

A MULTIFACTORIAL ANALYSIS ON THE SYNTACTIC VARIATIONS IN ONE KIND OF CHINESE VERB-DIRECTION CONSTRUCTION (VDC)

TAO LIN

University of Colorado Boulder

This research investigates the linguistic features that can predict native speakers' choice of three word order alternations of one kind of Verb-Direction Constructions (VDCs) in Mandarin Chinese, i.e., object initial, middle, and final types. Previous work on the contributing features is done through qualitative observations, and the direction of some predictions is unclear. This paper provides corpus based quantitative studies on both monofactorial and multifactorial analyses, which clarify the effectiveness of most of these parameters. In addition to the discussed features, some new features, such as the use of certain classes of verbs and imperatives, are also introduced in order to enhance the model; further directions and improvements are also.

Keywords: multinomial logistic regression, word order, constructions, Chinese

1. INTRODUCTION

Like many languages, such as English and Arabic, word order variation in Chinese is a common device in syntactic alternations. English verbal particle constructions (VPCs) can allow different placements of the verbal particles, which are realized in prepositions (Dehé 2002, Jackendoff 2002 & Wasow 2002).

- (1) a. John took out a book.
- b. John took a book out.

In the Chinese translation of sentence 1, instead of using the preposition 'out,' Chinese can take two verbal particles: *chu* 'exit' and *lai* 'come' marked as Particle 1 and 2 (P₁ and P₂) in the examples. Their origins all came from directional verbs (Lamarre 2007). There are three possible alternations.

- (2) a. (Object initial type)
 - ta na yi ben shu chu lai
 - he take one CL book exit(P₁) come(P₂)

b. (Object middle type)

ta na chu yi ben shu lai
 he take exit (P₁) one CL book come(P₂)

c. (Object final type)

ta na chu lai yi ben shu
 he take exit(P₁) come(P₂) one CL book

‘He took out a book.’

These three alternations share the same meaning in example 2. The only difference is the placement of the object ‘a book.’

This study aims at adopting a quantitative approach to discover how native Chinese speakers prefer one alternation over another. In Chinese there are 28 directional particles (Liu 1998). However, in this research, I focus only on one combination (*chu lai*), because the linguistic factors related to it different orders have been well studied. To further narrow down the searching scope, I only choose one verb, *na* ‘to take,’ as the target verb. Next, I will introduce some linguistic factors that are believed to affect the choice of alternations, and use it to make up the first hypothesis (Model A). I collect the data in Section 3. Based on my observation on the corpora data, I discuss more possible factors in Section 4. I build up the second hypothesis (Model B) by adding these factors to Model A, which contains a set of hypotheses to be tested in factorial analysis. I identify the statistical techniques in Section 5. The results in Section 6 suggest that Model B is statistically well supported. While the results have many implications for future work, expanding to other kinds of VDC-related phenomena would be an obvious next step.

2. VARIABLES THAT PURPORTEDLY GOVERN THE ALTERNATIONS

A lot of work on the syntactic alternations of VDCs has been done within the last twenty years. Mostly using introspective data, researchers believe that the emergence of the three alternations can result from features in syntactic (2.3 & 2.4), semantic (2.2), phonological (2.1), and discursive (2.5 & 2.6) domains.

2.1. THE LENGTH OF NOUN PHRASES (NPS)

Sun (2012) argues that the syllable/character length is related to the use of the three alternations. THE OBJECT-INITIAL ALTERNATION TENDS TO EXCLUDE LONG OBJECTS, AND THE OBJECT-FINAL ONES ARE LESS LIKELY TO HAVE SHORT OBJECTS.

- (3) a. na chu banfa lai
 take go.out (P₁) method come (P₂)
 ‘to come up with a plan’
- b. na chu lai **yi ge feichang bianjie de caomiaoyi**
 take P₁ P₂ one CL very convenient ASSOC scan.machine
 ‘to take out a very user-friendly scanner’

The bolded parts are the object NPs. The object-middle example 3a contains a two-syllable NP, while the object-final example 3b has ten syllables.

