

## **Copyright Undertaking**

This thesis is protected by copyright, with all rights reserved.

### By reading and using the thesis, the reader understands and agrees to the following terms:

- 1. The reader will abide by the rules and legal ordinances governing copyright regarding the use of the thesis.
- 2. The reader will use the thesis for the purpose of research or private study only and not for distribution or further reproduction or any other purpose.
- 3. The reader agrees to indemnify and hold the University harmless from and against any loss, damage, cost, liability or expenses arising from copyright infringement or unauthorized usage.

## IMPORTANT

If you have reasons to believe that any materials in this thesis are deemed not suitable to be distributed in this form, or a copyright owner having difficulty with the material being included in our database, please contact <a href="https://www.lbsys@polyu.edu.hk">lbsys@polyu.edu.hk</a> providing details. The Library will look into your claim and consider taking remedial action upon receipt of the written requests.

Pao Yue-kong Library, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong

http://www.lib.polyu.edu.hk

# A COMPARABLE CORPUS-BASED STUDY OF GRAMMATICAL VARIATION BETWEEN DIFFERENT VARIANTS OF MANDARIN CHINESE

JIANG MENGHAN

PhD The Hong Kong Polytechnic University 2018

# THE HONG KONG POLYTECHNIC UNIVERSITY Department of Chinese and Bilingual Studies

A Comparable Corpus-based Study of Grammatical Variation between Different Variants of Mandarin Chinese

# JIANG MENGHAN

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

August 2017

## **CERTIFICATE OF ORIGINALITY**

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

\_\_\_\_\_(Signed)

\_\_JIANG MENGHAN\_\_\_\_\_(Name of Student)

## Abstract

This thesis explores the grammatical variations between two of the most important variants of Mandarin Chinese, i.e. Mainland China Mandarin and Taiwan Mandarin. The lexical differences between cross-strait are well documented from different perspectives (e.g., Huang, Simon and Hsieh, 2007; Li, 2010), while the studies of grammatical variations are relatively less advanced. The lack of comprehensive and systematic syntactic variation studies to a large extent is due to the complexity as well as the probabilistic nature of grammatical variations. Therefore this thesis adopts a comparable corpus-based statistical approach, based on the framework of probabilistic syntactic model. Two grammatical constructions -- light verb construction and  $VO_1+O_2$ construction are investigated.

For the study of light verb construction, the statistical result of mixedeffect logistic regression model and Chi-square test both show that light verbs from different variants can be effectively distinguished by annotated features. Taiwan and Mainland light verbs display variations not only in the syntactic type of the taken complement, but also in the degree of transitivity. Taiwan light verb tends to have higher degree of transitivity, in terms of transitivity frequency as well as semantic/syntactic properties of the taken complements. Variations in word order have also been observed in light verb construction, the theme of event tends to be appeared after the light verb in Taiwan Mandarin while the theme in Mainland Mandarin significantly prefer to be appeared before the light verb.

For the study of VO compounds, the syntactic differences can also be discovered by mixed-effect logistic regression model as well as Chi-square test. Taiwan VO compounds have shown a higher degree of transitivity, in both transitivity frequency and semantic/syntactic properties of the taken objects. The correlation between transitivity and separation ability is also examined to prove the grammatical variations of VO compounds are to some extent dependent on the degree of lexicalization of these compounds. The relationship between transitivity and alternative competing pattern is also examined to further study the transition mechanism of VO compounds as well as to provide the basis for study of correlation between language change and language variation (based on the Lexical Diffusion Theory (e.g., Wang, 1969)). A multilevel Competition Model is also proposed to account for the variation differences observed in corpus data. The correlation between transitivity and word frequency is investigated to further discover the reason that may affect VO transitivization. The analyses indicate that alternative pattern is the most important factor which can influence the transition rate of VO compound, while word frequency only matters when it does not involve internal competition between alternations.

Overall, in this thesis, comprehensive account of grammatical variations on two grammatical constructions between Mainland and Taiwan Mandarin are provided. Large scale comparable corpora, as well as the statistical tool are very effective in underlining the complex and subtle syntactic differences. Furthermore, the results of both constructions show that there is a close interaction between language alternation and language variation, in the sense that grammatical variations are not as simple as feature or behavior difference, but in fact difference in the patterns of alternations. Therefore, the correlation between language variation and language change is also possible to be explored.

## Publications arising from the thesis

- Jiang, M. H., Shi, D. X., and Huang, C. R. (2016). Transitivity in Light Verb Variations in Mandarin Chinese – A Comparable Corpus-based Statistical Approach. Proceedings of PACLIC 2016: The 29th Pacific Asia Conference on Language, Information and Computation.
- Jiang, M. H., and Huang, C. R. (2017). Transitivity Variations in Mandarin VO compounds – A Comparable Corpus-based statistical approach. *Proceedings of the 18th Chinese Lexical Semantics Workshop*.
- Jiang, M. H., and Huang, C. R. (2017). Lexicalization, Separation and transitivity: A comparative study of Mandarin VO compound Variations. *Proceedings of PACLIC 2017: The 29th Pacific Asia Conference on Language, Information and Computation.*

## Acknowledgements

My deepest gratitude belongs to my Chief supervisor, Prof. Huang Chu-ren. He is very patient to answer all my na ve questions and always help me to become a better researcher. I would also like to thank Prof. Shi Dingxu, for inspiring me and encouraging me. Dr Yao Yao, Dr. Lin Jingxia and Dr. Xu Hongzhi also deverse my thanks, for discussing with me and offering me workable solutions for my thesis. I also thank my parents, my family and also my friends, who always there to support me. My gratitude also goes to everyone who ever helped me during this process. Please forgive me if I cannot list all of your names.

# TABLE OF CONTENT

Abstract1
Publications arising from the thesis4
Acknowledgements
LIST OF TABLES13
LIST OF FIGURES16
LIST OF ABBREVIATIONS17
Chapter 1 Introduction
1.1 Introduction/Research Background18
1.2 Previous Studies on Language Variation19
1.2.1 Lexical Variation between Mainland and Taiwan Mandarin .21
1.2.2 Grammatical Variation between Mainland and Taiwan Mandarin23
1.3 Motivation and Significance of the Study26
1.4 Significance of this Dissertation27
1.5 Organization of the Thesis
Chapter 2 Theoretical Framework and Methodology
2.1 Theoretical Framework
2.1.1 Probabilistic Syntactic Model
2.1.1.1 Review the Work of Bresnan et al. (2007a, 2008)31
<ul><li>2.1.1.2 Probabilistic Syntactic Model in Underling Chinese data 35</li></ul>
2.1.1.3 Motivation
2.1.1.4 Constructions to be Explored in this Dissertation37
2.1.1.5 Variation in Constructions
2.1.2 Correlation among Language Alternation, Language Variation and Language Change– from the Perspective of Lexical Diffusion Theory
2.1.2.1 Correlation between Language Alternation and Language Variation
2.1.2.2 Correlation between Language Alternation and Language Change

2.1.2.3 Correlation between Language Variation and Language Change
2.2 Methodology44
2.2.1 Corpora
2.2.1.1 Introduction of Annotated Chinese Gigaword Corpus
2.2.1.2. Acadamia Sinica Palancad Corrus of Modern Chinese
2.2.1.2 Academia Sinica Balanced Corpus of Wodern Chinese
2.2.2 Comparable Corpus-based Approach
2.2.3 Data Extraction and Collection
2.2.4 Data Annotation
2.2.5 Test for the Validation and Reliability of Annotation Schema
2.2.6 Statistical Tests
Chapter 3 Light Verb Variations
3.1 Introduction55
3.2 Previous Studies on Light Verbs
3.2.1 Light Verb Studies in English
3.2.2 Previous Studies on Chinese Light Verbs / Light Verb Studies in Chinese
3.2.2.1 Light Verb Alternation on 进行 jinxing and 加以 jiayi
3.2.2.2 Previous Research on Chinese Light Verb Variation 66
3.2.3 Unsolved Problems
3.2.4 Definition of Chinese Light Verb in this Study
3.2.4.1 Syntactic Type of Light Verb in Mandarin Chinese69
3.2.4.2 Definition of Light Verb in Mandarin Chinese
3.2.4.2.1 Differentiation of TLV and VAV usages in Chinese Light Verbs
3.2.4.2.2 Criteria of Data Collection74
3.2.4.2.3 The Range of Light Verbs76
3.3 Methodology77

3.3.1 Data Extraction77
3.3.2 Data Annotation78
3.4 Data Analysis and Statistical Tool80
3.4.1 Mixed-effect Logistic Regression Model80
3.4.2 Chi-square Test
3.4.2.1 Light Verb Alternations: Variety Internal Comparison
3.4.2.1.1 Differentiation between 进行 jinxing and 加以 jiayi91
3.4.2.1.2 Difference between 做 zuo and 搞 gao
3.4.2.1.3 Difference between 展开 zhankai and 开展 kaizhan98
3.4.2.1.4 Summary101
3.4.2.2 Light Verb Variation: Variety External Comparison101
3.4.2.2.1 进行 <i>jinxing</i> in Taiwan and Mainland101
3.4.2.2.2 加以 <i>jiayi</i> in Taiwan and Mainland103
3.4.2.2.3 做 <i>zuo</i> in Taiwan and Mainland105
3.4.2.2.4 搞 <i>gao</i> in Taiwan and Mainland107
3.4.2.2.5 展开 <i>zhankai</i> in Taiwan and Mainland108
3.4.2.2.6 开展 kaizhan in Taiwan and Mainland110
3.4.3 Summary
3.5 The Syntactic Type of Taken Complement112
3.5.1 Analysis and Discussion112
3.5.2 Application and Implication120
3.5.3 Summary
3.6 Transitivity Variations in Light Verb Constructions
3.7 Variation in Alternation of Light Verb Construction127
3.7.1 Introduction127
3.7.2 Research Question and Methodology129
3.7.3 The First Research Question131
3.7.4 The Second Research Question134
3.7.4.1 Annotation Scheme134

3.7.4.2 Data analyses and discussion
3.7.4.2.1 For Light Verb 进行 <i>jinxing</i> in Mainland138
3.7.4.2.2 For Light Verb 加以 jiayi inTaiwan140
3.7.4.2.3 For Light Verb 进行 jinxing in Taiwan141
3.7.5 Correlation between Alternation Choice and Length of Theme
3.7.6 Commonality and Difference146
3.8 Summary of Chapter 3148
Chapter 4 Grammatical Variations in VO compound150
4.1 Introduction
4.2 Literature Review on the Transitivization of VO Compounds152
4.2.1 Literature Review on the Study of VO <sub>1</sub> +O <sub>2</sub> Construction in Mandarin
4.2.1.1 Research Gap of Studies on VO <sub>1</sub> +O <sub>2</sub> Construction 158
4.2.2 Literature Review on VO <sub>1</sub> +O <sub>2</sub> Variations between Mandarin Varieties
4.2.2.1 Research Gap of Variations Studies
4.2.3 Summary
4.3 Study 1: Transitivity Variations in Mandarin VO compounds163
4.3.1 Grammatical variations in transitivity frequency of VO compounds
4.3.2 Grammatical Variation in Semantic/Syntactic Properties171
4.3.2.1 Annotation Schema171
4.3.2.2 Methodology
4.3.2.3 Mixed-effect Logistic Regression Model179
4.3.2.4 Chi-square Test
4.3.3 Transitivity Degree in VO Variations
4.3.4 Summary of Study 1198
4.4 Study 2: Lexicalization and Transitivity199
4.4.1 Introduction199
4.4.2 Literature Review

4.4.2.1 Research Gap20	)4
4.4.3 Methodology20	)5
4.4.3.1 Data Collection and Calculation20	)5
4.4.3.1.1 Measurement of Separation Ability20	)5
4.4.3.1.2 Measurement of Transitivity Degree	)6
4.4.3.2 Statistical Method and Tool20	)8
4.4.4 Data Analysis and Result20	)8
4.4.4.1 Test 1: Comparison between Separable an Inseparable VO Compounds	ıd )8
4.4.4.2 Test 2: Separation Frequency and Transitivity21	1
4.4.4.3 Summary for the Two Tests21	3
4.4.5 Grammatical Variation and Lexicalization21	.4
4.4.6 Summary of Study 222	20
4.5 Evolutionary Mechanism of VO Transitivization and Languag Variations – Correlation among Language Alternation, Languag Variation and Language Change	ge ge 21
4.5.1 Introduction	21
4.5.2 Lexical Diffusion Theory22	25
4.5.2.1 Introduction to Lexical Diffusion Theory22	25
4.5.2.2 Lexical Diffusion Theory in Syntactic Change22	28
4.5.2.3 Lexical Diffusion in Transitivization of V Compound	0 29
4.5.3 Transition Patterns of VO Compound23	\$1
4.5.4 Research Question23	\$2
4.5.5 Methodology23	3
4.5.5.1 Data Collection and Annotation23	3
4.5.5.2 Statistical Tests and Tool23	\$5
4.5.6 Results and Discussions23	\$5
4.6.6.1 Result of Taiwan Data23	\$5
4.5.6.2 Result of Mainland Data23	;9
4.5.6.3 Line Chart to Present Tendency – Variation Difference	ce 10

4.5.7 The Evolutionary Route for VO Compound243
4.5.7.1 For VC Alternation:
4.5.7.2 For AV Alternation:
4.5.7.3 No Alternative Pattern
4.5.7.4 Both AV and VC Alternation:
4.5.8 Variation Differences in Group 1 – Multilevel Competition Model
4.5.9 Other Factors that May Affect the Process of Change257
4.5.10 Summary
4.6 Study 4: The Correlation between Transitivity and Word Frequency
4.6.1 Introduction
4.6.2 Literature Review
4.6.2.1 For Sound Change
4.6.2.1.1 Changes Affecting High-Frequency Words Earlier
4.6.2.1.2 Changes Affecting Low-Frequency Words Earlier
4.6.2.2 For Syntactic Change
4.6.3 Research Question
4.6.4 Transitivity and Word Frequency
4.6.5 Word Frequency and Alternation Type
4.6.5.1 Alternation Preference for Alternation Group 1271
4.6.5.2 Why Alternation 1 has Such Tendency?276
4.6.6 Summary276
4.7 Summary of Chapter 4
Chapter 5 Conclusion
5.1 Summary
5.1.1 Probabilistic Nature of Grammatical Variations
5.1.2 Probabilistic Syntactic Model
5.1.3 The Interaction between/of Language Alternation and Language Variation
5.1.4 Higher Transitivity of Taiwan Mandarin

5.1.5 Language Conservatism Issue	
5.1.6 Variation Differences in Word Order	
5.2 Possible Applications in Computational Linguistics	290
5.3 Future Studies	291
5.3.1 Correlation between Language Variation and Change for the Study of Light Verbs	l Language
5.3.2 Other Constructions in Other Languages	292
APPENDIX I	294
APPENDIX II	295
References	296

## LIST OF TABLES

Table 2.1 The 14 explanatory variables for annotation	32
Table 2.2 Model A accuracy	33
Table 2.3 The schematized version of lexical diffusion via synchronic	
alternation	42
Table 2.4 Intergroup agreement of syntactic features	51
Table 2.5 Intergroup agreement of semantic features	52
Table 2.6 Intergroup agreement of discourse feature	52
Table 2.7 Average intergroup agreement	52
Table 3.1 Differences in syntactic properties between 进行 jinxing and 加以	
jiayi	63
Table 3.2 Semantic Features that differentiate 进行 jinxing and 加以 jiayi	64
Table 3.3 The difference in complexity of object between 进行 jinxing and 加	
以 jiayie	65
Table 3.4 Differentiation between TLV and VAV in English (Kearns, 1998)	72
Table 3.5 The list of Verbal classifier in Mandarin Chinese (Huang and Ahrens	5,
1995)	76
Table 3.6 The annotation standard of light verb construction	80
Table 3.7 Result of logistic regression model for light verbs  8	83
Table 3.8 Interpretation of standardized person residuals	91
Table 3.9 Comparison between Mainland 进行 jinxing and 加以 jiayi	92
Table 3.10 Comparison between TW 进行 jinxing and 加以 jiayi	94
Table 3.11 Comparison between Mainland 做 zuo and 搞 gao	96
Table 3.12 Comparison between TW 做 zuo and 搞 gao	97
Table 3.13 Comparison between Mainland 展开 zhankai and 开展 kaizhan9	99
Table 3.14 Comparison between Taiwan 开展 kaizhan and 展开 zhankai10	00
Table 3.15 Comparison between 进行 jinxing in ML and TW10	02
Table 3.16 Comparisonbetween 加以 jiayi in ML and TW10	04
Table 3.17 Comparison between 做 zuo in TW and ML10	06
Table 3.18 Comparison between 搞 gao in TW and ML10	07
Table 3.19 Comparison between Mainland and Taiwan 展开 zhankai10	09
Table 3.20 Comparison between Taiwan and Mainland 开展 kaizhan1	10

