP1 is animate, if the event is inherently unavoidable (such as destruction or death), the use of *ràng* is even considered ungrammatical, as in (10).

(9) Example (28) in Chappell (1983, 319) marker=*ràng*

[墨水]$_{NP1}$瓶		**让**	[弟弟]$_{NP2}$	打	翻	了。
mòshuǐ	*píng*	***ràng***	*dìdi*	*dǎ*	*fān*	*le*
ink	bottle	PASS	brother	knock	RES	PFV

'The ink bottle got knocked over by my younger brother.'

(10) Example (i) in Chappell (1983, 316) marker=*bèi*

唐山	好	几	万	人	**被**
tángshān	*hǎo*	*jǐ*	*wàn*	*rén*	***bèi***
Tangshan	good	several	10,000	people	PASS

*** (让)**	地震	夺	去	了	生命。
**** (ràng)***	*dìzhèng*	*duó*	*qù*	*le*	*shēngmìng*
* (PASS)	earthquake	take	RES	PFV	life

'Tens of thousands of people in Tangshan were killed by the earthquake.'

The preference for *ràng* passives with regard to predicate semantics and polarity also highlights this "avoidable event" characteristic. The *ràng* passive construction is more likely when the predicate indicates the change/state of cognition and perception, and when preceded by a negator or in an interrogative sentence. For instance, in (11), the predicate 看扁

kànbiǎn ‘be looked down upon’, signifying people’s negative perception, 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 NP1 to avoid being looked down upon. However, it should be noted that if NP1 is inanimate, as in (12), the ability to prevent the ongoing event from occurring lies not with NP1 itself, but with the speaker or the owner of NP1, who did not act. The concept of “avoidable event” aligns with NP1 responsibility in the *get*-passive mentioned by Thompson and Scheepers (2013), who further explain that this characteristic stems from an association with the causative *get*, and is not an inherent feature of the semantics of the *get*-passive. As the passive marker *ràng* also derived from a causative marker, this explanation regarding the *get*-passive may also account for the preference for avoidable events in the *ràng* passive.

(11) zhTenTen17 (Traditional) tabf.org.tw 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.’

(12) zhTenTen17 (Traditional) ellison.idv.tw marker=*ràng*

[连	我	对	蛇	的	感情]$_{NP1}$，
lián	*wǒ*	*duì*	*shé*	*de*	*gǎnqíng*
even	I	for	snake	REL	feelings
都	让	[您]$_{NP2}$	看	出来	了。
dōu	***ràng***	*níng*	*kàn*	*chūlái*	*le*
all	PASS	you	see	RES	PFV

‘Even my feelings for snakes can be seen by you.’

Furthermore, scrutiny of the mispredicted tokens[4] in the *ràng* vs. non-*ràng* model indicates that, of the 298 non-*ràng* tokens mispredicted as *ràng*, 225 (76%) feature cognitive or perceptual predicates, and 208 (70%) feature an inanimate NP1 and an animate NP2. To illustrate, example (13), a *bèi* passive construction, was mispredicted with a high probability (0.96) as *ràng*, featuring both an inanimate NP1, an animate NP2, and a cognitive predicate.

(13) zhTenTen (Simplified) pumch.cn marker=*bèi*

[这些	早期	症状]$_{NP1}$	[…]	常	会
zhèxiē	*zǎoqī*	*zhèngzhuàng*		*cháng*	*huì*
these	early	symptom		often	can
被	[病人]$_{NP2}$	误以为	是	老年	人
bèi	*bìngrén*	*wùyǐwéi*	*shì*	*lǎonián*	*rén*
PASS	patient	mistake	COP	old	people
缺	钙	[…]			
quē	*gài*				
lack	calcium				

'These early symptoms are often mistakenly thought by patients to be caused by calcium deficiency in the elderly.'

Compared to *bèi*, the passive markers *gěi* and *ràng* are more commonly used in spoken language (e.g., Li 2004; Zhang and Wang 2017). While our data is restricted to a single discourse medium (i.e., the internet), which makes impossible a clear distinction between spoken and written language, the web-based data we used encompass diverse discourse modes (Tao 1999), ranging from formal genres such as government documents to informal discourses such as two-way dialogues in novels and conversation-like comments on social media. This range provides a comprehensive reflection of passive markers' usage in contemporary Chinese.