2.2. CONCRETENESS OF NPs

WHEN THE OBJECT IS AN ABSTRACT ITEM, THE OBJECT-MIDDLE ALTERNATIONS ARE MORE COMMON (Liu 1998). In this condition, an object-final alternation can usually be replaced by an object-middle one, but it is not true for an object-initial alternation.

- (4) men-feng-li chuan chu lai yi zhen ge-sheng
 door-gap-inside pass P₁ P₂ one CL song-sound
 ‘Sounds of music came from the gap in the door.’

In example 4, it is grammatical to transform the object-final alternation into an object-middle one by putting the object ‘a song’ between Particle 1 and 2. However, the placement does not work for an object-initial alternation.

2.3. DEFINITENESS OF NPs

In Chinese, a definite NP can be indicated by demonstratives (i.e., *zhe* ‘this,’ or *na* ‘that’), proper nouns as genitive (i.e., *fuermosi de shu* ‘Holmes’ book’), and personal pronouns (i.e., *ta de* ‘his’). COMPARED TO THE NON-OBJECT-FINAL ALTERNATIONS, NPs IN OBJECT-FINAL ONES ARE LESS LIKELY TO BE DEFINITE (ZHANG 1991 & NIU 2002). My pilot search on Google shows that there are only 12 results for the object-final alternation *na chu lai zhe ben shu* ‘to take out this book,’ while the other two alternations with demonstratives *zhe* ‘this’ have much more results.

2.4. THE ASPECT MARKER LE AFTER THE VERB

The influence of the post-verb aspect marker *le* on the choice of the alternations is controversial. Based on introspection, Lu (2002) concludes that there is no alternation preference

for the aspect marker. Chen (1982) believes that *le* only appears in the object-initial alternation. Zhang (1991) argues that the other two alternations can also have *le*. His corpus study shows that THE OBJECT-INITIAL ALTERNATION TENDS TO HAVE THE POST-VERB ASPECT MARKER *LE* AND THAT THE OTHER ALTERNATIONS HAVE MUCH LESS USE OF *LE*.

- (5) ta na le zhi chu lai
 He take ASP paper P₁ P₂
 ‘He takes paper out.’

The object-initial type in example 5 can be replaced by an object-middle type, but cannot be replaced by an object-final one.

2.5. OLD AND NEW INFORMATION OF NPs

If an NP is definite or has a referent in the previous clauses, then I call it old information. An NP that has never been introduced before is new information. THE OBJECT-FINAL AND OBJECT-MIDDLE ALTERNATIONS TEND TO INTRODUCE NEW INFORMATION (ZHANG 1991).

- (6) wo yiwei ta hui na chu yi feng xialüdi de **xin** lai,
 I think she will take P₁ one CL Charlotte ASSOC letter P₂

 danshi bingbu jian ta na **xin** chu lai
 but NEG see she take letter P₁ P₂
 ‘I thought she would take out Charlotte’s letter, but I didn’t see that she took out any.’

In Example 6, there are two verb-direction constructions with the object *xin* ‘letter.’ Although the second ‘letter’ is indefinite, it is old information because of its prior mention.

2.6. ZERO-ANAPHORA

Zero-anaphora means a gap in a clause whose referent is the object of a VDC in the previous clause. OBJECT-FINAL ALTERNATION IS MORE HIGHLY CORRELATED TO THE USE OF ZERO-ANAPHORAS (SUN 2012), as in 7.

- (7) fuermosi na chu lai yi xiao pian qianbi mu-xie,
Holmes take P1 P2 one small CL pencil wood-dust,

shangmian you zimu 'n'
top have alphabet 'n'

'Holmes takes out a small piece of pencil shaving. It has the alphabet 'n' at the top.'

In the second clause, there is no subject, but one can recover the omitted subject by using the object in an object-final alternation (the bolded part).