Table 3.21 Transitivity Parameters (Hopper and Thompson, 1980: p.252)123
Table 3.22 The 5 types to introduce the theme of event131
Table 3.23 Result of Chi-square test for 进行 <i>jinxing</i> and 加以 <i>jiayi</i> in two
varieties133
Table 3.24 New annotation features for alternative study
Table 3.25 Chi-square test for syntactic choice of Mainland 进行 jinxing140
Table 3.26 Chi-square test for syntactic choices of 加以 jiayi140
Table 3.27 Result of Chi-square test for Taiwan 进行 jinxing142
Table 3.28 Result of Kruskal Wallis test for Mainland 进行 jinxing144
Table 3.29 Result of Kruskal Wallis test for Taiwan 进行 jinxing144
Table 3.30 The Post hoc comparisons of Taiwan 进行 jinxing145
Table 3.31 Result of Kruskal Wallis test for Taiwan 加以 jiayi145
Table 4.1 Likelihood ratio of 过境 guojing and 借道 jiedao168
Table 4.2 Result of likelihood ratio filter  168
Table 4.3 Examples of words whose likelihood ratio >=10169
Table 4.4 Examples of contrast difference  170
Table 4.5 Annotation schema for VO compound  173
Table 4.6 The result of logistic regression model for VO compound181
Table 4.7 Result of Chi-square test for VO compound
Table 4.8 Variation differences in syntactic type  189
Table 4.9 Number of separable/inseparable VO in both Taiwan and Mainland
Mandarin
Table 4.10 The result of Mann Whitney U test for TW data210
Table 4.11 The result of Mann Whitney U test for Mainland data210
Table 4.12 The result of Spearman's rank-order correlation for Taiwan data212
Table 4.13 The result of Spearman's rank-order correlation for Mainland data
Table 4.14 Comparison of average separation/transitivity frequency214
Table 4.15 Separation differences based on Z-test
Table 4.16 Separation comparison of 把关 baguan
Table 4.17 Examples of contrast difference between Taiwan and Mainland 219
Table 4.18 A schematized version of lexical diffusion via synchronic variation
Table 4.19 Type of alternations 234

Table 4.20 The result of Kruskal-Wallis test for 8 groups in Taiwan23	37
Table 4.21 The post hoc comparison for 8 groups in Taiwan23	37
Table 4.22 The result of Kruskal-Wallis test for 468 groups in Taiwan23	38
Table 4.23 The result of Kruskal-Wallis test for 123 groups in Taiwan23	38
Table 4.24 The post hoc comparison for group123 in Taiwan23	39
Table 4.25 The result of Kruskal-Wallis test for 468 groups in Mainland23	39
Table 4.26 Post hoc comparison for 123 groups in Mainland	40
Table 4.27 Comparison of mean, mean rank and median between Group 1 and	13
	50
Table 4.28 Alternation preference for words of Group 1 in Taiwan25	54
Table 4.29 Alternation preference for words of Group 1 in Mainland	
(transitivity <5%)	55
Table 4.30 The result of Spearman's correlation in Taiwan	65
Table 4.31 The result of Spearman's correlation in Mainland20	65
Table 4.32 Comparison of Taiwan transitivity between high and low frequency	у
words	67
Table 4.33 Comparison of Mainland transitivity between high and low	
frequency words	67
Table 4.34 Alternation type 1, 2 and 3	69
Table 4.35 Comparison of transitivity for each group in Mainland27	70
Table 4.36 Comparison of transitivity for each group in Taiwan27	70
Table 4.37 Alternation preference for Taiwan data	73
Table 4.38 Alternation preference for Taiwan data  2'	74

## LIST OF FIGURES

Figure 2.1 The model of formula (A)	33
Figure 2.2 Interaction of recipient animacy with variety: New Zealand spectrum	oken
('nz'), US spoken ('s') and US written ('w')	34
Figure 3.1 Frequency distribution of 进行 <i>jinxing</i> and 加以 <i>jiayi</i> in two	
varieties	132
Figure 4.1 Result of Z-test	167
Figure 4.2 Synchronic clines of lexicality and grammaticality	202
Figure 4.3 Line chart with log for Group 2 and 3 in both varieties	241
Figure 4.4 Line chart with log for Group 1 in both varieties	243
Figure 4.5 the transition process for Group 3	247
Figure 4.6 The transition process for Group 2	248
Figure 4.7 Multilevel Competition Model	252
Figure 4.8 Line chart with log for Group 1 in Taiwan	255
Figure 4.9 Line chart with log for Group 1 in Mainland	256

## LIST OF ABBREVIATIONS

1 SG	First person singular pronoun
1 PL	First person plural pronoun
2 SG	Second person singular pronoun
2 PL	Second person plural pronoun
3 SG	Third person singular pronoun
3 PL	Third person plural pronoun
BA	Marker of disposal construction ba
BEI	Passive marker <i>bei</i>
VP	Verbal Phrase
NP	Nominal Phrase
DE	prenominal modifier marker, post-verbal resultative marker, SFP
LE	perfective marker <i>le</i>
ZHE	progressive markerzhe
GUO	experiential markerguo
MEI	negatormei
ZHENG	progressive markerzheng
ZHENGZAI	progressive markerzhengzai
CL	classifier
XIAQU	progressive marker <i>xiaqu</i>
*	Ungrammatical sentence
?	Odd sentences

## **Chapter 1 Introduction**

### **1.1 Introduction/Research Background**

Mandarin or Mandarin Chinese, or simply Chinese nowadays, is the one of the most commonly learned first or second languages in the world now. It is the official language for Mainland China, Hong Kong and Taiwan, one of the four official languages in Singapore.

Although Mandarin Chinese spoken in different regions have the same linguistic heritage, variations and differences between different variants do exist because of political, economic, cultural and social reasons (Huang, Lin and Zhang, 2013). The variation differences have been noticed by some Chinese linguists. Numbers of work have been conducted to help resolve any misunderstanding problems and to facilitate better communication among different Chinese-speaking communities (Huang et al., 2013).

This dissertation mainly focuses on two of the most important variants of Mandarin Chinese, Mainland Mandarin and Taiwan Mandarin. The lexical differences between cross-strait have become a popular topic among linguistics for many years (e.g., Xu, 1995a; Zeng, 1995; Wang and Li, 1996). Grammatical difference of these two variants, however, was not well studied beyond anecdotal observation, partly because of the near identity of their grammatical systems. In this dissertation, a large-scale comparable corpus-based statistical approach will be adopted to systematically investigate the syntactic variation between Mainland and Taiwan Mandarin. Comprehensive and definitive accounts of two widely used constructions (Light Verb Construction and  $VO_1+O_2$  Construction) will be provided. Theoretical issues related to degree of transitivity, process of lexicalization, etc., will be discussed. Correlation among language variation, language alternation and language change will also be explored based on the empirical tendency.

### **1.2 Previous Studies on Language Variation**

Previous studies on Language variation are conducted from different perspectives: sociolinguistics, theoretical linguistics as well as computational linguistics. Language variation is a very core concept in sociolinguistics. Sociolinguistic variation is the study of the way language varies and changes in communities of speakers and concentrates in particular on the interaction of social factors (e.g., a speaker's gender, ethnicity, age, degree of integration into their community) and linguistic structures (e.g., sounds, grammatical forms, intonation features, words) (Britain, 2016). The most well-known work is conducted by William Labov (1966/2006), in which he analyzes the social stratification of English in New York City, based on detailed observations of phonological features. He demonstrated that linguistic variation is socially stratified, such that the use of pronunciation features (e.g., dropping of postvocalic /r/ correlates with social class, ethnicity, etc. in regular patterns (Hazen, 2011). Systematic multivariate statistical model together with the notion of linguistic variable has also been generated, and first implemented by Cedergren and Sankoff (1974). The Labovian model of patterned variation has been confirmed in numerous studies of sociolinguistic variations (Labov and Wang,

2014).

Besides studies in sociolinguistics, plenty of grammarians in theoretical linguistics have also been interested in the study of grammatical variations in varieties of English or other languages. Bresnan and Hay (2008) argue that since the grammatical variations are usually subtle and gradient, they are better to be observed by quantitative analyses. Therefore, logistic regression model was adopted in their paper to investigate syntactic variation between New Zealand and US English on the 'give'dative construction. The result of logistic regression shows that NZ English is subtly different from US English in that NZ English speakers appear to be more sensitive to the role of animacy. Details of this study will be discussed in Section 2.1 of Chapter 2.

In addition to the studies in thereotical linguistic field, computational linguistic also concern about how language varies. Researches on automatic identification and classification of similar languages have been conducted in different languages. For instance, Doğruöz, Nakov's and Tower (2014) show that light verb constructions can be automatically classified between Turkish spoken in Netherlands and Turkish spoken in Turkey.

Apart from the studies of language variations in other languages, the variations between different Mandarin speaking communities, especially the differences between Mainland and Taiwan Mandarin, have also attracted the interests of previous scholars. The variation studies between Mainland and Taiwan Mandarin are usually conducted in two perspectives: lexical variation differences and grammatical variation differences.

#### **1.2.1 Lexical Variation between Mainland and Taiwan Mandarin**

In these years, a variety of studies have been conducted for lexical variations between Mainland and Taiwan, such as region-specific neologism, meaning variations of the same word, and the use of different words to express the same meaning (Huang et al., 2013).

A number of dictionaries have been compiled to collect and record the lexical differences. Qiu (1990) is among the first to collect a list of cross-strait variations in lexical. Wei and Sheng's dictionary (2000) focuses on the cross-strait differences in word form, sound, and semantic meaning. Shi, Li and Li (2003) compile a larger scale dictionary that contains more than 40,000 common entries for Mainland and Taiwan Mandarin; with additional 1,300 Mainland specific entries, and 1,000 Taiwan specific entries. More recently, aiming to promote better understanding and to standardize Modern Chinese word usage, the *New Century Global Chinese Dictionary* (Li, 2010) collects about 10, 000 entries that are used in more than seven Chinese-speaking communities, including Mainland China and Taiwan.

In addition to dictionaries, plenty of case studies have also been conducted on the regional variations of lexicon between Mainland and Taiwan. For instance, Tang (1995) and Zhao (2008) provide examples showing how one concept is expressed differently between Mainland and Taiwan Mandarin, as well as among other varities. Wang and Li (1996), Xu (1995a), Xu (1995b), Zeng (1995), among many others, discoverthe lexical differences between Mainland and Taiwan Mandarin. Even lexical differences in transliteration of particular fields have been investigated: including Mao (2002) on movie titles, Sheng (1996) on computer terminologies, and Zhou (2009) on person and place names.

While most of the studies introduced above are based on researchers' introspection and observation, a number of lexical comparisons have also been carried out based on corpus, between Mainland and Taiwan. In recent years, the availability of comparable corpora (i.e. two or more corpora with similar topics) of different varieties of Chinese has enabled wider coverage and more comprehensive lexical studies (Lin, Shi, Jiang and Huang, to appear). For example, corpus-based lexical comparisons between Mainland and Taiwan have been carried out on topics such as thesaurus (Kwong and Tsou 2006, 2008; Tsou and Kwong, 2006), Chinese news headlines (Chin, 2007), celebrity coverage in the Chinese press (Tsou, Yuen, Kwong, Lai and Wong., 2005), judgment terms (Kwong and Tsou, 2004) and automotive patents (Lu and Tsou, to appear, among others).

Even statistics or computational methods are adopted to compare the lexical differences cross strait. For instance, Huang, Simon and Hsieh (2007), as well as Simon, Huang, Hsieh and Hong (2007) adopt bootstrapping cooccurrence statistics from segmented and tagged Chinese corpus to automatically identify diverging transliterations of foreign named entities into Mainland and Taiwan Mandarin (Huang et al., 2013). Hong and Huang (2008), among others, develop methodologies to detect the lexicons and their collocations that are used differently in Mainland and Taiwan Mandarin. With these research outputs, it is fair to say that lexical variations between Mainland and Taiwan Mandarin are well documented.

### 1.2.2 Grammatical Variation between Mainland and Taiwan Mandarin

The studies of grammatical variation between Mainland and Taiwan, on the other hand, are relatively less advanced. A good number of existing grammatical studies focus on observations on individual grammatical constructions.

Wei (1984) and Kubler (1985) focus on the development of Mandarin in Taiwan and point out several grammatical distinctions between Taiwan Mandarin and Standard Mandarin, including 有 you 'have'+VP, increase in the use of 会/不会 hui/buhui 'can/cannot', use of 来/去 lai/qu 'come/go' and 来去 laigu 'come and go', use of 给 gei 'give', 用 yong 'use' with nominalized verbs, E zai 'at' as an auxiliary verbs, deletion of the second syllable of bi-syllabic verbs before  $\overline{\wedge}$  bu 'BU', use of  $\overline{\neg}\overline{\neg}$  vixia 'one time', omission of 的 de 'DE', use of 样子 yangzi 'the look', final 没 mei 'MEI' and 不 bu 'BU', increase in use of 而已 ervi 'nothing more', coexist of VV 看 VV kan 'have a look' and V 看看 V kankan 'have a look' and so on. Huang (1988) also investigates the variation of several grammatical constructions based on Taiwan novel: reduplication (also see Liao, 1999), use of classifier, comparative sentence, 有 you 'have' construction, 用 yong 'use' construction, 到 dao 'reach' construction and so on. Diao (2000, 2012a, 2012b, 2012c, 2012d, 2012e, 2013) has

conducted a number of studies on the variation differences in different grammatical constructions between Mainland and Taiwan Mandarin, including 有 you 'have'+VP (2012g), 获 huo 'receive' construction (2012a), 遭 zao 'suffer' and 被 bei 'BEI' construction (2012c, 2013), 将 jiang 'will' construction (2012f), the use of 而已 eryi 'nothing than' (2012d) and so on. Zhu and Zhou (1990) mainly focus on the transitivity of VO compound, different use of clitics, reduplications of adjectives and modal verbs between Mainland and Taiwan Mandarin. Qiu and Fan (1994) describe 有 you 'have' construction, V+neg+V question, word order of double object construction. Zhang's (1994) work mainly focuses on comparative sentences, 有 you 'have' construction, 被 bei 'BEI' construction and V-not-V question. Zhao and Shi (2014) talk about the different usages of 有信心 you xinxin 'have confidence' construction in Hong Kong, Taiwan and Mainland Mandarin.

Besides these, other constructions have also been included in the study of grammatical variation: 透过 *touguo* 'trough' (Chu and Zhang, 2013), classifier 通 *tong* 'CL' (Tsou and Mo, 2013), use of degree adverbs (Li, 2015; Fang, 2015), mood marker (Fang, 2013), demonstrative marker (Fang, 2014) and so on. Dictionaries are complied to record the grammatical differences. Global Chinese Grammar (Xing, to appear) is the first that aims to comprehensively describe the grammar of Mandarin Chinese used in six regions (including Taiwan, Hong Kong, Macau, Singapore, Malaysia, and USA).

Apart from the observations on individual grammatical constructions, Diao

(1998a) also discusses the variation differences in language characteristic/property between Mainland and Taiwan Mandarin. He points out that there are several characteristics of Taiwan Mandarin: the residue of classical Chinese and Japanese, the mixing of different grammatical forms and the influence of Southern Min. Other studies also mention the influence of Southern Min, English for Taiwan Mandarin (Dai and Zhao, 2009; Mao, 2007; Huang, 1988 and Kubler, 1985; among others).

The existing studies of grammatical variation between Mainland and Taiwan share three features. One feature is most of the studies consider the variation differences as contrast differences, meaning that Taiwan and Mainland are using the constructions in different ways. But according to the corpus observation, variation differences in grammatical construction to a large extent tend to differences in tendency/preference instead of dichotomy contrast. In that sense, the language variations are more appropriate to be studied in a continuous and gradient way (Tseng, 2003). The second feature is, in terms of methodology, although more and more recent studies rely on authentic data such as corpora for identifying possible variation (e.g., Fang, 2013, 2014, 2015), only a few exceptions (e.g., Huang, Lin, Jiang and Xu, 2014; Lin, Xu, Jiang and Huang, 2014) adopt computational and statistical tools to process the data. In addition, the third feature is, most of the studies are still at the descriptive level, while linguistic exploratory generalization and analyses are very necessary.

I agree with Huang et al (2013) that, the lack of comprehensive and systematic grammatical variation studies between Mainland and Taiwan, to a

large extent is due to the nature of grammatical variation. The target of grammatical variation studies are larger units such as phrases, constructions, and sentences. The aim is to investigate whether diversity exists in distributional and selectional constraints on how these units are constructed. Compared to the lexical differences, grammatical variations are more complex and thus, technically more difficult to be detected. Moreover, grammatical variations are very subtle to be observed in the sense that they more tend to be differences in frequency or preference (more conventionalized versus less used) instead of un/grammatical differences (Huang et al., 2013).

### **1.3 Motivation and Significance of the Study**

As introduced above, unlike lexical differences, less work has been conducted on grammatical variations between Mainland and Taiwan Mandarin. The complexity and subtleness of grammatical differences make it difficult to detect the variations by traditional approach of introspective judgment. Hence a more objective, quantitative, data driven research methodology is needed to identify and analyze the frequency of collocational differences (i.e. whether the two varieties differ in selecting types of collocating arguments and whether diversity exists in selectional constraints) and/or the preference of certain alternatives among different varieties (Lin et al., to appear). Nowadays, with the availability of comparable corpora, e.g., LIVAC (T'sou and Kwong, 2006) and Tagged Chinese Gigaword (Huang, 2009), and computational/statistical analytical tools, it is possible now to effectively identify the subtle and tendency differences among multiple Chinese varieties and to precisely describe the nature and motivation of variations (Lin et al., to appear)). In this dissertation, a comparable corpus based, statistical modeling based approach is adopted to examine the grammatical variations between Mainland and Taiwan Mandarin. Theoretical issues including degree of transitivity, word order, syntactic type, and degree of lexicalization will be discussed. The correlation among language alternation, language variation and language change will also be investigated based on the statistical results.

Two grammatical constructions --- light verb construction and  $VO_1+O_2$  construction are explored. These two constructions are very important in traditional linguistic studies and have been studied by many researchers from different perspectives (e.g., Zhu, 1985; Cai, 1982). The variations of these two constructions between different Mandarin variants have also been observed by previous studies (e.g., Huang and Lin, 2013; Huang et al., 2013; Wang, 1997; Diao, 1998), but have not been comprehensively and empirically studied.