A register preference for *gěi* and *ràng* is further supported by syntactic predictors, including NP1 definiteness and explicitness. The *gěi* passive is more likely to be used when NP1 is unclear, either as a generic noun, as in (14), or referring to the speaker, as in (15). The former typically appears in public warnings, while the latter is common in dialogues (e.g., in fiction) and social media comments (e.g., blogs). Similarly, *ràng* passives tend to occur with implicit NP1s, where the NP1 is specified elsewhere in the context rather than immediately preceding the marker. Both indefinite and implicit NP1s are characteristic of minor sentences (Chao 1968, 80–90; Tao 1996, 171–72), which are typical of spoken language. These syntactic patterns thus reinforce the colloquial preference for *gěi* and *ràng.*

(14) zhTenTen17 (Simplified) marker=*gěi*
qingdao315.com.cn

小心	**给**	[人]$_{NP2}$	骗	钱	了。
xiǎoxīn	***gěi***	*rén*	*piàn*	*qián*	*le*
cautious	PASS	people	cheat	money	PFV

'Watch out, (don't) get ripped off by others.'

(15) zhTenTen17 (Traditional) ezla.com.tw marker=*gěi*

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

'Useless! (I) was almost frightened to death by you.'

In a nutshell, long passive constructions exhibit some stable features across language varieties. *Bèi* is preferred in atypical passive constructions, favoring an animate NP1 and an inanimate NP2. In contrast, *ràng* passives inherit the attribute of avoidable events or NP1 responsibility from *ràng* causatives and reduce ambiguity by favoring an inanimate NP1. Compared to *bèi*, both *gěi* and *ràng* tend to convey a more colloquial tone.

4.2 Fluid Constraints across Language Varieties

The results of the three parallel models indicate that eight constraints differ across varieties of contemporary Chinese. For reasons of space, we selected the four most significant interactions[5] for further discussion.

Definiteness in language indicates information status. Definite nouns refer to old information, while indefinite nouns refer to new information. According to Song et al. (2007), the primary feature of information delivery in Mandarin *bèi* passives aligns with old-before-new information structure. This tendency is reflected in the primary pragmatic role of the NP2, which is to introduce new information (Guo and Chow 2014). The definiteness of NP2 is not a significant main effect in our models, indicating that there is no preference for the definiteness of NP2 with all three markers in Mainland Chinese. Nevertheless, as an interaction term with Variety according to our models, an indefinite NP2 significantly attracts the *bèi* passive in Taiwan Chinese (e.g., (16)), implying that NP2 in the long *bèi* passive in Taiwan Chinese tends to prominently convey new information. Thus, Taiwan Chinese may adhere more closely to old-before-

new information structure with respect to the use of the prototypical passive marker *bèi*.

(16) zhTenTen17 (Traditional) reports.tw marker=*bèi*

老百姓	常	**被**	[一些	伪	健康
lǎobǎixìng	*cháng*	***bèi***	*yìxiē*	*wěi*	*jiànkāng*
people	often	PASS	some	false	health
信息]$_{NP2}$	所	误导。			
xìnxī	*suǒ*	*wùdǎo*			
information	AUX	mislead			

'People are often misled by false health information.'

Xing (2004) interprets the loss of source meaning and semantic broadening of the *bèi* passive as manifestations of grammaticalization. Guo and Chow (2014) agree, stating that *bèi* is taking on a broader functional role while retaining its original adversative sense. Authors (Submitted) also provide empirical support for this claim. An analysis of 893 randomly selected long *bèi* passives from the Corpus of Historical Chinese (CHC)[6] shows that 94.7% exhibit negative semantic prosody, confirming that in Modern Chinese, long *bèi* passives are predominantly adversative. In the present synchronic dataset, however, the proportion of long *bèi* passives with negative semantic prosody has significantly decreased to 47.5% overall. This shift suggests that the adversity effect on *bèi* passives has weakened over time, with *bèi* now more frequently used in non-negative contexts (53.5%). This weakening of the adversity effect can be attributed primarily to the Westernization of Chinese grammar (Kubler 1985, 89; Wang 1985; Hsu 1994).

Importantly, our statistical analysis reveals a significant interaction between semantic prosody and Variety, meaning that the weakening of the adversity effect does not apply uniformly across language varieties. In Taiwan Chinese, 60.2% of long *bèi* passives still exhibit negative semantic prosody, whereas in Mainland Chinese, this proportion is significantly lower at 34.9%. Comparing these figures with its 94.7% adversative usage in Modern Chinese, we posit that the *bèi* passive in Taiwan Chinese aligns more closely with the semantic expectations of Modern Chinese, as it continues to demonstrate a stronger adversative sense. Within the 167 mispredicted *bèi* passives in Taiwan Chinese, 12.5% were misclassified

due to their occurrence in positive contexts, as in (17). Additionally, the difference in positive semantic prosody between the two varieties is even more striking: in Mainland Chinese, 39.0% of long *bèi* passives carry positive semantic prosody, whereas in Taiwan Chinese, this figure drops to 12.3%. It seems that it is the *ràng* passive that is more likely to be used to express positive semantic prosody in Taiwan Chinese (e.g., (18)).