To summarize Section 2, I synthesize MODEL A: THE ALTERNATION CHOICE OF VERB-DIRECTION CONSTRUCTION DEPENDS ON LENGTH, CONCRETENESS, DEFINITENESS, AND INFORMATION TYPE OF AN OBJECT NP, THE USE OF ASPECT MARKER *LE*, AND ZERO-ANAPHORA.

3. DATA COLLECTION

This study aims at measuring how linguistic variables of interest can contribute to the alternation choice for VDCs. 350 tokens were collected and annotated with linguistic factors for each alternation. Only one verb *na* 'to take' and one direction combination *chu lai* 'come out' are searched in the corpora.

I planned to use Chinese Treebank 6.0 (CTB) (Xue et al. 2005) and Chinese Gigaword (Huang 2009) as the main source of corpora. However, they did not have enough data for this specific search task. For example, I found only 20 tokens of object-middle alternation in Chinese Gigaword, and no result of object-initial alternation in CTB. In order to find plenty of instances, other Modern Chinese corpora/search engines were used, as in Table 1.

TABLE 1. CORPORA AND SEARCH ENGINES USED IN THIS STUDY

Name	Size	Genre
Corpus in Center of Chinese Linguistics in Beijing University (CCL) (Zhan et al. 2003)	477 million characters	written
Chinese National Broadcast Language Media Resource	241 million characters	spoken
Chinese National Corpus (Xiao 2012)	19 million characters	written
Central China Normal University Corpus	44 million	written
Corpus zhTenTen11 (Jakubíček 2013)	1 billion words	written
Google, Baidu, & Sougou		

The collected 1,050 tokens came from different genres and formats, and are not balanced corpus data. For example, some tokens do not have a minimal context for the discursive annotation, which may have influence on the analysis of features in Section 2.5 and 2.6.

4. REVISED SET OF VARIABLES

Based on my observation, five more factors complementary to those in Section 2 are considered in my hypothesis. I AUGMENT THE ORIGINAL SIX FACTORS IN MODEL A BY ADDING THE FOLLOWING FIVE NEW FACTORS (4.1-4.5), AND IN SO DOING I TAILOR MODEL A TO MODEL B.

4.1. LOCATION/SOURCE (THEMATIC ROLE)

According to the argument structure of the verb ‘to take,’ the typical thematic role assignment is ‘AGENT take P₁ P₂ THEME.’ However, other thematic roles, such as location and source, also appear in the alternations. The expressions of location or source are usually featured in the preposition *cong* ‘from,’ postposition *li/limian/zhong* ‘inside,’ or both. WHEN P₁ AND P₂ ARE NOT ADJACENT (THE OBJECT-MIDDLE ALTERNATION), LOCATION AND SOURCE ARE LESS LIKELY TO APPEAR.

4.2. QUANTIFICATION

The object can be modified by quantifying expressions, such as classifiers (i.e., *ge* ‘the general classifier’), quantifiers (i.e., *xuduo* ‘many’ or *yixie* ‘some’), and numbers, or the object itself is conceptually a quantity such as certain amounts of money (i.e., *ji qian wan yuan* ‘ten million yuan’). IN OUR DATA, THE OBJECT-FINAL ALTERNATION IS MOST LIKELY BE QUANTIFIED.

4.3. IMPERATIVES

Imperatives are ‘verb forms or construction types that are used to directly command the addressee to perform some action’ (Payne 1997: 303). LÜ (1992) PREDICTS THAT ONLY THE OBJECT-FINAL TYPE PREVENTS IMPERATIVES, while the other two can have imperative functions. If there is no imperative for objective-final alternation in the data, then it is likely that this statement is true.

4.4. MODAL VERBS

Modal verbs are (auxiliary) verbs that can deal with judgments and evidentials, and so on (Palmer 2001: 100). Like English, Chinese also has a relatively close set of modal verbs, such as *ying/gai/yinggai/yingdang* ‘should,’ *bixu* ‘must,’ and *neng/nenggou* ‘can,’ as in 8.