### **1.4 Significance of this Dissertation**

**Empirically,** comprehensive account of grammatical variations on two grammatical constructions between Mainland and Taiwan Mandarin will be provided. The two widely used and dynamic constructions will be carefully examined, and solid empirical basis for the description of grammatical variations will be given.

Methodologically, probabilistic syntactic model and gradient grammar

concept (Bresnan et al., 2007a, 2007b, 2008, 2010) will be adopted to underline syntactic variation in Mandarin Chinese, based on large scale comparable corpora. In other words, validate statistical models for linguistic studies of language variation based on comparable corpora will be provided, thus facilitate more synergy between corpus/computational linguistics and theoretical linguistics. With the assistance of this model, some linguistic problems which are unlikely to be solved/discovered by traditional approach (e.g., grammatical variations) will be able to explore.

**Furthermore, theoretically,** different theoretical issues such as syntactic type, argument structure, degree of transitivity, word order and degree of lexicalization will be discussed based on the empirical tendencies. The correlation among language alternation, language variation and language change will also be explored on the basis of Lexical Diffusion Theory. Another crucial contribution of this thesis is that it shows that variations can be studied and captured in terms of aggregation of linguistic features as well as patterns of alternation.

### **1.5 Organization of the Thesis**

The rest of the dissertation is organized as follows. Chapter 2 introduces the methodology and the theoretical framework of the thesis. The motivation of adopting probabilistic syntactic model as well as Lexical Diffusion Theory will be discussed. The merits of the approach and the methodology of date collection, corpus construction, data annotation and statistical test selection will

also be introduced. The study of light verb construction will be provided in Chapter 3. Comprehensive descriptive accounts of light verb variations will be given. Theoretical issues such as variation in degree of transitivity, syntactic type of the taken complement and word order will also be studied. The study of VO compounds will be presented in Chapter 4. The syntactic differences are shown to be fully discovered by mixed-effect logistic regression model as well as Chi-square test. Theoretical issues including the variation difference in transitivity degree, the correlation between lexicalization and transitivity, the transition mechanism of VO compounds will also be explored. The relationship between transitivity and alternative competing pattern is also examined to further study the transition mechanism of VO compounds as well as to provide the basis for study of correlation between language change and language variation (based on the Lexical Diffusion Theory (e.g., Wang, 1969)). A summary of the thesis, together with the direction for future studies will be discussed in Chapter 5.

## **Chapter 2 Theoretical Framework and Methodology**

### **2.1 Theoretical Framework**

#### 2.1.1 Probabilistic Syntactic Model

As mentioned in Chapter 1, the studies of grammatical variations are less advanced. The lack of comprehensive and systematic grammatical variation studies to a large extent is due to the complexity and subtleness of syntactic differences. Furthermore, the grammatical variation has a gradient nature in the sense that it tends to be the difference in frequency or preference instead of un/grammatical difference (Huang et al., 2013), hence it is not easy to be detected by traditional observation of introspective judgment. Then the question should be addressed is, how can the variation in tendency be comprehensively studied and which approach should be adopted?

The gradient nature of grammatical variation leads us to the probabilistic/gradient grammar model. Probabilistic grammar, which lies in at the heart of probabilistic linguistics, always considers grammar to be quantitative and has the nature of gradient (i.e. gradient grammar). Probabilistic grammars are usually used to describe the probabilistic nature of a vast number of linguistic phenomena, such as phonological acceptability, morphological alternations, syntactic well-formedness, semantic interpretation, sentence disambiguation and sociolinguistics variation as well as language variation (Bod and Hay, 2003).

Several studies have already been conducted based on probabilistic grammar model from different perspectives (Bod and Hay, 2003). One good

example of using probabilistic syntactic model to solve traditional linguistic problems is the series of studies conducted by Bresnan et al. (2007a, 2007b, 2008, 2010). Bresnan (2007a) in the paper points out that theoretical linguists have traditionally relied on linguistic intuitions such as grammaticality judgments for their data, however, linguistic intuitions of grammaticality are deeply flawed and seriously underestimate the space of grammatical possibility.

Therefore, their studies were conducted based on the probabilistic models of grammar, which assume that grammar is quantitative and learned from exposure to other speakers. The concept of 'gradient grammar' they adopted (grammar has the nature of gradient), is also consistent with usage-based, probabilistic syntactic models. The grammar can be considered as a continuum instead of a binary system. Under this model, Bresnan et al. (2007a, 2007b, 2008, and 2010) have conducted a series of studies on English alternative dative structures. They have shown that language alternations as well as language variations can be effectively distinguished and predicted by modern statistical models. Also, linguistic data are shown to be much more probabilistic than has been widely recognized in theoretical linguistics.

#### 2.1.1.1 Review the Work of Bresnan et al. (2007a, 2008)

The basic model and the method of probabilistic grammar is first presented in Bresnan et al. (2007a), in which they have shown that the logistic regression model can predict United States (US) English speakers' syntactic choice with 'give'-type verbs extremely accurately (with more than 90% accuracy).

In English, there are two ways of expressing the dative --- prepositional dative double structure object structure. The problem of or traditional/theoretical linguistics in studying the alternative dative structures is, they usually separate grammar with actual usage. Some of the ungrammatical sentences shown in theoretical studies are actually very nature in actual usage. In that case, intuition is no longer reliable. Large quantity of data is needed to investigate how does people use this construction in natural language (Bresnan, 2007a).

Therefore their studys elected 2360 instances of dative structures used by speakers in the full Switch-board collection of recorded and transcribed telephone conversation. All the data were annotated based on 14 predicting explanatory variables which may influence the syntactic choice of alternative dative structures (shown in Table 2.1, originally in Bresnan, 2007a: p.13).

#### Response modeled as depending on

semantic class + accessibility of recipient + accessibility of theme + pronominality of recipient + pronominality of theme + definiteness of recipient + definiteness of theme + animacy of recipient + person of recipient + number of recipient + number of theme + concreteness of theme + structural parallelism in dialogue + length difference (log scale)

Table 2.1 The 14 explanatory variables for annotation

Table 2.2 (originally in Bresnan, 2007a: p.14) presents that the predicting accuracy provided by Logistic Regression Model is very high, which means the model is very effective and the explanatoryfeatures they choose for annotation are also efficient in making the prediction.
Classification Table for Model A (1 = PP; cut value = 0.50)					
Predicted: % Correct					
	0 1				
Observed:	0	1796	63	97%	
	1	115	386	77%	
Overall: 92%					

% Correct from always guessing NP NP (=0): 79%

Table 2.2 Model A accuracy

The result of logistic regression model also shows that different factors/features have different preference for certain structures, as shown in Figure 2.1 (originally in Bresnan, 2007a: p.17). Longer, nominal, inanimate, or non-given recipients favor the dative PP construction; otherwise favor the double NP constructions.

 $\label{eq:robability} \{ \text{Response} = 1 \} = \frac{1}{1 + e^{-X\beta}}, \ \text{where}$  $X\hat{\beta} =$ 0.95  $-1.34\{c\} + 0.53\{f\} - 3.90\{p\} + 0.96\{t\}$ (a) +0.99{accessibility of recipient = nongiven} (a) −1.1{accessibility of theme = nongiven} (b) +1.2{pronominality of recipient = nonpronoun} (b) −1.2{pronominality of theme = nonpronoun} (c) +0.85{definiteness of recipient = indefinite} (c) -1.4{definiteness of theme = indefinite} (d) +2.5{animacy of recipient = inanimate} +0.48{person of recipient = nonlocal} -0.03{number of recipient = plural} +0.5{number of theme = plural} -0.46{concreteness of theme = nonconcrete} (e) -1.1{parallelism = 1}  $-1.2 \cdot \text{length difference (log scale)}$ 

and  $\{c\} = 1$  if subject is in group c, 0 otherwise (and likewise for other categories).

# Figure 2.1 The model of formula (A)

What are more important, Bresnan and Hay (2008) also shows that such a model can be used to predict syntactic differences between language variants. Since the language variations are usually subtle and gradient, they are expected

to be observed not only in the use of clear dialectal features, but also in extremely subtle factors like the relative probabilistic weights of conditioning factors (Bresnan and Hay, 2008). In this study (i.e. Bresnan and Hay, 2008), Logistic Regression Model was also used to investigate syntactic variation between New Zealand and US English on the 'give' dative construction. Apart from the annotated US data (in Bresnan et al., 2007a), another 2842 tokens from New Zealand corpus were also annotated with the same predicting factors. The different sources of data were distinguished by adding another variable. The results of logistic regression are consistent with the ones shown in the earlier studies (Bresnan et al., 2007a). NZ English is subtly different from US English in that NZ English speakers appear to be more sensitive to the role of animacy (as shown in Figure 2.2, originally in Bresnan and Hay, 2008, p.9), which supports for the gradient nature of grammar, and are also consistent with usage-based, probabilistic syntactic models.



Figure 2.2 Interaction of recipient animacy with variety: New Zealand spoken ('nz'), US spoken ('s') and US written ('w')

Based on the two studies shown above (Bresnan et al., 2007a, 2008), gradient properties of grammar and language variation (difference in animacy) can both be revealed by the statistical model of the syntax. The variability captured in the statistical model is unlikely to be observed and explained by traditional approach.

# 2.1.1.2 Probabilistic Syntactic Model in Underling Chinese data

Probabilistic Syntactic Model is also proved to be effective in underlying Chinese data. Yao and Liu (2010) present a corpus-based statistical study on Chinese dative variation based on Bresnan's model. Since the dative variation phenomenon in Chinese is more complicated than in English (Yao and Liu, 2010), a two-level hierarchical schema has been adopted in their study. The upper level distinguishes preverbal ditransitive and post-verbal ditransitive while the lower level further differentiates the pre-IO (indirect object) and post-IO (indirect object). The results of Logistic Regression Model show that the upper level distinction is sensitive to the semantic features of direct object, as well as pronominality of the indirect object and the length difference between the two objects; while the lower level is conditioned by the length difference and the presence of a following verb (Yao and Liu, 2010). Yao (2014) further investigates the Chinese BA construction with two frequently verb 拿 na 'take' and 放 fang 'put', to examine which factor will influence the word order variation (BA or SVO). This study is also based on the probabilistic model. The results show that for both verbs, the use of BA construction is highly conditioned by prominence, weight of the object, as well as structural parallelism. The differences of these two models lie in the opposite preference for object NP weights. The model of 放 *fang* 'put' presents a negative effect of object weight on the use of BA construction while 拿 *na* 'take' shows a positive effect.

# 2.1.1.3 Motivation

The idea of probabilistic linguistics provides us a new direction/method for language studies and also a very important model for quantitative language studies. Since previous studies have shown that probabilistic syntactic model is effective in studying grammatical variation (Bresnan and Hay, 2008) and also efficient in underlying Chinese data (Yao and Liu, 2010; Yao, 2014), this model seems to be appropriate to be adopted for this dissertation to study the grammatical variation between Mainland and Taiwan Mandarin.

As mentioned above, grammatical variations are usually subtle differences which are very difficult to be observed by traditional approaches. For example, the light verb 从事 *congshi* 'engage in' in Taiwan sometimes can take informal event complements like 散步等运动 *sanbu deng yundong* take a walk\_and so on\_sport 'to take a walk and so on', while in Mainland such collocations are rare. This kind of differences is usually neglected by linguists. In addition, grammatical variation usually tends to be difference in preference/tendency rather than grammatical/ungrammatical distinction. For example, corpus data shows that the VO compound 把关 *baguan* 'guarantee' in Mainland usually

takes common NP like 产品质量 *chanpin zhiliang* product\_quality 'quality of product' while in Taiwan it prefers to take event denoting complements like 商 品管理 *shangpin guanli* good\_manage 'merchandise control'. However, these differences are not absolute, and examples of taking eventive complements may also be found in Mainland corpus, but with a relatively lower frequency. This is in accordance with the nature of probabilistic syntactic model.

# 2.1.1.4 Constructions to be Explored in this Dissertation

Two widely used and dynamic constructions are explored in the dissertation: light verb construction and VO<sub>1</sub>+O<sub>2</sub> construction. Chinese light verbs like 进行 *jinxing* 'proceed', 加以 *jiayi* 'inflict' are usually semantically bleached and contain no eventive information, the predicative content mainly comes from their taken complements (as in 进行讨论 *jinxing taolun* proceed\_discuss 'to discuss', the predicative information of 'discussion' comes from the object 讨 论 *taolun* 'discuss' instead of the light verb 进行 *jinxing* 'proceed'). VO<sub>1</sub>+O<sub>2</sub> construction is formed by one VO compound and one external complements taken by the compound, such as 把关商品质量 *baguan chanpin zhiliang* guarantee\_goods\_quality 'to guarantee the quality of goods'.

These two constructions are very important in traditional linguistic studies and have been studied by many researchers from different perspectives (e.g., Zhu, 1985; Cai, 1982). The variation differences of these two constructions between different variants have also been observed by previous studies (e.g., Wang, 1997; Diao, 1998), but is lack of comprehensively and empirically studied.

## 2.1.1.5 Variation in Constructions

Although Bresnan et al. (2007a, 2007b, 2008, and 2010) have examined language variations by probabilistic syntactic model, the main focus of their studies is language alternation. Yao and Liu (2010) also focus on the syntactic alternations instead of language variations. Different from these previous studies, my dissertation is aiming at exploring the grammatical variation between different language variants. The probabilistic model will be applied from a different perspective: while Bresnan et al., (2007a, 2007b, 2008, 2010) explore which features can influence people's syntactic choices between different alternations, my objective is mainly to investigate which factors may help to distinguish different variations for the same syntactic construction.

Let us take VO<sub>1</sub>+O<sub>2</sub> construction as an example. One VO may have different collocations and distributions in different language variants across regions. For instance, 把关 baguan 'guarantee' in Mainland may mainly take 产 品 质量 / 水 质 NP like chanpin zhiliang/shuizhi product\_quality/water\_quality 'quality of product/water' as the objects while in Taiwan it may prefer to take deverbal nouns as in 把关房地产开发/市场管理 baguan fangdichan kaifa/shichang guanli intendance\_realestate\_develop/market\_management 'to intendance real-estate development/market management'. Therefore the statistical analyses may reveal the significantly distributional difference in the feature 'syntactic type of the taken complement' between Mainland and Taiwan  $VO_1+O_2$  constructions (i.e. the feature 'syntactic type of the taken complement' may be very effective in distinguishing Mainland and Taiwan  $VO_1+O_2$  constructions).

# 2.1.2 Correlation among Language Alternation, Language Variation and Language Change– from the Perspective of Lexical Diffusion Theory

# 2.1.2.1 Correlation between Language Alternation and Language Variation

Bresnan et al. (2007a, 2007b, 2008, and 2010) focus on language alternation, aiming at investigating how to distinguish and predict the syntactic choices between different syntactic alternative patterns. While my study is focusing on language variations, aiming at distinguish and predict the syntactic differences between different language varieties. Basically, these two studies share a common nature. As mentioned above, language variation has the gradient nature and grammatical variations tend to be differences in tendency/preference instead of dichotomy contrast. My thesis will show that, the grammatical variations are not as simple as feature or behavior change, but in fact are differences in the patterns of alternations. In other words, the language variation across regions to a large extent lies in the differences in alternation preferences. The preference of alternative pattern differs in different varieties of the same language. For example, there are different light verb alternations 进行/加以研 究 jinxing/jiayi yanjiu proceed/inflict research 'to conduct research', 做/搞活 动 zuo/gao huodong do/do activity 'to do activity' to express the same meaning 'to conduct research', 'to organize an event'. Mainland and Taiwan may show different preferences in choosing light verbs under different contexts. In addition, I have observed that there are different alternative patterns to introduce the theme of the verbal complement taken by the light verb construction: 对可行性进行研究 dui kexingxing jinxing vanjiu for\_practicality\_proceed\_research/进行可行性研究 *jinxing kexingxing yanjiu* proceed\_practicality\_research/进行(对)可行性的研究 *jinxing (dui)* kexingxing de yanjiu proceed\_(for)\_practicality\_DE\_research/进行对可行性研 究 jinxing dui kexingxing vanjiu proceed for practicality research/进行研究 可行性 jinxing yanjiu kexingxing proceed\_research\_practicality 'to conduct research on practicability'. Mainland and Taiwan have shown variation differences in the alternation choices. Besides that, in terms of transitivity of VO compound, different VO compounds have different alternative patterns to introduce the external objects 媲美澳洲/与澳洲媲美/媲美于澳洲 pimei aozhou rival with\_Australia/yu aozhou pimei with\_Australia\_rival with/pimei yu aozhou rival with\_at\_Australia 'rival with Australia', and variation differences in alternative choices have also been observed between these two varieties.

# 2.1.2.2 Correlation between Language Alternation and Language Change

The relationship between alternation and language change has been clearly studied in the field of evolutionary linguistics, especially by Lexical Diffusion Theory (e.g., Wang, 1969, 1977, 1979).

The Lexical diffusion theory was originally created to describe the process of sound change, but later is applied to study syntactic change as well. The basic idea of lexical diffusion theory is for sound change, it is phonetically abrupt but lexically gradual. We can understand as each individual word undergoes the change at different rates or different times. The review of the theory will be presented in detail later in Chapter 4.