(17) zhTenTen17 (Traditional) atmovies.com.tw marker=*bèi*

“…	希望	[这	部	电影]$_{NP1}$	可以
	xīwàng	*zhè*	*bù*	*diànyǐng*	*kěyǐ*
	hope	this	CL	film	can

被	[更多	人]$_{NP2}$	看见…”
bèi	*gèngduō*	*rén*	*kànjiàn*
PASS	more	people	see

‘(I) hope that this film can be seen by more people.’

(18) zhTenTen17 (Traditional) slime.com.tw marker=*ràng*

[世	上	好处]$_{NP1}$，	都	**让**	[你们]$_{NP2}$
shì	*shàng*	*hǎochù*	*dōu*	***ràng***	*nǐmen*
world	on	benefits	all	PASS	you

捡	到	了。
jiǎn	*dào*	*le*
pick	RES	PFV

‘All the good in the world was got by you.’

Furthermore, two syntactic predictors, SynForm and Aspect, exhibit distinct preferences for the passive markers *bèi* and *gěi* across varieties. In Mainland Chinese, *bèi* is less likely to occur with the perfective marker *le* or *guo*, while *gěi* is more commonly used with a perfective marker (e.g., (19)). This pattern aligns with the corpus-based findings of Xiao et al. (2006) regarding the interaction between passives and aspect. However, in Taiwan Chinese, these aspectual preferences become less pronounced: 54% of long *gěi* passives occur without a perfective marker, as in (20). This deviation may be influenced by the passive marker *hoo* in Taiwanese, a widely spoken Southern Min dialect in Taiwan. Since *hoo*, meaning “give” (Cheng et al. 1999), is a counterpart to *gěi* in Mandarin Chinese, *gěi* in Taiwan Chinese may function more freely as a passive marker and may be

less constrained by this aspectual constraint compared to *gěi* in Mainland Chinese.

(19) zhTenTen (Simplified) myliving.cn marker=*gěi*

整	个	小区	的	环境	都
zhěng	*gè*	*xiǎoqū*	*de*	*huánjìng*	*dōu*
whole	CL	residence	REL	environment	all
给	他们	破坏	了。		
gěi	*tāmen*	*pòhuài*	*le*		
PASS	they	destroy	PFV		

'The environment of the whole neighborhood has been destroyed by them.'

(20) zhTenTen17 (Traditional) marker=*gěi*
dragon-go.com.tw

后	山坡	**给**	海	水	环绕，
hòu	*shāngpō*	***gěi***	*hǎi*	*shuǐ*	*huánrào*
real	hillside	PASS	sea	water	surround
时常	水雾	缭绕。			
shícháng	*shuǐwù*	*liáorào*			
often	mist	shroud			

'The back hillside is surrounded by seawater and is often shrouded in mist.'

Nevertheless, *gěi* is significantly less favored in Taiwan Chinese when the predicate takes the form of a complex stative construction (i.e., verb+*de*) or a coordinating verb and serial verb construction. Among 31 long *gěi* passives with a complex verb in Taiwan Chinese, 74.2% were mispredicted, as in (21). The syntactic form of the predicate interacts with the length of the right periphery, as complex verbs tend to increase the right periphery's length. For instance, in (21), the right periphery extends to 12 characters, a complex structure that strongly favors *bèi* over *gěi*. According to the model results, *bèi* is preferred with longer right peripheries, while *gěi* is favored in shorter constructions. This preference for both complex verbs and a longer right periphery can be explained by the cognitive complexity principle (Rohdenburg and Mondorf 2003; Deshors and Gries 2014), which suggests that more explicit constructions are favored in cognitively more demanding contexts. Thus, the commonly used passive marker *bèi* might be preferred in complex clauses, as it helps reduce cognitive load during language processing. While dialectal influences in

Taiwan Chinese may contribute to syntactic preferences in passive choice, their effect appears secondary to the cognitive mechanisms that govern variant choice.