- (8) taiwan dang-ju yinggai na chu shiji xingdong lai
 Taiwan current-government should take P₁ practical action P₂
 ‘Taiwan government should take out practical action.’

IT IS POSSIBLE THAT OBJECT-MIDDLE ALTERNATION IS MORE CORRELATED TO MODAL VERBS THAN THE OTHER TWO TYPES. I use the wordlists from Peng 2005 and Fan et al. 1987 as a matching guideline to find possible modal verbs.

4.5. REQUEST VERBS

LIKE MODAL VERBS, IF THERE IS A VERB FOR REQUEST IN THE MAIN CLAUSE, THEN THE OBJECT-MIDDLE ALTERNATION IS MORE LIKELY TO APPEAR. There are two semantic sub-classes for request: verbal and non-verbal. If an imperative speech act is realized in the complement clause, a verbal request in the main clause is expected, such as ‘to order’ and ‘to require.’

- (9) ta-men yaoqiu pingguo na chu yixie shuju lai
 3P-PL require apple take P₁ some data P₂
 ‘They require that Apple take out some data.’

The non-verbal request class includes verbs like ‘to hope’ and ‘to expect,’ in which a speaker has a will to change other people’s action. The English translation of Chinese request verbs is roughly mapped to Classes order-60-1, urge-58.1, and wish-62 in English VerbNet (Kipper et al. 2008).

5. HYPOTHESES, VARIABLE ASSIGNMENT AND STATISTICAL TECHNIQUES

The hypotheses of this study is based on the eleven predictions in the alternation preference in Model B. Compared to Model A which has fewer parameters, I hope Model B can have a better explanation on the variation of the 1,050 tokens in our data.

Before any factorial analysis is performed, I need to translate various linguistic factors into variable types and values that are computable.

TABLE 2. VARIABLE ASSIGNMENT

Variables	Variable type	Value
alternation type	nominal	1='obj-initial,' 2='obj-middle,' and 3='obj-final'
syllable length of object	interval	any positive integer
concreteness of object	nominal	0='concrete,' 1='abstract'
definiteness of object	nominal	0='indefinite,' 1='definite'
aspect marker	nominal	0='no marker le,' 1='marker le'
information structure	nominal	0='old,' 1='new'
zero-anaphora	nominal	0='no,' 1='yes'
thematic role	nominal	1='source,' 2='location,' and 3='neither'
quantifying expression	nominal	0='no,' 1='yes'
imperative	nominal	0='no,' 1='yes'
modal verb	nominal	0='no,' 1='yes'
request verb	nominal	0='no,' 1='yes'

After the collected tokens were annotated with standards in Table 2, I use IBM SPSS software to make monofactorial and multifactorial analyses (Morgen et al. 2012; Sweet et al. 1999). The correlation coefficients are *lambda* and *eta*, which separately deal with the relationship between nominal variables and the alternation types, and that between nominal variables and interval variables. Multinomial logistic regression analysis will tell us how one factor value predicts the choice of an alternation type.

6. RESULTS

6.1. DESCRIPTIVE STATISTICS

Table 3 includes the counts and their percentage in the sample tokens. I tested whether our sub-hypotheses (2.1-2.6 & 4.1-4.5) are well indicated in the counts and recorded the results in the last column. IT IS OBVIOUS THAT THE FUNCTION OF ZERO-ANAPHORA IS REJECTED AND THAT IMPERATIVES DO NOT APPLY TO THE OBJECT-FINAL ALTERNATION. However, counts are easily affected by extreme values, and cannot tell us any correlationship, which makes monofactorial and multifactorial methods necessary.