In that sense, Wang (1969, 1977, 1979) argues that Leixcal Diffusion Theory emphasizes the importance of synchronic variations and language change proceeds by alternation. That is, before the replacement of a new rule, we should expect to observe the co-existence of different alternations (i.e. old and new forms co-exist). A schematized version of lexical diffusion via synchronic alternation is given in Table 2.3 (see original table in Wang, 1979: p. 362).

For any  $W_i$ , the notation  $\overline{W}_i$  represents the changed counterpart of that particular word. Hence a typical change involves three stages: u (unchanged), a (synchronic alternation) and c (changed). In Table 2.3,  $W_1$  is the most advanced word, having already reached the c stage.  $W_2$  and  $W_3$  are in a stage, while  $W_4$ and  $W_5$  are still in the u stage.

Stages	U	a	с
Words			
<b>W</b> <sub>1</sub>			$\overline{W}_1$
$W_2$		$W_2 \sim \overline{W}_2$	
<b>W</b> <sub>3</sub>		$W_3 \sim \overline{W}_3$	
$W_4$	$W_4$		
$W_5$	$W_5$		

Table 2.3 The schematized version of lexical diffusion via synchronic alternation

Therefore what is important for language studies is that, synchronic alternation can help us to examine the mechanism in the implementation of language change, whether the actuation was from inside or outside the linguistic system undergoing change. Once we recognize the critical role that synchronic variation plays in lexical diffusion, then such evidence is indeed easily found in every living languages (Wang, 1979). The study of language change might be more much easier to conduct. Although it is difficult to collect sizeable historical documentation, at least the synchronic data do give a time-slice of the change in progress.

# 2.1.2.3 Correlation between Language Variation and Language Change

Numbers of work have been conducted on language variation and language change respectively, but few studies have focused on the relationship between language variation and language change. Now with the help of Lexical Diffusion Theory, it is possible to build the bridge between these two.

As presented above, syntactic alternations provide us the evidence to examine the evolutionary mechanism/route of language change. Language variation to a large extent lies in the differences in alternation preferences. Therefore to examine the alternative preferences in different language variations is very crucial, as this may display the variation differences in the process of language/transition change.

In Chapter 4, the correlation among language alternation, language variation and language change will be carefully examined through the study of VO transitivization.

To be particular, in Mandarin Chinese, some VO compounds can be used transitively and followed by an external object (O<sub>2</sub>) directly, which yield VO1+O2 construction (e.g., 媲美澳洲 *pimei aozhou* rival with\_Australia'rival with Australia'). There are also some other alternative patterns to introduce the external objects (O2): Prep+O2+VO1 (e.g., 与澳洲媲美 yu aozhou pimei with\_Australia\_rival with 'rival with Australia'), VO<sub>1</sub>+(Prep)+O<sub>2</sub> (e.g., 媲美于 澳洲 pimei yu aozhou rival with\_at\_Australia 'rival with Australia'), V+O<sub>2</sub>+O<sub>1</sub> (e.g., 过河南境 guo henan jing pass\_He'nan\_border 'transit through He'nan border'). Mainland and Taiwan Mandarin actually show differences in preferences for these alternative patterns. According to Lexical Diffusion Theory, different preferences for alternations may further imply the variation differences in evolutionary route of VO compound transitivization (VO is transited from intransitive to transitive). Therefore in Chapter 4, the variation differences in alternation preferences as well as transition route between Mainland and Taiwan Mandarin will be explored in detail.

# 2.2 Methodology

As discussed above, due to the gradient/probabilistic nature of grammar as well as the subtleness of grammatical variations, syntactic variations between different variants of Mandarin Chinese are very difficult to be observed by traditional approach. Therefore, probabilisticsyntactic model (Bresnan et al., 2007a, 2007b, 2008, 2010) is adopted to investigate the subtle variation differences from a quantitative/probabilistic perspective. To be specific, the grammatical variations in Light Verb Constructions and VO<sub>1</sub>+O<sub>2</sub> Constructions between Mainland and Taiwan are examined using a comparable corpus-based statistical approach. Further linguistic generalization and explanation will also be provided based on the result of statistical analyses.

# 2.2.1 Corpora

The main corpus used in this dissertation is the Annotated Chinese Gigaword Corpus (Huang, 2009) maintained by LDC which contains over 1.1 billion Chinese words, consisting of 700 million characters from Taiwan Central News Agency (CNA) and 400 million characters from Mainland Xinhua News Agency (XNA). As can be seen, the dataset is large enough to cover different usages and reflects the actual situation of language use. Moreover, the Mainland and Taiwan sub-corpus allow me to make effective comparison between the two varieties. Sinica Balanced Corpus of Modern Chinese is also used as a supplementary corpus (Sinica Corpus user manual, 2006).

# 2.2.1.1 Introduction of Annotated Chinese Gigaword Corpus

Chinese Gigaword was produced by Linguistic Data Consortium (LDC) catalog number LDC2003T09 and ISBN 1-58563-230-9. This is a comprehensive archive of newswire text data that has been acquired from Chinese news sources by the LDC over several years. Two distinct international sources of Chinese newswire are represented here: Central News Agency of Taiwan (can) and Xinhua News Agency of Beijing (xin). All sources have been categorized into four distinct 'types'. (Huang, 2009, p.1):

this type of DOC represents a coherent report on a particular topic or story event, consisting of paragraphs and full sentences this type of DOC contains a series of unrelated 'blurbs', each of which briefly describes a particular topic or event: 'summaries of today's multi news', 'news briefs in ...' (some general area like finance or sports), and so on these are DOCs which the news service addresses to news editors, they advis are not intended for publication to the 'end users' these DOCs clearly do not fall into any of the above types; these are other things like lists of sports scores, stock prices, temperatures around the world, and so on The general strategy for categorizing DOCs into these four classes was, for each source, to discover the most common and frequent clues in the text stream that correlated with the three 'non-story' types. When none of the known clues was in evidence, the DOC was classified as a 'story'.

# 2.2.1.2 Academia Sinica Balanced Corpus of Modern Chinese

In addition, the Sinica Balanced Corpus has also been used as a supplementary

corpus:

'Academia Sinica Balanced Corpus of Modern Chinese', simplified as Sinica Corpus, is the first Balanced Modern Chinese Corpus with part-ofspeech tagging. The new version targeted at 10 million words is completed in 2006. Texts are collected from different areas and classifiedaccording to five criteria: genre, style, mode, topic, and source. Therefore, this corpus is a representative sample of modern Chinese language. Sinica Corpus is developed and maintained by Institute of Information Science and CKIP group in Academia Sinica. The CKIP group is managed by Prof. Keh-jiann Chen (Institute of Information Science) and Prof. Chu-Ren Huang (Institute of Linguistics) (cited from Sinica Corpus user manual, 2006: P1).'

#### 2.2.2 Comparable Corpus-based Approach

According to Sinclair (1996), a comparable corpus is composed of 'similar texts in more than one language or variety' (Huang et al., 2013). The comparable corpus contains texts from different languages or varieties based on the same criteria such as size, genre, and time (Sinclair, 1996). It is different from parallel corpora in the sense that the latter one consists of corpora translated into a different language. According to Huang et al. (2013), the most important advantage of constructing large-scale comparable corpus is that it enables the comparison between different languages or varieties in 'similar circumstance of communication, but avoiding the inevitable distortion

introduced by the translations of a parallel corpus' (Sinclair, 1996) or by direct quotation of data from another variety or language. In that sense, comparative studies based on a large-scale comparable corpus can best objectively discover the variation differences in the usage of the constructions under investigation.

As introduced in Section 2.2.1, the Annotated Chinese Gigaword corpus contains over 1.1 billion Chinese words, consisting of data from both Central News Agency (Taiwan, about 700 million characters), and Xinhua News Agency (Mainland, about 400 million characters) (Huang, 2009). The availability of Chinese Gigaword corpus makes it possible for the construction of comparable corpora for grammatical variation investigation (Huang et al., 2013).

# **2.2.3 Data Extraction and Collection**

For both Light Verb Construction and VO compound, in both Mainland and Taiwan, the data was extracted from the Chinese Annotated Gigaword Corpus.

For light verb constructions, 7 verbs were included in data extraction: 进行 *jinxing* 'proceed', 加以 *jiayi* 'inflict', 做 *zuo* 'do', 搞 *gao* 'do', 从事 *congshi* 'engage in', 展开 *zhankai* 'carry out' and 开展 *zhankai* 'carry out'. The 7 light verbs were chosen because they are usually considered as the most frequently/commonly used light verbs among previous scholars (e.g., Diao, 2004) and can be divided into near-synonym pairs (e.g., 进行 / 加以 *jinxing/jiayi* 'procced/inflict', 做 / 搞 *zuo/gao* 'do/do' and 开展 / 展开 *kaizhan/zhankai* 'carry out') for further analyses. For each light verb in each

variety, about 200 tokens were randomly extracted from corpora, which result in about 2800 tokens in total. For VO compound, 171 verbs were selected from previous researches (e.g., Wang, 1997; Diao, 1998; Liu andLi, 1998; Qian, 2011). Since there are numbers of debates on the syntactic status of VO (i.e. whether it is a word or a phrase), to avoid the ambiguity, intersection was made between these 171 verbs and the wordlist of Gigaword and Sinica corpus (to ensure the VO is considered as a word in each variety). The procedure of intersection can also ensure the word is not unique in one variety (i.e. the word is used in both Taiwan and Mainland). After intersection, a wordlist which contains 155 VO compounds is generated for data extraction. For each VO compound in each variety, around 1000 tokens were randomly extracted from the Gigaword corpus. If the VO does not have enough 1000 tokens in corpus, I include as many as possible in the corpus. In total, more than 200000 tokens were extracted from both Mainland and Taiwan sub-corpus.

#### 2.2.4 Data Annotation

For light verb construction and VO compounds, different features (including features at semantic, syntactic and discourse levels) were selected to differentiate between different varieties.

For light verb construction, 14 features were selected from previous studies (e.g., Zhu, 1985; Zhou, 1987a; Zhou, 1987b; Cai, 1982; Huang et al., 2013, among others), to help identifying different light verbs as well as light verb variations. These features cover semantic, syntactic as well as discourse

levels. The annotation schema will be explained in detail in Chapter 3. All of the extracted 2800 sentences were manually annotated with the 14 features. The main annotator is a trained expert on Chinese linguistics. All ambiguous cases were discussed with another two experts in order to reach an agreement. Another two annotators were also included to test the validation and reliability of the annotation schema/result (which will be discussed in detail in Section 2.2.5).

For the study of VO compound, transitivity frequency for each VO in each variety was calculated before the data annotation. Since the transitivity frequencies of 46 words were too low (smaller than 0.1%) in both varieties, they were considered as intransitive VO compounds and excluded from further analyses in both varieties. Annotation was done on the other 109 transitive VO compounds.10 potential features (including features at semantic, syntactic and discourse level) were selected to help to identify the syntactic differences between Mainland and Taiwan Mandarin. The annotation schema will be explained in detail in Chapter 3. For each VO compound in each variety, 100 transitive tokens were randomly selected for annotation. For the ones that may not have enough 100 transitive tokens, I include all the transitive tokens as many as possible. Over 20000 transitive tokens were manually annotated by the main annotator and the reliability of annotation schema has also been tested.

# 2.2.5 Test for the Validation and Reliability of Annotation Schema

As mentioned above, all the tokens from both Mainland and Taiwan corpus are annotated with numbers of features by a linguistic expert. To test the validity and reliability of our annotation scheme and annotation result, two more annotators are also involved to annotate a portion of the data, which have already been fully annotated by the main annotator. About 652 examples were randomly selected, and the selection follows the principle that it could cover both Mainland and Taiwan usages. The 652 examples were divided into two parts, one part contains 381 examples and another part contains 271 examples. The two subsets of the corpus are annotated by two different annotators respectively. The two annotators are Year-4 undergraduate students from the Department of Chinese Bilingual Studies in The Hong Kong Polytechnic University, and are Mandarin Speakers with very good linguistic background (have recived systematic linguistic training on Chinese Syntax and Semantics). The two annotators have recived multiple training sessions from the main annotator, to make sure they are familizrized wih the purpose of the annotation as well as annotation scheme, before the test. Firstly, a very detailed annotation instruction was given to the annotators. Secondly, before the real annotation, 100 examples were given to both of them as an exercise. The annotation results of the exercise were carefully discussed by the main annotator and the other annotators, in order to reach agreement. Then the 652 examples were given to the two annotators respectively.

Cohen's Kappa was run to test the inter-rater agreement. Comparison was

done on every linguistic feature. To be specific, the annotations involve 13 features in total, including 7 syntactic features, 5 semantic features and 1 discourse feature. The agreement results are reported as below.

First for the annotation of syntactic feature, as shown in Table 2.4, there is a very good agreement between annotator 1 and the main annotator (average k=0.957, P<0.001). The agreement between annotator 2 and the main annotator is also very good (average k= 0.848, P<0.001). The two datasets are combined to examine the agreement between the two new annotators and the main annotator. The average of k coefficient also reveals a very good agreement (k=0.9272, P<0.001)

	POS	VO	Subje	other	aspe	Accom	Attribut	Average
			ct	lv	ctual	plishm	ive	
					mark	ent		
					er			
Annota	0.913	0.949	1	1	1	0.864	0.971	0.95671
tor 1								429
Annota	0.876	0.729	0.957	0.798	1	0.663	0.913	0.848
tor 2								
Annota	0.897	0.877	0.971	0.940	1	0.852	0.954	0.92728
tor 1+2								5

Table 2.4 Inter-group agreement of syntactic features

The agreement on semantic features is a little bit lower than that on syntactic features, but is still very good, either between annotator 1 and the main annotator (k=0.8414, P<0.001), between annotator 2 and the main annotator (k=0.8378, P<0.001), or between both new annotators and the main annotator (k=0.878, P<0.001).

Type of	Spontaneo	Durative	Formal	Psychol	Interacti	Averag
event	us			ogical	on	e
Annotator	Not valid	0.720	0.709	0.932	0.846	0.8414
1	1					
Annotator	Not valid	0.722	0.577	Not	0.890	0.8378
2	1			valid 1		
Annotator	Not valid	0.720	0.637	0.933	0.877	0.878
1+2	1					

 Table 2.5 Intergroup agreement of semantic features

For the annotation of discourse feature, as presented below, there is very good agreement between annotator 1 and the main annotator (k=0.836, P<0.001). Although the agreement between annotator 2 and the main annotator is moderate (k=0.589, P<0.001), there is a very good agreement between the two new annotators and the main annotator (k=0.807, P<0.001).

Annotator	Polarity of context
Annotator 1	0.836
Annotator 2	0.589
Annotator 1+2	0.807

Table 2.6 Intergroup agreement of discourse feature

If all the features are combined and the average k is calculated, we can see that overall theres areveery good agreements between the three annotators (between annotator 1 and main annotator: k=0.903, P<0.001; between annotator 2 and the main annotator: k=0.824, P<0.001; between annotator 1+2 and the main annotator: 0.882, P<0.001).

	Average
Annotator 1	0.903076923/0.895
Annotator 2	0.824153846/0.792181818
Annotator 1+2	0.881923077
T 11 07 4	• ,

 Table 2.7 Average intergroup agreement

Therefore, it can be summarized as, the inter-rater agreement Cohen's Kappa test among the three annotators can show that the annotation schema and annotation result are very reliable.

# **2.2.6 Statistical Tests**

Following the framework of Bresnan et al. (2007a, 2007b, 2008, 2010), logistic regression model was conducted to differentiate the syntactic variations, for both Light verb construction and VO compound. In my study, I used Mixedeffect Logistic Regression Model instead of the general binary logistic regression model. Since the dataset includes different verbs, the variability in the verbs should not be disregarded. Therefore the verb was treated as a random effect in this model. Using a mixed-effect regression approach ensures the results are generalizable and enabled the study to quantify and examine the variation of individual words. The variation differences between Taiwan and Mainland were analyzed by using mixed-effect logistic regression model (GLMM) (Pinheiro and Bates, 2000) implemented in the R package lme4 (Bates and Maechler, 2009). In addition, to better illustrate the subtle difference of each verb or each pair of verbs, Chi-square test was also included. The statistical tool I used is IBM SPSS V.22. The reason of conducting Chi-square test is because the Mixed-effect Logistic Regression Model can only give a general tendency of the grammatical differences between different variants, but cannot display the specific subtle difference for each verb or each verb pairs.

For each individual study, some other statistical tests were also utilized for different purposes. For the example, for the study of VO compound, Z-ratio test was conducted to compare if there is significant difference in transitivity frequency between Mainland and Taiwan. Likelihood ratio test was also included to measure the degree of differences between varieties. In the study of VO transitivization in Chapter 4, Kruskal-Wallis test was utilized to compare whether there are any statistical significant differences in transitivity frequencies of VO compound between different alternation groups. In the study of lexicalization, a Mann-Whitney U test was run for the data in each variety to determine if there were differences in transitivity frequency between separable and inseparable VO compounds. The Spearman's rank-order correlation was also used to assess the relationship between transitivity frequency and separation frequency in both Taiwan and Mainland Mandarin. All the tests mentioned above will be discussed in detail in the later chapters, respectively. Based on the result of statistical analyses, linguistic analyses were conducted and linguistic explanation will be provided.

# **Chapter 3 Light Verb Variations**

# **3.1 Introduction**

The notion of light verb is originated in Jespersen (1965) to describe English VO constructions like 'have a rest', 'give an advice' and 'do an interview'. The property/charactersitic of light verb is that the verb itself is semantically bleached without containing any eventive information. The predicative information mainly comes from its taken complement. From this perspective, the function of 'do an interview' is very similar to verb 'interview'.