(21) zhTenTen17(Traditional) tabf.org.tw marker=*gěi*

买	保险	最	傻	了，	钱
mǎi	*bǎoxiǎn*	*zuì*	*shǎ*	*le*	*qián*
buy	insurance	most	stupid	PFV	money
都	**给**	保险	公司	赚	去
dōu	***gěi***	*bǎoxiǎn*	*gōngsī*	*zhuàn*	*qù*
all	PASS	insurance	company	earn	RES
买	大	楼，	炒	地皮。	
mǎi	*dà*	*lóu*	*chǎo*	*dìpí*	
buy	big	building	speculate	land	

'Buying insurance is the stupidest thing. All money was earned by the insurance companies to buy buildings and speculate in land.'

In summary, passive choice differs across varieties of Mandarin Chinese. Taiwan Chinese favors *bèi* with indefinite NP2s and perfective markers, while *gěi* is less likely to co-occur with perfective markers or complex predicates. *Ràng* is more likely to be used than *bèi* in contexts with positive prosody. These patterns reflect both dialectal influence on Taiwan Chinese and its stronger adherence to old-before-new information structure and the semantic expectations of Modern Chinese, though the cognitive complexity principle plays a greater role in Taiwan Chinese.

5. CONCLUSION

This paper has presented a big, richly annotated dataset for studying the alternation of Chinese long passive constructions marked by *bèi*/*gěi*/*ràng*. Our analysis of three Chinese long passives confirms that the probabilistic conditioning of long passives is complex indeed. Through regression modeling, we delineated the different usages of these three variants in Mandarin Chinese as follows. *Bèi* is preferred in atypical passive constructions (i.e., those with an animate NP1 and an inanimate NP2). *Gěi* and *ràng* have a more colloquial tone, as both are preferred when NP1 is implicit (in which case NP1 refers to a generic noun or the speaker).

The *ràng* passive inherits the property of avoidable events from *ràng* causatives and differs from *bèi* and *gěi* in favoring predicates indicating cognition/perception and non-affirmative sentences. *Ràng* passives also mitigate ambiguity by preferring an inanimate NP1.

Long passive constructions also exhibit significant differences between Taiwan Chinese and Mainland Chinese. In Taiwan Chinese, the typical passive marker *bèi* adheres more closely to old-before-new information structure and the adversative reading of passives in Modern Chinese, whereas *ràng* tends to be used in contexts with positive prosody. The syntactic usage of long *gěi* passives is influenced by dialectal variation, resulting in a less strict aspectual requirement in Taiwan Chinese. Usage of *gěi* passives also conforms to the cognitive complexity principle, as *gěi*, being a non-prototypical passive marker, is less likely to be used with complex predicates.

This study, to the best of our knowledge, is the first multifactorial analysis of the Chinese long passive alternation. However, it is important to note the limitations of this study. Firstly, due to the widespread use of the typical passive marker *bèi* and the lack of a large-scale balanced Chinese corpus, it is currently difficult to empirically investigate this alternation across different registers or genres. Secondly, although the passive markers *bèi* and *gěi* can be used in short passive constructions, our analysis only focused on long passive constructions. Therefore, it would be beneficial to conduct additional multifactorial analyses of the choice of Chinese passive constructions marked by *bèi*/*gěi*, as their overall usage, across short and long passives, remains unclear.

NOTES

1. In contemporary Chinese, there are three strategies to indicate passive voice in total:
   1) the above-mentioned passive constructions marked by *bèi*/*gěi*/*ràng*/*jiào*, which is also the most common strategy (Zhu 1982, 178−179);

2) the automatic passive (Zhang 1953) formed with *zāo* 'suffer, meet with' as a representative, which is also used as a full lexical verb (Tian 2010);
3) the notional passive, also labeled "semantic passive sentences" (Wang 1989), a patient-topic clause without an overt passivizing element, similar to the active sentence, but with NP1 being the topic preceding the predicate.

These latter two types are not included in our study because neither of them has an overt auxiliary verb functioning as a passive marker.

2. All examples in this paper are from the Chinese Web Corpus 2017 (zhTenTen17, which will be introduced in detail in Section 2.1), unless otherwise specified.

3. A reviewer questioned grouping verb+*le*/*zhe*/*guo* with RVCs, as aspect markers may not inherently express results. However, a supplementary inference tree analysis (see supplementary materials) shows they fall under the same branch, indicating a similar influence on long passive choice. While these two forms may convey different meanings, the grouping suggested by the supplementary inference tree analysis helps mitigate data sparsity and ensures more robust regression modeling.