TABLE 3. BASIC COUNTS I

		Obj-initial	Obj-middle	Obj-final	Good match?
thematic roles	source (1)	4%(13/350)	1%(4/350)	5%(18/350)	Y
	location (2)	11%(38/350)	3%(11/350)	7%(25/350)	Y
abstract obj (1)		15%(54/350)	63%(222/350)	8%(27/350)	Y
definite obj (1)		7%(23/350)	4%(14/350)	5%(18/350)	?
ASP marker (1)		15%(51/350)	0%(1/350)	1%(4/350)	Y
old information (1)		27%(94/350)	3%(12/350)	11%(38/350)	?
zero anaphora (1)		9%(31/343)	3%(9/348)	4%(13/348)	N
quantifying expression (1)		61%(212/350)	34%(120/350)	87%(306/350)	Y
imperative (1)		8%(28/350)	5%(19/350)	0%(0/350)	Y
modal verb (1)		15%(52/350)	47%(163/350)	2%(8/350)	Y
request verb (1)		7%(24/350)	22%(76/350)	0%(2/350)	Y

Counts for the length of objects alternations is in Table 4.

TABLE 4. BASIC COUNTS II (SYLLABLE LENGTH OF OBJECTS)

Alternation	1 to 5	6 to 10	11 to 15	More than 16	Total
obj-initial	326	23	0	1	350
obj-middle	259	70	17	4	350
obj-final	248	81	15	6	350
total	833	174	32	11	1050

The distribution pattern supports Sun's observation on the influence of the object length in Section 2.1. Almost all the objects for object-initial alternation (349 out of 350) have less than five syllables, while object-middle and final alternations can allow some objects longer than 11 syllables.

6.2. MONOFACTORYAL RESULT

The first monofactorial technique I use is *lambda*, which computes the correlation between any of the nominal linguistic factors and the alternation type. $\text{Pr}(>|z|)$ shows whether the correlation is significant. If it is less than 0.001 or 0.01, the correlation is reliable. Table 3 indicates that except for thematic roles (location and source) and definiteness, other linguistic factors do have a low correlation with the choice of alternations. The second monofactorial technique is *eta*, which is not included in Table 5. The *eta* value for syllable length is 0.35, which means that there is an intermediate correlation between the syllable length of objects and the alternation choice.

TABLE 5. RESULTS OF MONOFACTORIAL ANALYSIS

	Lambda	Pr(> z)	Significance	Good match?
abstract obj	0.279	0.000	***(<0.001)	Y
quantifying expression	0.266	0.000	***(<0.001)	Y
old information	0.122	0.000	***(<0.001)	Y
request verb	0.106	0.008	**(<0.01)	Y
ASP marker	0.072	0.000	***(<0.001)	Y
thematic roles	0.059	0.111	ns	N
modal verb	0.037	0.000	***(<0.001)	Y
zero anaphora	0.032	0.000	***(<0.001)	Y
definite obj	0.011	0.756	ns	N

Although monofactorial analysis establishes a steady relationship between alternation choice and linguistic factors, it has weaknesses too. First, *lambda* only works for any single factor and alternation preference. That being said, it is likely that the interrelationship between two factors (i.e., the correlation between concreteness and quantifying expressions) may be underestimated (Gries 2001 & 2003). Since abstract nouns rarely combine with quantifying expressions, the correlation between quantifying nouns and alternation choice may not be reliable. I need a multifactorial method that can rule out the effect of possible mediators and support a cognitively realistic account of alternations.

6.3. MULTIFACTORIAL RESULT

Following the tradition of using regression models (Bresnan et al. 2007), multinomial logistic regression (MNR) is used as a multifactorial method in this study, and shows the logistic coefficient between each linguistic factor (independent variable) and alternation preference (dependent variable) in an interrelation. MNR usually assumes ‘one outcome category as a default case against which other categories are contrasted’ (Mari-Sanna et al. 2012: 5). All significant results are shown in Table 6.