In modern Chinese, there exists a kind of semantically bleached verb which is also called light verb (the most typically used ones are 进行/加以/做/ 搞/从事 *jinxing/jiayi/zuo/gao/cong* 'proceed/inflict/do/do/engage'). They are similar to English light verbs (e.g., take rest, give advice) in the sense that the predicative content mainly comes from its taken complement (e.g., Zhu, 1985) while the light verb itself may only contributes aspectual information, without containing any eventive information. For example, for the construction 进行讨论 *jinxing taolun* proceed\_discuss 'to discuss', the predicative information all comes from the complement 讨论 *taolun* 'dicuss', while the light verb 进行 *jinxing* 'proceed' itself only contributes information about event shape and indicates the event type that the event of 'discuss' is an activity which has a duration/process.

In terms of previous researches, light verb in English has been comprehensively studied in both theoretical (e.g., Jespersen, 1965; Butt and Geuder, 2001) and computational linguistics (e.g., Tu and Dan, 2011; Nagy, Vincze and Farkas, 2013; Hwang et al., 2010). Some literature will be briefly reviewed in the next section (Section 3.2).

However, compared to English light verbs, little work has been done for Chinese light verbs. This may due to several reasons. The first reason may be the semantic impoverishment of Chinese light verbs. As the light verb is semantically bleached, it does not have strong collocation constrains. The Chines light verb usually can take a wide range of objects, including deverbal 进行详细的研究 noun (e.g., jinxing xiangxi de vanjiu proceed detailed DE research 'to conduct comprehensive research'), eventive nouns (进行赛事 *jinxing saishi* proceed game 'play a game'), and sometimes concrete numbers with eventive meaning. Furthermore, while the light verb can take a variety of objects, their morphological statuses of these categories are typically unmarked (Lin et al., 2014). Therefore there are plenty of debates on which syntactic type does the taken complement (e.g., 研究/考量 yanjiu/kaoliang 'research/considerate' in 进行研究/加以考量 jinxing kaoliang proceed reserach/inflict considerate yanjiu/jiayi <sup>•</sup>to conduct research/give consideration') belong to: either it is a deverbal noun (e.g., Zhu, 1985), a verb or a typical noun (e.g., Li, 1990). The lack of morphological status makes the identification and differentiation of light verbs much more complex. Another reason for the lack of Chinese light verb study is that, Chinese light verb are often interchangeable with the same object due to their versatility, e.g., 进行/加以研究 jinxing/jiayi yanjiu 'to conduct research'; 做/搞

科研工作 zuo/gao keyangongzuo do/do research work 'do research'. However, collocation constraints are sometimes found with these light verbs, e.g., 进行/\* 加以会议 jinxing/\*jiayi huiyi proceed/inflict meeting 'have a meeting', \*进行/ 加以提高\*jinxing/jiayi tigao proceed/inflict improve 'make an improvement', \*进行/加以/\*做/\*搞/\*从事考虑 jinxing/\*jiayi/\*zuo/\*gao/\*congshi kaol ii\*proceed/inflict/\*do/\*do/\*engage considerate 'give consideration'. What makes the study of light verbs even more complicated is that, the collocations pattern usually lies in the absence/presence of a tendency, instead of strict constrains/rules (which is consistent with the idea of gradient grammar (Bresnan et al., 2007a, 2007b, 2008, 2010)). For example, 加以 jiayi 'inflict' has a tendency of taking psychological event as the complement (e.g., 加以考 虑/考量 jiavi kaol i/kaoliang inflict considerate 'to considerate'), while other light verbs may also have the usages of co-occurring with psychological event complement (in fact, 进行反省 jinxing fanxing proceed reflect on 'to reflect on' can also be found in corpus), but with a much lower frequency. In that sense, the collocation differences of light verbs are very difficult to be observed by traditional approach, therefore are less studied by previous researches.

From the perspective of language variation, light verb constructions are found in most varieties of contemporary written Chinese (e.g., Mainland Mandarin Chinese, Taiwan Mandarin Chinese, and Hong Kong Mandarin Chinese). However, the actual usages of these constructions have gone different pathways, even though the syntactic structures remain almost the same.

Therefore in different Mandarin varieties, even with the very limited collocation constrain, variations still exist. For example, previous studies (Huang et al., 2013; Diao, 2012b) have observed that Taiwan light verbs tend to take more types of Simple nouns and even VPs as the complements. For instance, light verb construction like 进行开票 jinxing kaipiao proceed\_ballot counting 'to proceed with ballot counting' can only be found in Taiwan corpus. However, besides these findings, to the best of my knowledge, very few investigations have been put forward to explore the variation differences in light verb usages among different variants of Mandarin. The lack of comprehensive and systematic grammatical variation studies in light verb constructions is mainly due to the following two reasons: the first reason is that the light verb variations are very complex and subtle to be observed. For example, according to the corpus investigation, Taiwan light verbs are often observed to be collocated with informal events (e.g., 从事散步等运动 congshi sanbu deng yundong engage walk and so on sport 'to go for a walk') while such collocations are not often be found in Mainland corpus. This subtle difference is easy to be neglected by previous researches. The second reason is that the light verb variations between different variants of Chinese tend to be differences in frequency or preference instead of un/grammatical differences (Huang et al., 2013, which is also consistent with the idea of gradient grammar (e.g., Bresnan et al., 2007a)). For example, 进行 jinxing 'proceed' in both Taiwan and Mainland Mandarin can take aspectual marker, but the frequency of taking aspectual marker in Mainland is much higher than that of in Taiwan.

As illustrated above, the complexity, subtleness and probabilistic nature make the differentiation/identification of light verb as well as light verb variations a very challenging work for both theoretical and computational linguistics in previous researches.

# **3.2 Previous Studies on Light Verbs**

# **3.2.1 Light Verb Studies in English**

As mentioned in Section 3.1, light verb in English has been comprehensively studied from both theoretical linguistics (e.g., Jespersen, 1965; Butt and Geuder, 2001) and computational linguistics perspectives (e.g., Tu and Dan, 2011; Nagy et al., 2013; Hwang et al., 2010).

In theoretical linguistic research, since Jespersen (1965) has generated the term 'light verb' to describe the constructions like 'have a rest', the term has been adopted for analysis in a variety of researches. But light verbs have proven to be a challenge for theories of semantics and syntax, because they have properties that make it difficult to class them either with function words or with lexical verbs (Butt and Geuder, 2001). One approach is to treat light verbs as pure functional words (e.g., copulas in Curme, 1935) that serve as a licenser of predication only (also see Grimshaw and Mester, 1998). Another approach is to identify light verb as a type of auxiliary (e.g., Hook 1974, 1991). The other approach is to consider LVC as a complex predicate in which both light verb and main verb contribute to the argument structure (e.g., Rosen, 1990;

Mohanan, 1994; Butt, 1995, 2010; Butt and Geuder, 2001). In that sense, light verb is differentiated from verbal licenser or auxiliary. In other words, in this approach, light verbs contribute to a semantically complex but syntactically monoclausal predication and they form a syntactically distinct class (Butt, 2010).

In computational linguistics, identification and classification of LVC in different languages (especially in English) have been an important fundamental task for many Natural Language Processing (NLP) applications, such as machine translation and information retrieval. Machine learning models are also proposed to automatically identify English LVCs in context (Tu and Dan, 2011; Nagy et al., 2013). There is also computational research aiming at differentiating LVCs between variants of the same language. For example, Doğruöz, Nakov and Tower (2014) used machine learning technology to automatically identify LVCs between NL-Turkish (Turkish spoken in Netherlands) and TR-Turkish (Turkish spoken in Turkey).

# **3.2.2** Previous Studies on Chinese Light Verbs / Light Verb Studies in Chinese

In Chinese linguistic studies, light verbs are usually called 形式动词 *xingshi dongci* 'dummy/form verb' (Li, 1990; Shen and Zhang, 2013; Zhou, 1987b; Liu, 2005, etc.) 虚化动词 *xuhua dongci* 'dummy verb' (Zhu, 1985, etc.), or 虚义动词 *xuyi dongci* 'light meaning verb' (Diao, 2004). Despite the different

terminologies, they share a very similar definition in the sense that the verb itself appears to be semantically bleached and the predicative information comes from its taken complement (e.g., 进行研究 *jinxing yanjiu* proceed\_research 'to do research', 研究 *yanjiu* 'research' carries the eventive information). The range of light verbs (i.e. which verbs should be considered as light verbs) varies in different studies. While some studies include verbs such as 给以 *jiyu* 'render', 予以 *yuyi* 'give', 展开 *zhankai* 'carry out', 开展 *kaizhan* 'carry out' and even some dialectical verbs like 干 *gan* 'do' and 弄 *nong* 'do', most of studies only focus on the most typical and commonly used light verbs (e.g., 从事 *congshi* 'engage', 做 *zuo* 'do',搞 *gao* 'do'), especially the most frequently used light verb pair 进行 *jinxing* 'proceed' and 加以 *jiayi* 'inflict' (e.g., Tang, 2005; Xie, 1990).

# 3.2.2.1 Light Verb Alternation on 进行 jinxing and 加以 jiayi

Being the most frequently used light verbs, 进行 *jinxing* and 加以 *jiayi* have been the focus of many linguistic researches. These two light verbs are found to differ in two major perspectives: syntactic behavior and collocation constrains (e.g., Zhou, 1987a; Cai, 1982; Tang, 2005).

In terms of syntactic behavior, 进行 *jinxing* behaves more like a typical/common verb compared to 加以 *jiayi* (Zhou, 1987a). The first evidence is 进行 *jinxing* can take aspectual markers like 着 *zhe* 'progressive marker', 了

*le* 'perfective marker' and 过 *guo* 'experiential marker' while 加以 *jiayi* cannot (as shown in Table 3.1). The second evidence is only 进行 *jinxing* can be negated by 没 *mei* 'negation' (shown in Table 3.1). Finally, 进行 *jinxing* allows topicalization of the verbal complement while 加以 *jiayi* cannot (also shown in Table 3.1).

Syntactic	Typical/commo	Light verb 进行	Light verb 加以
behavior	n Verb		
	注意讨论	进行讨论	加以讨论
	zhuyi taolun	jinxing taolun	jiayi taolun
	pay attention	proceed_discuss	inflict_discuss
	to_discuss	'to discuss'	'to discuss'
Take	'pay attention to		
aspectual	the discussion'		
marker	注意了讨论	进行了讨论	*加以了讨论
	zhuyile taolun	jinxingle taolun	*jiayile taolun
	pay attention	proceed_LE_discuss	Inflict_LE_discuss
	to_LE_discuss	'had a discussion'	'had a discussion'
	'have paid		
	attention to the		
	discussion'		
	注意过讨论	进行过讨论	*加以过讨论
	zhuyiguo taolun	Jingxingguo taolun	Jiayiguo taolun
	pay attention	proceed_GUO_discuss	inflict_GUO_discuss
	to_GUO_discus	'have had a discussion'	'have had a
	S		discussion'
	'have already		
	paid attention to		
	the discussion'		
	不注意讨论	不进行讨论	不加以讨论
	bu zhuyi taolun	bu jinxing taolun	bu jiayi taolun
	BU_pay	BU_proceed_discuss	BU_inflict_discuss
Negation	attention	'do not discuss'	'do not discuss'
of light	to_discuss		
verb	'not pay		
	attention to the		
	discussion'		
	没注意讨论	没进行讨论	*没加以讨论
	mei zhuyi taolun	mei jinxing taolun	mei jiayi taolun
	MEI_ pay	MEI_proceed_discuss	MEI_inflict_discuss

	attention	'have not had a	'have not had a
	to discuss	discussion'	discussion'
	'have not paid		
	attention to the		
	discussion'		
	要坚持改革	要讲行改革	要加以改革
	vao iianchi	vao jinxing gaige	vao jiavi gaige
	gaige	need proceed reform	need inflict reform
Topicaliz	need adhere	'need to carry on	'need to carry on
ation of	to reform	reform'	reform'
verbal	'need to adhere		
complem	to reform'		
ent	改革要坚持	改革要讲行	*改革要加以
	gaige vao	gaige vao iinxing	gaige vao jiavi
	iianchi	reform need proceed	reform need inflict
	reform need ad	'reform needs to be	'reform needs to be
	here to	carried on'	carried on'
	'reform needs to		
	be adhered to'		
	要把改革坚持	要把改革进行下去	*要把改革加以下去
	下去	yao ba gaige jinxing	yao ba gaige jiayi
	vao ba gaige	xiaqu	xiaqu
	iianchi xiaau	need_BA_reform_proc	need_BA_reform_inf
	need BA refor	eed_XIAQU	lict_XIAQU
	m adhere	'need to keep carrying	'need to keep
	to XIAOU	on reform'	carrying on reform'
	'need to adhere		
	to reform'		
	要坚持的改革	要讲行的改革	*要加以的改革
	vao ijanchi de	van jingxing de onige	vao jaivi de gaige
		need proceed DE refo	need inflict DF ref
	need adhere	rm	orm
	to DE refrom	'the reform needed to	'the reform needed to
	'the reform	he carried on'	he carried on'
	needed to be		

Table 3.1 Differences in syntactic properties between 进行 *jinxing* and 加以 *jiayi* 

With respect to collocation constrains, 进行 *jinxing* and 加以 *jiayi* mainly differ in syntactic type/argument structure of the complement. For example, 进行 *jinxing* can take event nouns (one place predicate) as the complement while

加以 *jiayi* cannot (e.g., 进行战争 *jinxing zhanzheng* 'to wage war'; \*加以战争 *jiayi zhanzheng* 'to wage war') (Liu, 2005; Cai, 1982). 进行 *jinxing* and 加以 *jiayi* can also be differentiated by several semantic properties of the complement, as in Table 3.2 (Zhou, 1987a; Zhou, 1987b; Cai, 1982; Xie, 1990; Tang, 2005 and among others). Moreover, these two light verbs can also be distinguished in the complexity of the complement, i.e. the complement of 进行 *jinxing* can be modified in different ways while 加以 *jiayi* prefers to co-occur with the bare object (example as shown in Table 3.3).

Semantic features of complement	进行 jinxing	加以 jiayi
Non-durative events	*进行解决	加以解决
	jinxing jiejue	jiayi jiejue
	proceed_solve	inflict_solve
	'to solve'	'to solve'
Psychological activity, e.g.,	?进行关注	加以关注
feeling/emotion/mental process	jinxing guanzhu	jiayi guanzhu
	proceed_concern	inflict_concern
	'to concern'	'to concern'
event involving interaction of the	进行交往	??加以交往
agent and patient	jinxing jiaowang	jiayi jiaowang
	proceed_associate	inflict_associate
	with	with
	'to associate with'	'to associate
		with'
accomplishment event (usually	??进行实现	加以实现
complex eventuality consisting	jinxing shixian	jiayi shixian
of a process of successive stages	procced_implement	inflict_implement
and a natural point, with a	'to implement'	'to implement'
change of state as an outcome)	-	_

Table 3.2 Semantic Features that differentiate 进行 jinxing and 加以 jiayi

	进行 jinxing	加以 jiayi	
Take VP	进行欺骗宣传	?加以欺骗宣传	
as	jinxing qipian xuanchuan	jiayi qipian xuanchuan	
attributiv	proceed_lying_ propaganda	inflict_lying_ propaganda	
e	'to conduct lying propaganda'	'to conduct lying propaganda'	
Take	进行一项研究	?加以一项研究	
NumP as	jinxing yixiang yanjiu	jiayi yixiang yanjiu	
attributiv	proceed_one_CL_research	inflict_one_CL_research	
e	'to conduct a research'	'to conduct a research'	
Take	进行感情真挚的描写 jinxing	?加以感情真挚的描写	
Subject-	ganqingzhenzhi de miaoxie	jiayi ganqingzhenzhi de miaoxie	
Predicat	proceed_affectionate_DE_descrip	inflict_affectionate_DE_descrip	
e Phrase	tion	tion	
as	'to give affectionate description'	'to give affectionate	
attributiv		description'	
e			
Take	进行反对政府的宣传	?加以反对政府的宣传	
Predicat	jinxingfanduizhengfu de	jiayifanduizhengfu de	
e-Object	xuanchuan	xuanchuan	
Phrase	proceed_anti-government_DE_	inflict_anti-government_DE_	
as	propaganda	propaganda	
	to proceed with anti-government	to conduct anti-government	
C	propaganda	propaganda	
Take PP	进行关于政治制度的改革	?加以关于政治制度的改革	
as	jinxing guanyu jingjizhidude	jiayi guanyu jingjizhidu de	
attributiv	gaige	gaige	
e	proceed_about_political	inflict_about_political	
	system_DE_reform	system_DE_reform	
	'to conduct reform of political	'to conduct reform of political	
	system'	system'	

Table 3.3 The difference in complexity of object between 进行 *jinxing* and 加以 *jiayi* 

However, in actual language use, the distinctions proposed by previous studies are often observed to be differences in tendency/preferences instead of grammatical dichotomy. For example, the corpus data shows that 加以 *jiayi* can also take a complement which denotes interactive events (e.g., 加以合作/讨论

jiayi hezuo/taolun inflict\_cooperate/discuss 'to cooperate /discuss'), even though the frequency is much lower than 进行 jinxing. Moreover, these differences may not only limit/restricted in differentiating 加以 jiayi and 进行 jinxing. They may also lie among other light verbs. For example, 搞 gao and 做 zuo also differ in the semantic property of the taken complement, in the sense that 搞 gao is more likely to co-occur with interactive event (e.g., 搞合作/辩论 gao hezuo/bianlun do\_cooperate /debate 'to cooperate /debate') while 做 zuo may not (e.g., \*做合作? /辩论 \*zuo hezuo/bianlun do\_ cooperate /debate 'to cooperate /debate'). Therefore, the features proposed by previous scholars to differentiate 进行 jinxing and 加以 jiayi may also be effective in differentiating other light verbs. But studies on the differentiation among other light verbs are very rare.