4. We are grateful to an anonymous reviewer for suggesting the discussion of misclassified tokens.

5. This selection of interactions is based on the Bonferroni correction, a statistical method used to adjust *p*-values when conducting multiple comparisons to control for the overall probability of a Type I error. The adjusted *p*-value is calculated by dividing the original significance level (0.05) by the number of tests performed (16, as 16 interactions were tested sequentially using the drop1() function in each model, as described in Section 2.4). Therefore, only those four interactions with adjusted *p*-values lower than 0.003 (0.05/16) are selected for further discussion.

6. The Corpus of Historical Chinese (CHC), the largest corpus of historical Chinese data, is the diachronic component of the CCL Corpus (Center for Chinese Linguistics Corpus, Peking University, 2014 Version, accessed in 2023, (Zhan et al. 2019).

REFERENCES

AUTHORS. Submitted. A Diachronic Study on Chinese Analytic Long Passive Constructions Marked by *Bèi*, *Gěi*, *Jiào* and *Ràng*.

BAAYEN, R Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge: Cambridge University Press.

BENNETT, Paul A. 本涅特. 1981. Beidongshi Chuzhishi de fazhan 被动式处置式的发展 (The evolution of passive and disposal sentences). *Journal of Chinese Linguistics* 9(1): 61–90.

BOHMANN, Axel, Julia Müller, Mirka Honkanen, and Miriam Neuhausen. 2023. A Large-Scale Diachronic Analysis of the English Passive Alternation. In *Language and Linguistics in a Complex World*, edited by Beatrix Busse, Nina Dumrukcic, and Ingo Kleiber, 32:31–56. Berlin, Boston: De Gruyter.

BRESNAN, Joan. 2007. Is Syntactic Knowledge Probabilistic? Experiments with the English Dative Alternation. In *Roots: Linguistics in Search of Its Evidential Base*, edited by Sam Featherston and Wolfgang Sternefeld, 75–96. Berlin, New York: Mouton de Gruyter.

BRESNAN, Joan, and Marilyn Ford. 2010. Predicting Syntax: Processing Dative Constructions in American and Australian Varieties of English. *Language* 86(1): 168–213.

CARTER, Ronald, and Michael McCarthy. 1999. The English *Get* -Passive in Spoken Discourse: Description and Implications for an Interpersonal Grammar. *English Language and Linguistics* 3(1): 41–58.

CHAO, Yuen Ren. 1968. *A Grammar of Spoken Chinese*. Berkeley: University of California Press.

CHAPPELL, Hilary. 1983. *A Semantic Analysis of Passive, Causative and Dative Constructions in Standard Chinese*. Canberra: Australian National University.

CHEN, I-hsuan 陈怡璇 and Chinfa Lien 连金发. 2011. Minnanyu shidong yu beidong zhi lishi yanbian - cihui yihan yu jiegou de jiemian 闽南语使动与被动之历时演变 - 词汇意涵与结构的界面 (A diachronic

perspective on causative variants and related passives in Southern Min). *Journal of Chinese Linguistics* 39(2): 311–344.

CHENG, Lisa L.-S., C.-T. James Huang, Y.-H. Audrey, and C.-C. Jane Tang. 1999. Hoo, Hoo, Hoo: Syntax of the Causative, Dative, and Passive Constructions in Taiwanese. *Journal of Chinese Linguistics Monograph Series*, (14): 146–203.

COLLINS, Peter C. 1996. Get-Passives in English. *World Englishes* 15(1): 43–56.

CORRIGAN, Roberta. 1988. Children's Identification of Actors and Patients in Prototypical and Nonprototypical Sentence Types. *Cognitive Development* 3(3): 285–297.

DESHORS, Sandra C., and Stefan Th. Gries. 2014. A Case for the Multifactorial Assessment of Learner Language: The Uses of May and Can in French-English Interlanguage. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, edited by Dylan Glynn and Justyna A. Robinson, 179–204. Human Cognitive Processing 43. Amsterdam: Benjamins.

FEHRINGER, Carol. 2022. The *Get* -Passive in Tyneside English: A Highly Frequent yet Constrained Variant. *English World-Wide. A Journal of Varieties of English* 43(3): 330–356.

FENG, Shengli 冯胜利. 1997. Guanyuelilun yu Hanyu beidongju 管约理论与汉语被动句 (GB-Theory and the Passive Construction in Chinese). In *Zhongguo yuyanxue luncong* 中国语言学论丛, vol.1, edited by C.-T. James Huang, 1–28. Beijing: Beijing Language and Culture University Press.