TABLE 6. RESULTS OF THE MULTIFACTORIAL ANALYSIS

Reference	Dependent variable	Independent variable	Weight	Pr(> z)	Odds ratio
C	A	information (0)	1.567	<0.001***	4.794
		quantifying expression (0)	1	<0.001***	2.717
		length of NP	-0.384	<0.001***	0.681
		modal verb(0)	-2.477	<0.001***	0.084
		request_verb(0)	-2.53	<0.01**	0.08
		ASP marker (0)	-3.137	<0.001***	0.043
		imperative(0)	-17.68	<0.001***	2.10E-08
	B	quantifying expression (0)	2.09	<0.001***	8.082
		request_verb(0)	-0.395	<0.001***	0.019
		concreteness(0)	-1.724	<0.001***	0.178
		modal verb(0)	-3.776	<0.001***	0.023
B	A	information (0)	2.064	<0.001***	7.878
		request_verb(0)	1.419	<0.001***	4.135
		modal verb(0)	1.299	<0.001***	3.666
		concreteness(0)	1.248	<0.001***	3.483
		quantifying expression (0)	-1.09	<0.001***	0.336
		ASP marker (0)	-2.763	<0.01**	0.063
		length of NP	-2.95	<0.001***	0.745
	C	request_verb(0)	3.95	<0.001***	51.928
		modal verb(0)	3.776	<0.001***	43.632
		concreteness(0)	1.724	<0.001***	5.61
A	B	length of NP	0.089	<0.05*	1.093
		quantifying expression (0)	-2.09	<0.001***	0.124
		ASP marker (0)	2.763	<0.01**	15.852
		information (0)	2.064	<0.001***	1.27E-01
		quantifying expression (0)	1.09	<0.001***	2.975
		length of NP	0.295	<0.001***	1.343
		concreteness(0)	-1.248	<0.001***	0.287
	C	modal verb(0)	-1.299	<0.001***	0.273
		request_verb(0)	-1.419	<0.001***	0.242
		ASP marker (0)	3.137	<0.001***	23.033
		request_verb(0)	2.53	<0.01**	12.558
		modal verb(0)	2.477	<0.001***	11.902
		length of NP	0.384	<0.001***	1.468
		quantifying expression (0)	-1	<0.001***	0.368
		information (0)	-1.567	<0.001***	-0.209

The ‘reference’ contains different alternations as a baseline for comparison. For example, ‘reference = C’ and ‘dependent variable = A’ mean that relative to C, A has a certain tendency. Also, in this table, if the odd ratio is bigger than one, then the weight should be a positive number, which tells us the dependent variable has a positive effect on the alternation choice. Otherwise there would be a negative effect. The full interpretation of Table 6 is shown in Section 7 as the conclusion.

7. CONCLUSION

This study first introduces Model A (including eleven factors), and improves Model A into Model B by adding five more linguistic features. The basic aim is to use factorial analysis to testify the hypotheses in Model B. The monofactorial analysis shows that when there is no variable interrelationship, all linguistic factors correlate with the alternation preference, except for definiteness and the co-occurrence of location and source. However, multifactorial analysis gives us much more information about the weight of linguistic variables when any two of the three alternations are compared. The influence of zero-anaphora is rejected. Based on Table 6, I can draw a conclusive sketch for this study.

First, the object-initial and middle alternations are compared relative to the object-final alternation. The distinguishing factors of an object-initial alternation are: a shorter object, the aspect marker *le*, new information in the object, imperative, no quantifying expression, modal verbs and request verbs. The distinguishing factors of an object-middle alternation are an abstract object, no quantifying expression, modal verbs, and request verbs.

Second, the object-initial and final alternations are compared relative to the object-middle alternation. The distinguishing factors of an object-initial alternation are: a shorter and more concrete object, the aspect marker *le*, old information in the object, quantifying expression, and no modal verbs or request verbs. The distinguishing factors of an object-final alternation are longer and more concrete objects, quantifying expressions, and no modal verb or request verb.

Third, the object-middle and final alternations are compared relative to the object-initial alternation. The distinguishing factors of object-middle alternation have longer and more abstract object objects, no aspect marker or quantifying expression, old information in the object, modal verb, and request verb. The distinguishing factors of an object-final alternation are a longer

object, the aspect marker *le*, new information in the object, quantifying expressions, modal verb, and request verb.