# 3.2.2.2 Previous Research on Chinese Light Verb Variation

As mentioned in Section 3.1, even with the very limited collocation constrains, light verb variations still exist among different variants of the same language. However, little work has been conducted on light verb variations among different Mandarin-speaking communities, either from theoretical perspective or from the computational perspective.

Huang et al. (2013) explore the grammatical differences of the light verb 进行 *jinxing* in Mainland and Taiwan Mandarin from an empirical perspective. They observed that 进行 *jinxing* used in Taiwan differs from that in Mainland. The Taiwan one can collocates with more types of simple nouns<sup>1</sup> and even VPs<sup>2</sup> (e.g., the examples shown below, while such collocations are not found with Mainland 进行 *jinxing*).

Simple nouns:

感恩之旅 ganen zhilüthanksgiving trip'/学术之旅 xueshu zhilü 'academic trip', 票选 piaoxuan 'ballot election', 环评 huanping 'environmental impact assessment' and 此事 cishi 'this event/activity')

VPs:

进行抹黑 jinxing mohei 'mud-sling'/测谎 cehuang 'lie-detect'/

喊话 hanhua 'propaganda'/投票 toupiao 'voting')

Therefore, Mainland Mandarin speakers may find an 'overuse' of 进行 *jinxing* by Taiwan speakers in the sense that the latter can select more types of complements (cf. Wei and Sheng, 2000; Shi et al., 2003; Diao, 2004, among others). The limitation is their study only focus on one light verb whereas more words should be included.

From the computational perspective, Lin et al. (2014) and Huang et al. (2014) use both statistical analysis and machine learning algorithm to explore the possible differences existing between Mainland and Taiwan Mandarin light verbs. The analyses show that for each light verb, there is at least one context

<sup>&</sup>lt;sup>1</sup>Here according to Huang et al., 2011, 'simple noun' (e.g., 决赛 *juesai* 'final') refers to nouns that describe a situation (cf. nouns denoting an entity such as 苹果 *pingguo* 'apple' and 桌子 *zhuozi* 'table'). But unlike deverbal noun (研究 *yanjiu* 'research'), a simple noun never has a verbal use.

<sup>&</sup>lt;sup>2</sup>VP (e.g., 验票 *yanpiao* 'inspect votes') describes a situation and in general only has verbal uses (Huang et al., 2013).

where the two variants of Mandarin show differences in usage tendencies, and thus can be differentiated. Although the paper proposes a very important and effective methodology for variation differentiation, more theoretical input is still needed.

# **3.2.3 Unsolved Problems**

As can be seen, light verb variations as well as light verb differentiation are far from comprehensively studied. The complexity and subtleness of light verb variations/alternations increase the difficulty for the differentiation. In addition, this kind of differences are more tend to be frequency or preference difference instead of grammaticality dichotomies, which is unlikely to be studied by traditional approach.

In that sense, a more quantitative approach is needed to study the gradient/probabilistic nature of grammar. Therefore, for the light verb study in this chapter, a statistical corpus-based approach based on annotated comparable corpus is adopted, following the research paradigm set up in Bresnan et al., (2007a, 2007b, 2008, 2010), Lin et al. (2014) and Huang et al. (2014). The methodology will be discussed in detail in the following section.

In this chapter, it is shown that with the assistance of statistical tool and machine learning technology, the gradient/probabilistic nature of light verb alternations especially light verb variations between different variants of Mandarin can be effectively classified and differentiated.
#### **3.2.4 Definition of Chinese Light Verb in this Study**

#### **3.2.4.1** Syntactic Type of Light Verb in Mandarin Chinese

As discussed above (in Section 3.2.1), there are numbers of debates on the syntactic type of light verbs across languages (especially in English). Mainly three approaches have been proposed by previous studies. One approach is to treat light verbs as pure functional words (e.g., copulas in Curme, 1935) that serve as a licenser of predication only (also see Grimshaw and Mester, 1998). Another approach is to identify light verb as a type of auxiliary (e.g., Hook 1974, 1991). The other approach is to consider LVCs as a complex predicate, in which both light verb and main verb contribute to the argument structure (e.g., Rosen, 1990; Mohanan, 1994; Butt 1995, 2010; Butt and Geuder, 2001). In that case, light verb is differentiated from verbal licenser or auxiliary. In other words, light verbs contribute to a semantically complex but syntactically monoclausal predication and they form a syntactically distinct class (Butt, 2010).

Light verb in Mandarin Chinese actually supports Butt (2010) that light verb should be considered as a distinct category, different from a functional element which only serves as a licenser of prediction (e.g., Curme, 1935; Grimshaw and Mester, 1998), or an auxiliary (e.g., Hook 1974, 1991). Light verbs in Mandarin are part of a joint predication within a complex predicate. The most obvious difference which sets light verbs apart from auxiliaries as well as verbal licensers is that, light verbs exhibit subtle lexical semantic differences in terms of combinatorial possibilities with complement (Butt, 2010). This is mainly reflected in two perspectives in Mandarin. On one hand, when different light verbs are co-occurring with the same complement, there are still slight lexical differences between these LVCs. For example, there is an obvious meaning difference between 从事研究 *congshi yanjiu* engage\_research 'engage in doing research' and 做研究 *zuo yanjiu* do\_research 'to conduct research', in the sense that the latter one focuses on the specific event of 'doing research', while the former one refers to the career (doing research as a profession, more genetic reading). Therefore the latter one can be co-occurred with a specific time or location marker such as 正在 *zengzai* 'progressive marker' 在实验室 *zai shiyanshi* at\_lab 'at lab' (e.g., 他正在实验室做研究 *ta zhengzai shiyanshi zuoyanjiu* 3 SG\_ZHENGZAI\_lab\_do\_research 'He is doing research at the lab'), while the former one cannot (e.g., \*他正在实验室从事研

究 ta zhengzai shiyanshi congshi yanjiu SG\_ZHENGZAI\_at\_lab\_engage\_research 'He is engaged in research at the lab'). As 从事研究 congshi yanjiu engage research 'to conduct research' refers to a more generic meaning, it has semantic conflict with the time/location marker (e.g., 正在 zhengzai 'progressive marker' or 在实验室 zai shiyanshi at lab 'at lab'). On the other hand, due to the lexical semantic differences between light verbs, collocation constrains between light verbs and the complements can be detected, as the examples given above: 进行/\*加以会议 jinxing/\*jiavi huivi proceed/inflict meeting 'have a meeting', \*进行/加以提高 \**jinxing/jiavi tigao* proceed/inflict improve 'make an improvement', \*进行/加 以 /\* 做 /\* 搞 /\* 从 事 考 虑 jinxing/\*jiayi/\*zuo/\*gao/\*congshi kaolü proceed/inflict/do/do/engage\_considerate \*proceed/inflict/\*do/\*do/\*engage considerate 'give consideration'. The two distributional characteristics set light verbs apart from either verbal licensers or auxiliaries, and make them a distinct syntactic category.

In other words, based on the two evidences shown above, this study supports Butt (2010) and considers Chinese light verbs constructions as a semantically complex but syntactically monoclausal predicate, in which both light verb and the taken complement contribute to the argument structure.

### 3.2.4.2 Definition of Light Verb in Mandarin Chinese

In terms of the definition of Chinese light verbs, this study follows Huang and Lin (2013):

'Light verbs are semantically impoverished verbs that may contribute information about event shape, but specify little about the kind of event under description. The event, i.e. the predicative content of a light verb construction, mainly comes from the event-denoting element that is taken as complement by the light verb.' (Huang and Lin, 2013, p. 728)

Based on this definition, one thing should be noted is that, during the process of data extraction, different kinds of light verb usages should be distinguished.

### **3.2.4.2.1** Differentiation of TLV and VAV usages in Chinese Light Verbs

Kearns (1998) has differentiated English light verb into two types: True Light Verb (TLV) and Vague Action Verb (VAV). These two categories mainly differ in the following properties (summary and examples are shown in Table 3.4 below). Although most of the distinctions are not found in Mandarin light verbs, there is one important feature (Feature 5) can help with the Chinese light verb categorization.

Summary	True light verb	Vague action verb
1. The complement NP	No: *A pull was given	Yes: A demonstration of
becomes surface subject in	on the rope.	the new equipment will
passive		be given on Monday.
2. The complement NP	No	Yes
undergoes WH movement	*Which pull did John	Which offer did the
	give the rope?	finance company make?
3. The complement can be	No	Yes
pronominalized	*I gave the soup a	If you can give a
	head and then Bill	presentation after lunch,
	gave it one too.	I will give one after
		yours.
4. The complement NP must	Yes	No
be indefinite	The bike looks	I cannot find the report
	terrific. *Who gave it	and I don't know who
	the polish?	made the inspection.
5. The complement is	Yes	No
headed by N which	Give a pull	
(i).is a stem form identical		
to a verb		
(ii).need not occur freely in		
other nominal environments		
(iii). Cannot license		
argument-denoting NPs		
within its projection		
6. The expression has	Yes	No
aspectual and other		
differences from the		
corresponding simple verb		
expression		

Table 3.4 Differentiation between TLV and VAV in English (Kearns, 1998)

To be particular, some Chinese light verbs are found to have both true light verb usages and non-light verb usages, in terms of the syntactic type of the taken complement. These differences are often neglected by previous researches. For example, the light verb 做 zuo can take both common nouns and de-verbal nouns as complements. Constructions like 做点心 zuo dianxin do dessert 'make dessert' and 对环境做整治 dui huanjing zuo zhengzhi for environment do renovate 'to renovate the environment' can both be found in corpus. In the construction 做整治 zuo zhengzhi do renovate 'to renovate', 整治 *zhengzhi* 'renovate' as the taken complement is a stem form identical to a verb as in 整治环境 zhengzhi huanjing renovate environment 'to renovate the environment' (corresponds with Feature 5). Moreover, 整治 zhengzhi 'renovate' does not occur freely in other nominal environments (it only occurs as a deverbal noun in Light Verb Constructions). That is to say, the different usages within light verbs like 做 zuo actually in accordance with the distinction between TLV and VAV proposed by Kearns (1998).

Moreover, native speakers of Mandarin Chinese may have the intuition that 做 zuo in these two constructions (做点心 zuo dianxin do\_dessert 'make dessert' and 做整治 zuo zhengzhi do\_renovate 'to renovate') differ in the amount of content it contributes to the whole construction. 整治 zhengzhi 'to renovate' as a de-verbal noun contains the eventive information, which in accordance with the definition of light verbs. 做 zuo in this case can be omitted without changing proposition of the construction (e.g., 对环境 (做) 整治 dui huanjing (zuo) zhengzhi for\_ environment\_(do)\_renovate 'to renovate environment'), although sometimes the whole construction/sentence needs to be re-written to ensure the grammaticality. In contrast, 点心 dianxin 'dessert' is a common NP which refers to a concrete entity and does not contain any predicative information. Therefore the predicative information all comes from the verb 做 zuo. In this case, 做 zuo in 做点心 zuo dianxin 'make dessert' may represent a series of actions including 'stir' 'blend' 'knead' 'bake' and so on. Hence it cannot be omitted in the construction. From this perspective, the 做 zuo in 做点心 zuo dianxin do\_dessert 'make dessert' is not in accordance with the definition of light verbs.

The same distinction between TLV and VAV can also be detected within the usages of some other light verbs (e.g., 搞/弄/干 *gao/nong/gan* 'do'). That is to say, light verbs like 做 *zuo*, 搞 *gao*, 弄 *nong*, 干 *gan* can both be used as TLV and VAV. However, according to my research objectives, the VAV usages should not be considered as light verb usages. Therefore the VAV usages of some light verbs should be excluded in data collection.

### **3.2.4.2.2** Criteria of Data Collection

Based on the definition of light verbs as well as the distinction between TLV and VAV, a more detailed criterion has been established. VAV usages should be excluded in the process of data extraction. The criteria are shown below:

i. The taken complement should contain eventive information, it is either a

de-verbal noun or a verb. Event noun is also included in the sense that it contributes predicative information, and the frequencies of taking event noun are relatively small for both Mainland and Taiwan Mandarin (shown in Section 3.5.1).

- ii. As claimed in Huang and Ahrens (1995), the classifier correlates with the semantic properties of a noun, and event classifiers select event-type nominals. Classifier can coerce nouns to refer to kinds, individuals and events. Therefore the classifier can be a very good indicator to examine the semantic property of the taken complements. As in Mandarin Chinese, there is no morpho-lexical marking involved when the eventtype nominal is deverbaled, classifier is a good test to differentiate the noun which contains eventive information (deverbal noun or event noun) and the common noun. Usually, only the verbal classifier can co-occur with action/event (the list of verbal classifier is shown in Table 3.5).
- iii. The light verb can often be omitted without changing the proposition of the whole sentence (although the sentence often needs to be re-written to ensure the grammaticality).

波 bo	Of staggered event
班 ban	Of shift, scheduled flight/bus etc.
笔 bi	Of transaction
步 bu	Step (event procedures)
泡 pao	A brewing (of tea etc.)
盘 pan	A serving round (of a dish)
幕 mu	Act (of a play)
番 fan	Times (of a repeated event)
道 dao	Of dishes of procedures
档 dang	Duration of a run (of play, movie etc.)

段 du	Section (of play, etc.)	
顿 dun	The process of a meal	
台 tai	A run of a traveling troupe	
堂 tang	A class	
趟 tang	A journey	
通 tong	A phone call	
轮 lun	A round	
回 hui	A roundtrip	
节 jie	A class, a session	
届 jie	An annual event	
件 jian	Event	
局 ju	Game	
期 qi	Term	
起 qi	Event (especially a happening, an	
	accident)	
圈 quan	Round (of majong)	
席 xi	Lecture	
折 zhe	An act (in a Chinese play)	
阵 zhen	one of a sporadic events	
桩 zhuang	Event	
场 chang	A (scheduled) event (with beginning	
	and ending)	
出 chu	A play	
任 ren	Term (of a termed position)	
宗 zong	Trade/transaction	
餐 can	A meal	
次 ci	Once (re. frequency of event)	

Table 3.5 The list of verbal classifier in Mandarin Chinese (cite from Huang<br/>and Ahrens (1995))

# **3.2.4.2.3** The Range of Light Verbs

Then based on the definition and criteria, another issue needed to be identified is, which verb can be included into the category of light verbs in Mandarin

Chinese?

The range of light verbs (i.e. which verbs should be considered as light verb)

varies in different previous studies. While some studies include verbs such as 给以 *jiyu* 'render', 予以 *yuyi* 'give', 展开 *zhankai* 'carry out', 开展 *kaizhan* 'carry out' and even some dialectical verbs like 干 *gan* 'do' and 弄 *nong* 'do' (e.g., Diao, 2004), most studies only focus on the most typical and frequently used light verbs (e.g., 从事 *congshi* 'engage', 做 *zuo* 'do', 搞 *gao* 'do'), especially on the most frequently pair 进行 *jinxing* 'proceed' and 加以 *jiayi* 'inflict'. In my study, to avoid the ambiguity, only seven most frequently used light verbs are included (进行/加以/做/搞/从事/展开/开展 *jinxing/jiayi/zuo/gao/congshi/zhankai/kaizhan* 'proceed/inflict/do/do/engage/ carry out/carry out). VAV (Verbal Action Verb) usages of light verbs were excluded in the process of data extraction, as discussed in Section 3.2.4.2.1.

# **3.3 Methodology**

### **3.3.1 Data Extraction**

The corpus used in this study is the annotated Gigaword corpus, (Huang, 2009) which was collected and available from LDC and contains over 1.1 billion Chinese words, with 700 million characters from Taiwan Central News Agency and 400 million characters from Mainland Xinhua News Agency. For the seven light verbs that I focus on in this study (进行/加以/做/搞/从事/展开/开展 *jinxing/jiayi/zuo/gao/congshi/zhankai/kaizhan*), about 200 sentences were randomly extracted from the corpus for each light verb in each variety, which result in about 1400 sentences in each variety, and more than 2800 tokens in

total. The selection follows the principle that it could cover the different uses of each light verb.

#### **3.3.2 Data Annotation**

As has been reviewed in Section 3.2, previous studies (Zhu, 1985; Zhou, 1987b; Cai, 1982; Huang et al., 2013, among others) have proposed several features to identify the commonalities and differences between different light verb pairs, especially between the two most typical ones, i.e. 进行 *jinxing* and 加以 *jiayi*. The features covers semantic, syntactic, as well as discourse levels. For examples, in terms of the semantic type of the taken complements, 加以 *jiayi* usually takes the psychological event as the complement (e.g., 加以考虑 jiayi kaolv inflict considerate 'give consideration'), while 进行 jinxing disfavors this kind of complement (?进行考虑 jinxing kaol ü proceed considerate 'give consideration'). At syntactic level, 进行 jinxing has been observed to be compatible with aspectual marker (e.g., 进行了研究 jinxing le vanjiu proceed LE research 'has done research'), while 加以 jiayi has strong dispreference with 着 *zhe* 'progressive marker', 了 *le* 'perfective marker' and 过 guo 'experiential marker'. Furthermore, it has been observed that these linguistic features may also be effective in differentiating light verb variations. For instance, 从事 *congshi* in Taiwan, but not in Mainland Mandarin, can take informal event complements like 性交易 xingjiaoyi 'sexual trade' (Huang et al.,

2014).