———. 2012. Empty Operator Movement in Chinese Passive Syntax. In *Shuliluoji zhi mei: Fangli jiaoshou jinianwenji* 数理逻辑之美：方立教授纪念文集, edited by Lujiang Wang 王路江, 117–137. Beijing: Beijing Language and Culture University Press.

———. 2013. Short Passives in Modern and Classical Chinese. In *Breakingng down the Barriers: Interdisciplinary Studies in Chinese Linguistics and Beyond*, edited by Guangshun Cao, Hilary Chappell, Redouane Djamour, and Thekla Wiebusch, 2:595–623. Language and Linguistics Monograph Series 50. Taipei, Taiwan: Academia Sinica.

GAMER, Matthias, Jim Lemon, and Ian Fellows Puspendra Singh. 2019. Irr: Various Coefficients of Interrater Reliability and Agreement. Available at: https://cran.r-project.org/web/packages/irr/index.html.

GUO, Hongjie 郭鸿杰, and Daryl Chow. 2014. A Corpus-Based Variationist Approach to Bei Passives in Mandarin Chinese. *Corpus Linguistics and Linguistic Theory* 10(1): 139–173.

GUO, Hongjie 郭鸿杰, and Hong Han 韩红. 2012. Yuliaoku qudong de yinghan yuyan jiechu yanjiu: yi 'bei'ziju weili 语料库驱动的英汉语言接触研究: 以'被'字句为例 (A Corpus-Based Study of English-Chinese Language Contact: A Case Study of Bei (被) Passives). *Foreign Language Teaching and Research* 外语教学与研究 44(03): 359–370+479.

GRIES, Stefan Th. 2015. The Most Under-Used Statistical Method in Corpus Linguistics: Multi-Level (and Mixed-Effects) Models. *Corpora* 10(1): 95–125.

HASHIMOTO, Mantaro J. 1988. The Structure and Typology of the Chinese Passive Construction. In *Passive and Voice*, edited by Masayoshi Shibatani, 16: 329–54. Typological Studies in Language. Amsterdam: John Benjamins Publishing Company.

HSU, Jia-Ling. 1994. Englishization and Language Change in Modern Chinese in Taiwan. *World Englishes* 13(2): 167–184.

HUANG, C.-T. James. 1999. Chinese Passives in Comparative Perspective. *Tsing Hua Journal of Chinese Studies* 29(4): 423–509.

HUANG, C.-T. James, Y.-H. Audrey Li, and Yafei Li. 2009. *The Syntax of Chinese*. New York: Cambridge University Press.

JIANG, Lansheng 江蓝生. 2000. *Jindai Hanyu Tanyuan* 近代汉语探源 (The origin of Modern Chinese). Beijing: The Commercial Press.

JIANG, Shaoyu 蒋绍愚. 2002. 'Gei'ziju, 'jiao'ziju biao beidong de laiyuan - jian tan yufahua, keitui. he gongneng kuozhan '给'字句、'教'字句表被动的来源——兼谈语法化、 类推和功能扩展 (The Source of 'Gei' and 'Jiao' Passive Sentences: About Grammaticalization, Analogy and Extension). In *Yuyanxue luncong* 语言学论丛 26: 67–78.

KUBLER, Cornelius C. 1985. *A Study of Europeanized Grammar in Modern Written Chinese*. Taiwan: Student Book Company.

LABOV, William. 1972. *Sociolinguistic Patterns*. Vol. 10. Philadelphia: University of Pennsylvania Press.

LEVSHINA, Natalia. 2015. *How to Do Linguistics with R: Data Exploration and Statistical Analysis*. Amsterdam: John Benjamins Publishing Company.

LEVSHINA, Natalia, Dirk Geeraerts, and Dirk Speelman. 2014. Dutch Causative Constructions: Quantification of Meaning and Meaning of Quantification. In *Corpus Methods for Semantics: Quantitative Studies in Polysemy and Synonymy*, edited by Dylan Glynn and Justyna A. Robinson, 205–221. Human Cognitive Processing 43. Amsterdam: Benjamins.

LI, Chao 李朝. 2007. Hanyu 'bei'zi jiegou de yanbian 汉语'被'字结构的演变 (Evolution of the Bei Constructions in Chinese). *Journal of Chinese Linguistics* 35 (1): 98–127.

LI, Charles N., and Sandra A. Thompson. 1989. *Mandarin Chinese: A Functional Reference Grammar*. Berkeley: CA University of California Press.