IN GENERAL, I FIND THAT MOST LINGUISTIC FACTORS IN MODEL B (EIGHT OUT OF ELEVEN) CAN HELP US DISCRIMINATE THE THREE VERB-DIRECTION ALTERNATIONS.

8. FURTHER DIRECTIONS

The main contribution of this study is to firstly use statistical correlation and regression methods to discover how each syntactic, semantic, discursive, or phonological factor contributes to the choice of a verb-direction alternation, which is commonly believed to be interchangeable among three alternations. There are several directions I can improve this study and apply the result to other linguistic studies.

First, I plan to apply the same method to other types/senses of VDCs. According to Liu (1998), there are other directional verbs which involve multiple motion types such as ‘raising,’ ‘moving into,’ and ‘moving up/down.’ Object displacement can work for some of them. Also, VDCs can not only encode special movement, they can also express different resultative and aspectual interpretations. It is highly likely that even for the same direction particle(s), different meanings will bring different results in factorial analyses.

Second, the role of verb semantics in this study is not fully discovered. In order to get an in-depth analysis of alternation choice, I just choose one verb *na* ‘to get’ for neutralization. It is well known that different verbs conceptually evoke different argument structures. For example, the Chinese verb *zou* ‘to walk’ can also combine with P_1 and P_2 , but the thematic role of its object can be an Agent, which *na* ‘to get’ can never do. More verbs need to be collected in order to testify the effect of thematic roles in alternation choice. Also, semantic classification of verbs remains another question. For example, I can compare the factorial analysis results among verbs in various unrelated classes, using VerbNet. Also, there exists an interrelationship between verb’s semantic classes. Synsets with prototypical verbs like ‘to hope’ or ‘to ask’ do have a restriction on the alternations. These classes are positively correlated to the emergence of an object-middle alternation. It is still unknown how this interrelationship can improve the designing work of a computational lexicon, such as Chinese VerbNet.

Third, I disprove three factors in Model B, resulting in the development of Model C. Model C can be further used to discriminate more alternations related to VDCs, such as those carrying

passive marker *bei* and disposal marker *ba*. More interestingly, in our search of the CCL corpus, there is a new alternation whose form is ‘Verb P₁ Object P₂ P₃,’ which appeared in low frequency and was not discussed by the previous references. It is likely that this alternation will be acquired and used by more Chinese speakers. Factorial analysis will help us discover how one language evolves by looking at the change in alternation choice.

Finally, from the perspective of natural language processing (NLP), our statistical method in multifactorial analysis needs to be improved. The output of regression in SPSS is a comparison between two alternations. This ‘two against one’ technique should be changed so it can be compatible with the standard multi-class classification in NLP. Additionally, the frequency distribution of VDCs is equally important as the coefficients in this study.

REFERENCES

- Chen, Xinchun. 1982. Tong fuhe buyu bingjian de quxiang buyu de weizhi (The position of object in co-occurrence with compounding direction particles), *Newsletter of Chinese Language* 5.
- Dehé, Nicole. 2002. *Particle verbs in English: syntax, information structure and intonation* (Vol. 59). John Benjamins Publishing.
- Fan, Xiao; Du, Gaoyin; and Chen, Guanglei. 1987. *Han yu dong ci gai shu (An investigation on Chinese verbs)*. Shanghai: Shanghai Education Press.
- Huang, Chu-Ren. 2009. Tagged Chinese Gigaword Version 2.0, LDC2009T14. Linguistic Data Consortium.
- Gries, Stefan. 2001. A multifactorial analysis of syntactic variation: particle movement revisited. *Journal of quantitative linguistics* 8(1). 33-50.
- Gries, Stefan. 2003. *Multi-factorial analysis in corpus linguistics*. Continuum International Publishing Group.
- Jackendoff, Ray. 2002. English particle constructions, the lexicon, and the autonomy of syntax. *Verb-particle explorations*. 67-94.
- Kipper, Karin; Anna Korhonen; Neville Ryant; and Martha Palmer. 2008. A large-scale classification of English verbs. *Language Resources and Evaluation Journal* 42(1). 21-40.
- Lamarre, Christine. 2007. The linguistic encoding of motion events in Chinese: With reference to cross-dialectal variation. *Typological Studies of the Linguistic Expression of Motion Events* 1. 3-33.