Therefore, in this study for light verbs, 12 features which may help to differentiate different light verbs variations were selected, based on previous studies as well as my own observations. All the 12 selected features and examples are shown in Table 3.6. To simplify the annotation and statistical analyses, most of the factors are binary (with yes or no alternations). Over 2800 sentences in both varieties have been manually annotated by a main annotator who is a trained expert on Chinese linguistics. Any ambiguous cases were discussed with another two experts in order to reach an agreement. Two more annotators were also involved to test the validation and reliability of the annotation schema, which is shown in Section 2.2.5 of Chapter 2.

Feature	Example	Feature	Example
Co-occur with	开始进行比赛	Spontaneous/contr	进行投票
other light	kaishi/jinxing/bisai	ollable event	jinxing/toupiao
verbs 'OTHE	start_proceed_game	'SPONTEVT'	proceed_vote
RLV'	'start the game'		'carry on voting'
Take	昨天进行了比赛	Durative event	进行比赛
aspectual	zuotian/jinxing/le/bisa	'DUREVT'	jinxing/bisai
marker:着,	i		proceed_game
了,过	yesterday_procced_L		ʻplay a game'
'ASP'	E_game		
	'played the game		
	yesterday'		
Event	比赛在学校进行	Formal event	进行访问
complement	bisai/zai/xuexiao/jinxi	'FOREVT'	jinxing/fangwen
is at subject	ng		proceed_official
position	game_at_school_proc		visit
'EVECOMP'	eed		'pay an official
	'play the game at		visit'
	school		
POS	进行比赛(N)	Psychological	加以考虑
'POS'	jinxing/bisai	activity	jiayi/kaol ü
	proceed_game	PSYEVT	inflict_considera
	'play the game'		te

			'give consideration'
Argument structure 'ARGSTR'	进行调查(two) <i>jinxing/diaocha</i> proceed_investigate 'carry on investigation'	Involving interaction of agent and patient 'INTEREVT'	进行沟通 <i>jiayi/goutong</i> proceed_commu nicate 'to communicate'
VO compound as argument 'VOCOMP'	进行投票 <i>jinxing/toupiao</i> proceed_vote 'carry on voting'	Accomplishment complement 'ACCOMPEVT'	进行修正 <i>jinjing/xiuzheng</i> proceed_correct 'make corrections'

Table 3.6 The annotation standard of light verb construction

# **3.4 Data Analysis and Statistical Tool**

Based on the annotated data, both Chi-square test and mixed-effect logistic regression model are used, to differentiate the light verb variations between different variants as well as to differentiate between light verbs.

Mixed-effect logistic regression model is conducted, following the paradigm of Bresnan (2007a, 2007b, 2008, 2010). The result and discussion are shown in Section 3.4.1. Moreover, light verbs are also divided into three near-synonym pairs (进行/加以 *jinxing/jiayi* 'proceed/inflict'; 做/搞 *zuo/gao* 'do' and 展开/开展 *zhankai/kaizhan* 'carry out'). Chi-square tests are used to differentiate between light verb pairs as well as between language varieties. The results and discussion of Chi-square tests are shown in Section 3.4.2.

#### **3.4.1 Mixed-effect Logistic Regression Model**

For the statistical analysis, the variation differences between Taiwan and

Mainland were analyzed by using mixed-effect logistic regression model (GLMM) (Pinheiro and Bates, 2000) implemented in the R package lme4 (Bates and Maechler, 2009).

Since the dataset includes different verbs, the variability in the verbs should not be disregarded. Therefore the verb is treated as a random effect in this model because the study are not meant to be restricted to the verbs collected, but to generalize across the entire population of possible verbs (Bresnan et al., 2010). For this reason, verb is a random variable and is treated as a random effect in the model. Using a mixed-effect regression approach ensures our results are generalizable and enabled us to quantify and study the variation of individual words.

For the model in the current study for light verbs, the 'item' refers to different verbs, which is considered as the random-effect factor. The dependent variable is source (the source of the data: from Taiwan or Mainland), and the independent variables are the features we annotated (as shown in the annotation schema in Table 3.6): polarity of the context (polarity), POS of the complement (POS), whether the complement is VO compound or not (VOcomp), whether the complement is at the subject position (subject), whether the light verb occurs with other light verb (otherlv), whether the light verb takes aspectual marker (AspectualMarker), whether the complement denotes durative event (durative), formal event (formal), psychological event (psychological), interactive event (interactive), whether the complement is an accomplishment event (accomplishment), whether the complement has attributive or not (attributive). It should be noted that one of the features 'spontaneous' only has one level (all light verbs take spontaneous events as their complements), therefore this feature is not effective in differentiating the variation differences. It is excluded for further statistical analysis. To have sufficient data for simultaneous comparisons, most variables were simplified into binary values: yes and no (respectively denoted by 1 and 0). For example, 'AspectualMarker' was simplified to 'has' and 'not has', 'subject' was simplified to 'at subject position' or 'not at subject position'.

The specification of this model is shown below:

Formula: source ~ polarity + POS + VOcomp + subject + other lv+ AspectualMarker + durative + formal + psychological + interactive + accomplishment + attributive + (1 | item)

The result of model is shown in Table 3.7. Among all the 12 features which were included in the model, 7 of them show significance: 5 features are significant at P<0.01 level while 2 features are significant at P<0.05 level (POS of the complement; Attributive of the complement). The estimates in the table represent the regression coefficients. The signs of the coefficients show the direction of the effects. To be particular, positive coefficients favor the 1 source (here refers to Taiwan) and negative coefficients favor the 0 source (here refers to Mainland).

	Estimate	Std. Error	z value	Pr(> z )		
(Intercept)	0.23878	0.54148	0.441	0.659233		
Polarityneutral	-1.77436	0.18638	-9.520	<2e-16 ***		
Polaritypositive	-2.86264	0.23273	-12.300	<2e-16 ***		
POSV	-0.24495	0.09671	-2.533	0.011316 *		
POSVP	1.60120	0.60587	2.643	0.008222 **		
VOcomp1	0.78753	0.28441	2.769	0.005623 **		
subject1	-0.65310	0.14668	-4.453	8.48e-06 ***		
otherlv1	1.41503	1.17371	1.206	0.227973		
AspectualMarker1	-1.90630	0.19888	-9.585	<2e-16 ***		
durative1	1.57893	0.47182	3.346	0.000819 ***		
formal1	0.37082	0.21092	1.758	0.078726.		
psychological1	-0.40963	0.40917	-1.001	0.316762		
interactive1	-0.04192	0.12864	-0.326	0.744541		
accomplishment1	-0.43867	0.23916	-1.834	0.066627 .		
attributive1	-0.21312	0.10318	-2.066	0.038868 *		
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1						

Table 3.7 Result of logistic regression model for light verbs

In the following section, each feature will be examined one by one to discuss the variation difference between Mainland and Taiwan light verbs.

The main effect of 'polarity' is negatively significant. To be particular, the feature 'polarityneutral' has a negative coefficient -1.77436, indicating the neutral context favors the Mainland light verb (e.g., 如何进行战斗 *ruhe jinxing zhandou* how to\_proceed\_fight 'how to fight'; 加以改组 *jiayi gaizu* inflict\_reorganize 'to reorganize'; 做按摩 *zuo anmo* do\_massage 'to massage'; 搞教育 *gao jiaoyu* do\_educate 'engage in education'; 从事文学创作 *congshi wenxue chuangzuo* engage\_literary\_creation 'engage in literary creation'; 开展文体活动 *kaizhan wenti huodong* carry out\_cultural and sport\_activity 'carry out cultural and sport activity'; 展开了世界冠军争夺赛 *zhankai le shijie guanjun zhengduosai* carry out\_world\_champion\_competition 'have contested

the world championship'). Moreover, the Mainland light verbs are also favored by the positive context (the feature 'polaritypositive' also has a negative coefficient -2.86264). For example, 做了有益的准备 zuo le youvi de zhunbei do LE beneficial DE preparation 'have made helpful preparation'; 工作搞 得 好 搞 得 活 de hao gao de huo gongzuo gao work do DE good do DE lively 'job done well and lively'; 开展弘扬中华 民族文化的工作 kaizhan hongyang zhonghua minzu wenhua de gongzuo carry out carry forward Chinese culture DE work 'to carry forward the Chinese national culture' are frequently appeared in Mainland corpus. The result also indicates that the negative context favors Taiwan light verbs (e.g., 从事违法竞 选活动 congshi weifa jingxuan huodong engage illegal election activity 'engage in illegal election campaign'; 搞政治权谋 gao zhengzhi quanmou do political tactic 'utilize political tactics'). The feature 'POS' is also significant: the feature 'POSV' shows a negative coefficient -0.24495, indicating that the verb complement favors the Mainland light verbs (e.g., 进行 研究/加以讨论 jinxing yanjiu/jiayi taolun proceed research/inflict discuss 'to do research/to discuss'), while the feature 'POSVP' has a significant positive coefficient +1.60120, implying Taiwan light verbs is favored by the VP 进行修正法案 complement (e.g., jinxing xiuzheng fa'an proceed amendment bill 'to amendment bill'; 加以调整货币政策 jiayi tiaozheng huobi zhengce inflict adjust monetary policy 'to adjust monetary policy'). The result also indicates that nominal complement (event noun) favors

84

the Taiwan light verbs (e.g., 进行议程/感恩之旅 *jinxing yicheng/gan'en zhilü* proceed\_agenda/thanksgiving\_trip 'proceed with the agenda/have thanksgiving trip'). Moreover, the main effect of 'VOcomp1' is also positively significant, meaning that the VO complement is more likely to be occurred with Taiwan light verbs (e.g., 进行投票/开票 *jinxing toupiao/kaipiao* proceed\_vote/ballot\_counting 'to cast voting/ballot-counting'; 从事洗钱 *congshi xiqian* engage\_money\_laundering 'engage in money laundering'; 开展 '批苏' *kaizhan pisu* proceed\_animadvert\_the Soviet Union 'to animadvert the Soviet Union').

In terms of the syntactic position of light verb, the feature 'subject1' has shown a significant negative coefficient -0.65310, indicating that light verbs in Mainland are more likely to be occurred in the subject position (e.g., 战争仍在 进行 zhanzheng reng zai jinxing war\_still\_at\_proceed 'war is under way'; 把 工作做得好 ba gongzuo zuode hao BA\_work\_do\_DE\_good 'to a good job'; 把 经 济 建 设 搞 得 很 好 bajingji jianshe gaode henhao BA\_economic\_development\_do\_DE\_good 'did very well in economic development'; 活动在全国广泛开展 huodong zai quanguo guangfan kaizhan activity\_at\_nation\_widely\_carry out 'activities have been carried out nationwide'; 青工技能活动开始展开 qinggong ji'neng huodong kaishi zhankai young worker\_skill\_activity\_start\_carry out 'yong worker training activities start to be carried out').

With respect to aspectual marker (zhe, le and guo), the negative coefficient

-1.90630 indicates that Mainland light verbs are more likely to occur with aspectual marker (e.g., 进行了激烈的斗争 *jinxing le jilie de douzheng* proceed\_LE\_fierce\_DE\_fight 'had a fierce fight'; 做了大量工作 *zuo le daliang de gongzuo*do\_LE\_plenty of work 'have done plenty of work'; 搞了 几次对抗 *gaole jici duikang* do\_LE\_several\_CL\_confrontation 'had several confrontations'; 搞过房地产开发 *gaoguo fangdichan kaifa* do\_GUO\_real-estate\_develop 'have involved in development of real-estate'; 进行着探索 *jinxing zhe tansuo* proceed\_ZHE\_explorate 'having exploration'; 进行过维护 *jinxing guo weihu* proceed\_GUO\_maintain 'had maintained').

In the case of semantic properties of the complement, most features do not show significance in our model. Actually only the feature 'durative' shows a significant positive coefficient +1.57893, indicating it favors Taiwan light verbs (e.g., 加以长时间观察 *jiayi changshijian guancha* inflict\_long\_time\_observe 'to have long time observation'; 进行长期接触 *jinxing changqi jiechu* proceed\_long-term\_contact 'to have long-term contact'; 做健康检查 *zuo jiankang jiancha* do\_health\_check 'to have health check'). Other semantic features do not show significance (including formal, psychological, interactive and accomplishment event).

In terms of the complexity of the complement, the feature 'attributive1' has a significant negative coefficient -0.21312, indicating that the complex complement is favored by Mainland light verbs. For example, constructions like 进行了丰富多彩、别开生面的歌舞表演 *jinxing le fengfu duocai* 

biakaishengmian de biaovan gewu proceed LE various lively DE singing dancing performance 'carried out a variety of lively and wonderful dancing and singing performances'; 做了高难 度 的 搭 桥手术 zuole gaonandu de daaiao shoushu do LE difficult bypass operation 'had a difficult bypass operation'; 搞多头 重复检查 gao duotou chongfu jiancha do multiple repetitive check 'to have multiple and repetitive check' are frequently found in Mainland Corpus.

For the five features which do not show significance in our model, four of them refer to the semantic properties of the complement: whether the complement denotes formal event, psychological event, interactive event or accomplishment event. In other words, Taiwan and Mainland light verbs do not tend to have significant difference in the semantic properties of their taken complements. Another non-significant feature is 'otherlv', and if we examine our annotated data in detail, we could find that the number of co-occurrence of this feature is too small. To be particular, among more than 2800 annotated data, only 7 tokens of light verbs co-occurred with other light verbs (e.g., 活动开始 展开 *huodong kaishi zhankai* activity\_start\_carry out 'activities start to be carried out'; 着手加以研究 *zhuoshou jiayi yanjiu gaishan* start\_inflict\_research 'start to conduct research'; 继续进行议程 *jixu jinxing zhengjianhui* continue\_proceed\_agenda 'continue to proceed with the agenda'). Therefore the non-significance is not difficult to be accounted for.

In a word, the result can be summarized as shown below:

i. Features that favor Mainland light verbs:

Positive context; neutral context; Verb complement; complement at subject position; take aspectual marker; complement with attributive

ii. Features that favor Taiwan light verbs:

Negative context; VP complement; N complement; VO compound; durative event;

#### 3.4.2 Chi-square Test

In the previous section, the results of mixed-effect logistic regression model present us a general tendency of light verb variations between Mainland and Taiwan Mandarin. However, the subtle variation differences for each individual light verb/light verb pair also need to be examined. In Huang et al., (2013), as shown in Section 3.2.2.2, Taiwan 进行 *jinxing* are found to be more likely to take simple nouns (e.g., 进行议程/日程 *jinxing yicheng/richeng* proceed\_agenda/schedule 'to proceed with the agenda/schedule') and even VPs (e.g., 进行投票/开票 *jinxing toupiao/kaipiao* proceed\_vote/ballot counting 'to cast voting/ballot counting'), but other light verb variations have not been discussed in previous studies. To investigate the variation differences for each light verb, Chi-square test is used here to test for the significance of the co-occurrence of each factor with each individual light verb. This approach enables us to quantify and study the variation of individual light verbs.

Moreover, as mentioned above, the light verbs collected in this study can be divided into three near-synonym pairs: 进行/加以 *jinxing/jiayi*; 做/搞

*zuo/gao* and 开展/展开 *kaizhan/zhankai*. The reason of categorizing like this is because these three pairs are often considered as synonyms and are mutual glossing in the dictionary. For example:

- i. 搞 gao: 做/干/弄 zuo/gan/nong'do';《现代汉语八百词 Xiandai
  Hanyu Babai Ci 'Modern Chinese Eight Hundred Words', 1980, p. 219》
- ii. 开展 kaizhan: 展开 zhankai'carry out'.《现代汉语词典 Xiandai Hanyu Cidian 'Contermporary Chinese Dictionary', 2005, p. 722》

Because of the versatilities of their semantic meaning, the near-synonyms pairs are interchangeable under most contexts: 进行/加以讨论 *jinxing/jiayi taolun*proceed/inflict\_discuss 'to discuss'; 做/搞了很长时间的研究 *zuo/gao le henchang shijian de yanjiu* do/do\_LE\_very\_long\_time\_DE\_research 'have conducted research for a very long time'; 展开/开展学习活动 *zhankai/kaizhan xuexi huodong* carry\_out\_learning\_activity 'to carry out learning activity'. However, collocation constrains within the pairs can also be found: \*做/搞第三 产业 *zuo/gao disan chanye* do/do\_tertiary\_industry 'engage in tertiary industry'; ?进行/加以考虑 *jinxing/jiayi kaol ü* proceed/inflict considerate 'give consideration'. Very few studies were conducted to differentiate between these light verb synonym pairs (especially between 展开/开展 *zhankai/kaizhan*). Therefore, to differentiate between these light verb pairs is also a very important task for linguistic researches.

In this sense, dividing the light verbs into near-synonym pairs would allow us to make the distinction from two dimensions: from both variety internal (i.e. light verb alternations) and variety external (i.e. light verb variations) perspective. In terms of variety internal comparison, the subtle difference within the same pair (e.g., between 进行 *jinxing* 'proceed' vs. 加以 *jiayi* 'inflict' in Mainland) needed to be investigated. Meanwhile, the factors that influence the alternative choices between similar light verbs will also be discovered; with respect to variety external perspective, the variation differences of the same light verb in different varieties (e.g., 进行 *jinxing* in Taiwan vs. 进行 *jinxing* in Mainland) will be examined. In other words, by dividing the light verbs into synonym pairs, both light verb alternations (alternative choices between similar light verbs) and light verb variations can be explored.