LI, Wei. 2004. Qing zhongye yilai beijinghua de beidong 'gei' jiqi xiangguan wenti - jianji 'nanfang guanhua de beidong 'gei' 清中叶以来北京话的被动'给'及其相关问题——兼及'南方官话'的被动'给' (Passive 'Gei' in Peking Dialect since Mid-Qing Dynasty and Related Issues of Passive 'Gei' in Southern Mandarin Chinese). *Journal of Sun Yat-sen University (Social Sciences Edition)* 中山大学学报 (社科版) 44(03): 35–40+123.

LIESENFELD, Andreas, Meichun Liu, and Chu-Ren Huang. 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.

LIN, Tzong-Hong Jonah, and Yu-Shan Huang. 2015. Structures of the Mandarin Gei Constructions. *Journal of East Asian Linguistics* 24(3): 309–338.

LIU, Mei-Chun, and Ting-Yi Chiang. 2008. The Construction of Mandarin VerbNet: A Frame-Based Study of Statement Verbs. *Language and Linguistics* 9(2): 239–270.

LIU, Yuehua, Wenyu Pan, and Ye Gu. 2001. *Shiyong Xiandai Hanyu Yufa* 实用现代汉语语法 (*Practical Modern Chinese Grammar*). Beijing: The Commercial Press.

PEYRAUBE, Alain 贝罗贝. 1989. Shi shiji yiqian hanyu beidongju de lishi fazhan 10 世纪以前汉语被动句的历史发展 (History of the Passive Constructions in Chinese Until the 10th Century). *Journal of Chinese Linguistics* 17(2): 335–372.

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

ROHDENBURG, Günter, and Britta Mondorf, eds. 2003. *Determinants of Grammatical Variation in English*. Topics in English Linguistics 43. Berlin; New York: Mouton de Gruyter.

SHI, Dingxu 石定栩. 2008. Changduan 'bei'ziju zhi zheng 长短'被'字句之争 (Probing into the Long and Short "Passive" Words and Expressions). *Journal of Qinghai Nationalities Institute* 青海民族学院学报 (03): 111–117.

SHI, Yuzhi. 2023. *The Evolution of Chinese Grammar*. Cambridge: Cambridge University Press.

SONG, Wenhui 宋文辉, Zhengjing Luo 罗政静, and Jingchao Yu 于景超. 2007. Xiandai hanyu beidongju shishi yinxian de jiliang fenxi 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* 中国语文 (02): 113–124+191.

SZMRECSANYI, Benedikt. 2013. Diachronic Probabilistic Grammar. *English Language and Linguistics* 19 (December): 41–68.

TAGLIAMONTE, Sali A. 2012. *Variationist Sociolinguistics: Change, Observation, Interpretation*. Malden & Oxford: Wiley-Blackwell.

TANG, Sze-Wing. 2001. A Complementation Approach to Chinese Passives and Its Consequences." *Linguistics* 39(2): 257–295.

———. 2004. Zuogehua he hanyu beidongju 作格化和汉语被动句 (Ergativization and Chinese passives). *Zhongguo Yu Wen* 中国语文 (04): 291–301+383.

TAO, Hongyin. 1996. *Units in Mandarin Conversation: Prosody, Discourse, and Grammar*. Studies in Discourse and Grammar. Amsterdam/Philadelphia: John Benjamins Publishing Company.

———. 1999. Shi lun yuti fenlei de yufaxue yiyi 试论语体分类的语法学意义 (Discourse taxonomies and their grammatico-theoretical implications). *Contemporary Linguistics* 当代语言学 (3): 15–24+61.

THOMPSON, Dominic, Fernanda Ferreira, and Christoph Scheepers. 2018. One Step at a Time: Representational Overlap Between Active Voice, Be-Passive, and Get-Passive Forms in English. *Journal of Cognition* 1(1).

THOMPSON, Dominic, and Christoph Scheepers. 2013. Harmonizing the Passive: A New Proposal for Passive Constructions in Generative Grammar. *Newcastle Working Papers in Linguistics* 19 (September):74–96.

TIAN, Chunlai 田春来. 2010. Biao beidong de 'zao' de lishi kaocha 表被动的'遭'的历时考察 (Diachronic Study on Zao). *Research in Ancient Chinese Language* 古汉语研究 (01): 68–73+96.

TING, Jen. 1998. Deriving the Bei-Construction in Mandarin Chinese. *Journal of East Asian Linguistics* 7(4): 319–354.