- Liu, Yuehua. 1998. *Quxiang buyu tongshi (Chinese directional particles: An Investigation)*. Beijing: Beijing Language and Culture University Press.
- Lu, Jianming. 2002. Dongci hou quxiang buyu de weizhi wenti (*On the positions of the directional particles after the Verbs*). *Chinese Teaching in the World* 1.5-17.
- Lü, Shuxiang. 1992. Tongguo duibi yanjiu yufa (Using Comparison to Study Chinese Grammar). *Chinese Language Teaching and Research* 2.4-18.
- Jakubiček, M., Kilgarriř, A., Kovář, V., Rychlý, P. and Suchomel, V., 2013, July. The tenten corpus family. In 7th International Corpus Linguistics Conference CL. 125-127.
- Mari-Sanna, Paukkeri; Väyrynen, Paukkeri,; and Arppe, Antti. 2012. Exploring extensive linguistic features sets in near-synonym lexical choice. *Computational Linguistics and Intelligent Text Processing: 13th International Conference, CICLing 2012, New Delhi, India, March 11-17, 2012, Proceedings II*. 1-12
- Morgan, George A., Nancy L. Leech, and Karen C. Barrett. 2012. *IBM SPSS for introductory statistics: Use and interpretation*. Routledge.
- Niu, Yanmin. 2002. *A Word Order Study on Chinese Verb-direction construction*. (Master's thesis). China Capital Normal University (Beijing, China).
- Palmer, Frank Robert. 2001. *Mood and modality*. Cambridge University Press.
- Payne, Thomas E., and Thomas Edward Payne. 1997. *Describing morphosyntax: A guide for field linguists*. Cambridge University Press.
- Peng, Lizhen. 2005. *On modality of modern Chinese*. (Doctoral dissertation). Fudan University (Shanghai, China).
- Sun, Yanling. 2012. *A Study of Order of Compound Directional Complements and Objects in Teaching Chinese as a Second Language*. (Master's thesis). Beijing University.
- Sweet, Stephen A., and Karen Grace-Martin. 1999. *Data analysis with SPSS*. Allyn & Bacon.
- Wasow, Thomas. 2002. *Postverbal behavior*. Center for the Study of Language and Inf.
- Xiao, Hang. 2012. *Guojia yuwei xiandai hanyu yuliaoku jieshao* (An Introduction of Modern Chinese Corpus by the Institute of Applied Linguistics of Ministry of Education in PRC). Online: <http://corpus.zhonghuayuwen.org/resources/CorpusIntroduction2012.pdf>
- Xue, Naiwen, Fei Xia, Fu-Dong Chiou, and Marta Palmer. 2005. 'The Penn Chinese TreeBank: Phrase structure annotation of a large corpus.' *Natural language engineering* 11 (2). 207-238.
- Zhan, Weidong., Guo, Rui., Chen, Yirong., 2003. The CCL Corpus of Chinese Texts: 700 million

Chinese Characters, the 11th Century B.C. - present, Available online at the website of Center for Chinese Linguistics (abbreviated as CCL) of Peking University. Online:

http://ccl.pku.edu.cn:8080/ccl_corpus

Zhang, Bojiang. 1991. Dongqushi li binyu weizhi de zhiyue yinsu (*On the restricting factors of objects in Chinese verb-direction constructions*). *Chinese Learning* 6. 4-8.

APPENDIX – ABBREVIATIONS

3P: the third person

ASP: aspect

ASSOC: associative *de*

CL: classifier

NEG: negation

PL: plural

P_n: the n-th directional particle