VIT, Suchomel. 2021. Genre Annotation of Web Corpora: Scheme and Issues. In *Proceedings of the Future Technologies Conference (FTC) 2020. Volume 1*, edited by Kohei Arai, Supriya Kapoor, and Rahul Bhatia, 1: 738–754. Advances in Intelligent Systems and Computing 1288. Cham: Springer.

WANG, Li 王力. 1985. *Zhongguo Xiandai Yufa* 中国现代语法 (*Chinese Modern Grammar*). Beijing: The Commercial Press.

———. 1989. *Hanyu yufa shi* 汉语语法史 (*The History of Chinese Grammar*). Beijing: The Commercial Press.

WANG, Mengxiang 汪梦翔. 2022. Jiyu guize de feidianxing youbiao beidongju de yuyi juese zidong biaozhu yanjiu 基于规则的非典型有标被动句的语义角色自动标注研究 (Rule-Based Automatic Labeling of Semantic Roles of Atypical Marked Passive Sentences). *Yuyan wenzi yingyong* 语言文字应用 (02): 122–132.

XIAO, Richard, Tony McEnery, and Yufang Qian. 2006. Passive Constructions in English and Chinese: A Corpus-Based Contrastive Study. *Languages in Contrast* 6(1): 109–149.

XING, Fuyi 邢福义. 2004. Chengcixing ‘bei’ziju 承赐型‘被’字句 (On the favor-accepting Type of Passive Sentences with the Marker ‘bei’). *Studies in Language and Linguistics* 语言研究 24(01): 1–11.

YANG, Xiaoming 杨晓明. 2018. Xiandai hanyu kouyu zhong de beidong biaojici yanjiu - jiyu yuliaoku de yanjiu 现代汉语口语中的被动标记词研究——基于语料库的研究 (A study of passive markers in spoken Chinese: a corpus-based study). *Journal of Shandong Agricultural University (Social Science Edition)* 山东农业大学学报(社科版) 20(01): 91–96.

ZHAN, Weidong 詹卫东, Rui Guo 郭锐, Baobao Chang 常宝宝, Yirong Chen 谌贻荣, and Long Chen 陈龙. 2019. Beijing daxue CCL yuliaku de yanzhi 北京大学 CCL 语料库的研制 (The Building of the CCL Corpus: Its Design and Implementation). *Corpus Linguistics* 语料库语言学 6(1): 77–86.

ZHANG, Weiwei 张炜炜, and Fang Wang 王芳. 2017. Cong jiyu yangli de gainian kongjian kan goushi jiaoti - yi ‘rang’ he ‘gei’ de beidong yongfa weili 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* 外语教学与研究 6: 22–23+145.

ZHANG, Zhigong 张志公. 1953. *Hanyu Yufa Changshi* 汉语语法常识 (*General Knowledge of Chinese Grammar*). Beijing: Chinese Youth Publishing House.

ZHU, Dexi 朱德熙. 1982. *Yufa Jiangyi* 语法讲义 (*Grammar notes*). Beijing: The Commercial Press.

ZUUR Alain, Elena N. Ieno, Neil Walker, Anatoly A. Saveliev, and Graham M. Smith. 2009. *Mixed Effects Models and Extensions in Ecology with R*. New York: Springer Science & Business Media.

现代汉语中“被”/“给”/“让”长被动构式的多因素研究

管苏敏，Benedikt Szmrecsanyi，张炜炜
荷语鲁汶大学

摘要

目前，汉语被动构式尚未通过语言变异的视角进行系统分析。本研究基于变异社会语言学的框架，利用中文网络语料库，采用混合效应逻辑斯蒂回归，考察了影响汉语“被”/“给”/“让”长被动构式选择的句法、语义和语言变体等因素。研究发现：非典型被动结构更倾向于选择“被”标记；“给”和“让”更具口语色彩，且“让”保留了致使构式中表达事件避免性的一面，但多用于无生命的 NP1 以避免歧义。在台湾汉语和大陆汉语的比较中，台湾汉语“被”的使用更符合旧信息-新信息的信息结构以及现代汉语对被动构式带消极义的要求；“让”在积极语义韵中逐渐替代了“被”；“给”被动的使用受到台湾方言中被动标记的影响，但也遵循认知上的复杂性原则。本文揭示了汉语长被动构式的复杂性，通过多因素分析考察了语言内部和外部因素，为进一步研究汉语被动构式提供了价值。

关键词
长被动　**让**被动　**给**被动　**变**异社会语言学　**多**因素分析