The feature set adopted in this section is the same as that has been used in Mixed-effect logistic regression model (shown in Table 3.6). The results of Chisquare test show that statistical model and annotated features are effective in differentiating both light verb alternations and light verb variations.

For the Chi-square test, the tool I used is SPSS V.22. All the results are transformed (Standardized pearson residuals  $e_{ij}$  or < 1.9 (Agresti and Kateri, 2011)) into signs (see Table 3.8). The results of test are presented in the following sections, from both variety internal and variety external perspectives.

Sign	Standardized	Interpretation
	pearson residuals	
+	$e_{ij} > 2$	statistically
		significant
		overuse of the
		light verb with
		the factor
-	$e_{ij} < -2$	statistically
		significant
		underuse of the
		light verb with
		the factor
0	$e_{ij}[-2,2]$	lack of statistical
	- -	significance

Table 3.8 Interpretation of standardized person residuals

### 3.4.2.1 Light Verb Alternations: Variety Internal Comparison

For the variety internal comparison within each light verb pairs, the results show that all three pairs in both varieties can be differentiated by most of the annotated features (i.e.those where they have contrasting positive/negative tendencies to appear (P-value<0.05)).

### 3.4.2.1.1 Differentiation between 进行 jinxing and 加以 jiayi

Chi-square test allows us to make the differentiation of 进行 *jinxing* and 加以 *jiayi* from a preference/tendency perspective. The Chi-square test results of Mainland data are first presented, as shown in Table 3.9 (some features which do not show significant correlation would be omitted in the table).

Features		进行	加以
Complement	Event	+	-
Types	noun		
	De-verbal	-	+
	noun		
Event complement	at subject	+	-
position			
Take aspectual mar	ker (le)	+	-
Take marker (guo)		0	0
Durative event		+	-
Formal event	0	0	
Psychological even	-	+	
Event involving in	+	-	
the agent and patien	nt		
Accomplishment ev	vent	-	+
Attributive of the co	+	-	
Connotation of Positive		-	+
the Complement	the Complement Neutral		-
	Negative	-	+

Table 3.9 Comparison between Mainland 进行 jinxingand 加以 jiayi

According to Table 3.9, 进行 *jinxing* and 加以 *jiayi* in Mainland can be differentiated by most of the factors. For example, Mainland 进行 *jinxing* 'proceed' has a preference of taking Event Noun (e.g., 进行会议/赛事 *jinxing huiyi/saishi* proceed\_meeting/game 'have a conference/play a game') as the complement, while 加以 *jiayi* prefers to take verbal complements (e.g., 加以提高/修改 *jiayi tigao/xiugaiinflict\_improve/revise* 'to improve/revise'). 进行 *jinxing* differs from 加以 *jiayi* 'inflict' in that it also has preference in taking aspectual marker le (进行了激烈的战斗 *jinxing le jilie de zhandou* proceed\_LE\_fierce\_DE\_fight 'have a fierce fight'), taking durative events (进行会议 *jinxing huiyi* proceed\_meeting 'have a conference'), interactive events (e.g., 进行磋商 *jinxing cuoshang* proceed negotiate 'have a negotiation') and

complements which have attributives (进行长时间的讨论 *jinxing* changshijiande taolunproceed long time DE discuss 'have a long-time discussion'), whereas 加以 jiayi shows dis-preference over these features. In contrast, 加以 *jiayi* prefers to take psychological event as the complement (加 以考虑 jiayi kaol ü inflict considerate 'give consideration') while 进行 jinxing has the opposite preference (?进行考虑 jinxing kaolü proceed\_considerate 'give consideration'). 加以 jiayi has a strong tendency of taking accomplishment event (e.g., 加以澄清/提高 jiayi chengqing/tigao inflict clarify/improve 'to clarify/improve'), while 进行 jinxing seems to disfavor this feature (e.g., ?进行澄清/提高 jinxing chengqing/tigao proceed clarify/improve 'to clarify/improve'). In terms of polarity of the context, 进行 *jinxing* is more likely to be occurred in neutral context (e.g., 进行 技术改造 jinxing jishu gaizao proceed\_technical\_improve 'to have technical improvement'), while 加以 *jiayi* prefers the positive (e.g., 加以表扬 *jiayi* biaoyang inflict praise 'to praise') or negative context (e.g., 加以指责 jiayi zhize inflict blame 'to blame').

进行 *jinxing* and 加以 *jiayi* in Taiwan can also be effectively differentiated, according to the result of Chi-square test, as presented in Table 3.10.

Feature	Features		TW 进行	TW加以
POS	Event not	In	+	-
	Deverbal	noun	-	+
VO			+	-
Event c	omplement	at	+	-
subject	position			
Take as	pectual ma	rker (le)	0	0
Take as	pectual ma	rker (guo)	0	0
Durativ	e event		0	0
Formal	event		0	0
Psychol	logical even	nt	-	+
Event in	nvolving in	teraction	+	-
of the a	gent and pa	ntient		
Accom	plishment e	event	-	+
Attribut	tive of the		+	-
comple	ment			
Polarity	7	Positive	-	+
		Neutral	+	-
		Negative	0	0
Occur v	Occur with other lv		0	0
Spontaneous event		0	0	

Table 3.10 Comparison between TW 进行 jinxing and 加以 jiayi

Similar to Mainland 进行 *jinxing* and 加以 *jiayi* pair, 进行 *jinxing* and 加 以 *jiayi* in Taiwan also have the contrast that 进行 *jinxing* is more likely to take nominal complement (e.g., 进行议程 *jinxing yicheng* proceed\_agenda 'proceed with the agenda'), while 加以 *jiayi* favors the verbal complement (e.g., 加以辅 导 *jiayi fudao* inflict coach 'to coach'). Moreover, Taiwan 进行 *jinxing* has shown the preference of taking VO compound as the complement (e.g., 进行投 票 / 开 票 *jinxing toupiao/kaipiao* proceed\_vote/ballot counting 'to cast voting/ballot counting'), while 加以 *jiayi* disfavors this feature. The complements of Taiwan 进行 *jinxing* tend to appear in subject position (e.g., 演 习在日本进行 *yanxi zai riben jinxing* maneuver at Japan proceed 'maneuver

is carried out in Japan'), while 加以 jiayi does not allow this distribution (\*批 评加以 piping jiavi criticize inflict 'to criticize'). Taiwan 进行 jinxing and 加 以 *jiayi* have not shown statistical differences in the preference of aspectual marker, durative event and formal event. But Taiwan 加以 jiayi has shown its favor for psychological event (e.g., 加以考虑 jiavi kaol ü inflict considerate 'to considerate') and accomplishment event (e.g., 加以改进 jiayi gaijin inflict improve 'to improve'), while 进行 *jinxing* is more likely to co-occur with interactive event (e.g., 进行合作 jingxing hezuo proceed\_cooperate 'to cooperate') and complex complement (e.g., 进行长期的技术研究 *jinxing* changqi de jishu yanjiu proceed long-term DE technical research 'to conduct long-term technical research'). In terms of polarity, in Taiwan, 加以 jiayi prefers the positive context (e.g., 加以称赞 jiavi chengzai inflict praise 'to praise'), while 进行 *jinxing* is more likely to be appeared in the neutral context (e.g., 进行合作 *jinxing hezuo* proceed corporate 'to cooperate').

# 3.4.2.1.2 Difference between 做 zuo and 搞 gao

The alternative choice between 做 *zuo* and 搞 *gao* is very lack of studies. The Chi-square tests in both varieties show that the light verb 做 *zuo* and 搞 *gao* can be differentiated effectively. As presented in Table 3.11 below:

Features		ML 做	ML 搞	
POS	Ν		+	-
	Deverbal	noun	-	+
VO con	nplement		0	0
Event c	omplement	t at	0	0
subject	position			
Take as	pectual ma	rker (le)	+	-
Take as	pectual ma	rker (guo)	0	0
Durativ	e event		0	0
Formal	event		0	0
Psychol	logical even	nt	0	0
Event in	nvolving in	teraction	0	0
of the agent and patient				
Accom	plishment e	event	0	0
Attribut	tive of the		+	-
complement				
Polarity	7	Positive	-	+
		Neutral	+	-
		Negative	-	+

Table 3.11 Comparison between Mainland 做 zuo and 搞 gao

做 zuo 'do' and 搞 gao 'do' in Mainland are also differed in POS feature in the sense that 做 zuo has the preference in taking noun as the complement (e.g., 做工作/手术/活动 zuo gongzuo/shoushu/huodong do\_work/operation/activity 'do job/operation/activity'), while 搞 gao tends to take verbal complement (e.g., 搞批发/改革 gao pifa/gaige do\_wholescale/reform 'do wholescale trade/to reform'). Moreover, 做 zuo also significantly prefers the complements which have attributive (e.g., 做了很长时间的工作 zuo le henchang shijian de gongzuo do\_LE\_very\_long\_time\_DE\_work 'have been working for a very long time') while 搞 gao shows the opposite preference.

Features		TW 做	TW 搞	
POS	Ν		-	+
	Deverbal	noun	+	-
VO con	nplement		0	0
Event c	omplement	at	0	0
subject	position			
Take as	pectual ma	rker (le)	+	-
Take as	pectual ma	rker (guo)	0	0
Durativ	e event		0	0
Formal	event		0	0
Psychol	logical even	nt	+	-
Event in	nvolving in	teraction	+	-
of the agent and patient		atient		
Accom	plishment e	event	+	-
Attribut	tive of the		+	-
complement				
Polarity	7	Positive	+	-
		Neutral	+	-
		Negative	-	+

Table 3.12 Comparison between TW 做 zuo and 搞 gao

As presented in Table 3.12, Taiwan 做 *zuo* and 搞 *gao* have the opposite contrast in the preference of POS of the complement, compared to the Mainland ones, in the sense that 搞 *gao* in Taiwan is more likely to take the nominal complement (e.g., 搞 霸 权 主义 *gao baquanzhuyi* do\_hegemonism 'seek hegemony') while 做 *zuo* shows its favor of verbal complement (e.g., 做宣传 *zuo xuanchuan* do\_propaganda 'to conduct propaganda'). Beside this, Taiwan 做 *zuo* is also significantly favored by aspectual marker (e.g., 做了以上表示 *zuo le yishang biaoshi* do\_LE\_above\_statement 'to issue above statement'), psychological event (e.g., 做考虑 *zuo kaolü* do\_considerate 'give consideration'), interactive event (e.g., 做交流 *zuo jiaoliu* do\_communicate 'to communicate'), accomplishment event (e.g., 做修正 *zuo xiuzheng* do correct

'to correct') and complex complement (e.g., 做具体的宣传工作 *zuo juti de xuanchuan gongzuo* do\_specific\_DE\_propaganda\_work 'to conduct specific propaganda'). In terms of polarity of the context, Taiwan 做 *zuo* shows its favor of positive (e.g., 做正确的选择 *zuo zhengquan de xuanze* do\_right\_DE\_choice 'make the right choice') and neutral context (e.g., 做评论 *zuo pinglun* do\_comment 'give comment'), while 搞 *gao* prefers negative context (e.g., 搞 分裂 *gao fenlie* do dispute 'to dispute').

# 3.4.2.1.3 Difference between 展开 zhankai and 开展 kaizhan

The difference between 展开 *zhankai* and 开展 *kaizhan* is very subtle to be observed. It seems very difficult to find the factors which can influence the alternative choices between these two light verbs (e.g., 开展/展开交流活动 *kaizhan/zhankai jiaoliu huodong* carry out\_interchange\_activity 'carry out interchange activity'; 开展/展开讨论 *kaizhan/zhankai taolun* carry out\_discuss 'to discuss'). So far, to the best of my knowledge, no research has been found in differentiating this synonym pair. In this study, the Chi-square test may help us to discover the subtle differences. The result of statistical analysis for  $\mathbb{R}$ 开/开展 *zhankai/kaizhan* is shown below in Table 3.13, Mainland data is first presented.

Features		ML展开	ML开展
POS	N	-	+
	Deverbal	+	-
	noun		
	VP	-	0
VO		0	0
Event complement at su	bject position	+	-
Take aspectual marker (	(le)	+	+
Formal event		0	0
Event involving interaction of the		+	-
agent and patient			
Accomplishment event		0	0

Table 3.13 Comparison between Mainland 展开 zhankai and 开展 kaizhan

As shown in Table 3.13, compared to Mainland 开展 kaizhan, Mainland 展开 zhankai tends to take verbal complement (e.g., 展开讨论/教育/宣传 zhankai taolun/jiaoyu/xuanchuan carry out discuss/educate/propaganda 'carry out discussion/education/propaganda'). 开展 kaizhan prefers the nominal object: e.g., 开展文体活动/社教工作 kaizhan wenti huodong/shejiao gongzuo carry out\_cultural\_and\_sport\_activity/social\_education\_work 'carry out cultural and sport activity/social education'. The feature 'event complement at subject position' is favored by 展开 zhankai (e.g., 建设工程全面展开 jiashe gongcheng quanmian zhankai construction engineering overall carry out 'construction engineering is fully carried out') while it is dis-favored by 开展 *kaizhan*. Also, 展开 *zhankai* prefers the aspectual marker (perfective marker) 了 le (展开了热烈的讨论 zhankai le relie de taolun carry out LE lively discussion 'carry out lively discussion'), while 开展 kaizhan is showing dis-preference over 了 le. 展开 zhankai in Mainland also has the feature 'interaction event' in its favor (e.g., 展开讨论/合作/谈判 zhankai

*taolun/hezuo/tanlun* carry out\_discuss/cooperate/negotiate 'carry out discussion/cooperation/negotiation') while 开展 *kaizhan* shows the opposite tendency.

Features		<b>TW</b> 展开	TW开展
POS	Ν	-	+
	Deverbal	+	-
	noun		
	VP	0	+
VO		0	0
Event complement		-	+
at subject position			
Take aspectual		-	-
marker (le)			
Formal event		0	0
Event involving		-	+
interaction of the			
agent and patient			
Accomplishment		0	0
event			

Table 3.14 Comparison between Taiwan 开展 kaizhan and 展开 zhankai

As presented in Table 3.14, similar to Mainland data, Taiwan 开展 kaizhan prefers the verbal complement (e.g., 开展合作 kaizhan hezuo carry out\_cooperate 'to cooperate') while 展开 zhankai tends take nominal complement (e.g., 展开决战 zhankai juezhan carry out\_decisive\_battle 'fight a decisive battle'). Compared to 展开 zhankai, the complement of 开展 kaizhan tends to occur at the subject position (e.g., 两岸贸易的正常开展 liang'an maoyi de zhengchang fazhan cross-strait\_trade\_DE\_normal\_carry out 'trade cross-strait is carried out normally'). Also, 开展 kaizhan has shown its favor for accomplishment event (e.g., 开展交流活动 kaizhan jiaoliu huodong carry out interchange activity 'carry out interchange activity').

#### 3.4.2.1.4 Summary

The results of variety internal comparison show that each near-synonym pairs can be differentiated by certain features. What is more important, the features that can differentiate near-synonym pairs actually show differences between varieties. For example, Mainland 加以 *jiayi* and 进行 *jinxing* are differ in syntactic type of the complements, word order, aspectual marker, semantic properties of the complement, polarity of the context while Taiwan 进行 *jinxing* and 加以 *jiayi* show difference in taking VO compound, psychological event, accomplishment event and the complexity of the complement as well as the polarity of the context. In other words, Mainland and Taiwan light verbs show differences in the factors that influence the light verb alternative choices.

### 3.4.2.2 Light Verb Variation: Variety External Comparison

With respect to the external comparison of the same light verb in different varieties of Chinese, the result of statistical analyses can also be able to present the tendency differences. As shown below:

Features		ML进行	TW 进行
POS	Ν	0	0
	Deverba	+	-
	l noun		
	VP	-	+
VO complement		-	+
Event		0	0
complement at			

3.4.2.2.1 进行 jinxing in Taiwan and Mainland

subject position			
Take aspectual		+	-
marker (le)			
Take asp	ectual	0	0
marker (	guo)		
Durative event		0	0
Formal event		0	0
Psychological		0	0
event			
Event involving		+	-
interaction	on of the		
agent and patient			
Accomplishment		0	0
event			
Attributive of the		0	0
complement			
polarity	positive	0	0
	neutral	+	-
	negative	-	+
Occur with other		-	+
lv			
Spontaneous		0	0
event			

Table 3.15 Comparison between 进行 jinxing in ML and TW

As presented in Table 3.15, the usage of 进行 *jinxing* in Mainland and Taiwan have differences at semantic, syntactic as well as discourse level. The most obvious difference is that the Taiwan 进行 *jinxing* can take VPs as the complement, as in 处理 临时提案 *chuli linshi ti'an* process\_interim\_proposal 'process the interim proposal', while Mainland 进行 *jinxing* prefers deverbal noun as the complement (e.g., 进行长期的合作 *jinxing changqi de hezuo* proceed\_long-term\_DE\_cooperate 'to establish a long-term cooperation'). This specific usage of Taiwan might shed light on the classification of light verb complements. It may indicate different status of taken complements in each

variety is located in the process of de-verbalization. I will discuss this in detail in Section 3.5.

Furthermore, the result demonstrates that TW 进行 *jinxing* is more likely to take VO compounds, as in 进行开票/投票 *jinxing kaipiao/toupiao* proceed\_ballot-counting/vote 'ballot-counting/voting'. With respect to the semantic property of the taken complement, 进行 jinxing in Mainland tends to more with interactive complements like 磋商/商討 co-occur cuoshang/shangtao 'negotiation'. In terms of the polarity of the context, Mainland 进行 *jinxing* is more likely to be co-occurred in neutral context (e.g., 进行研究 jinxing yanjiu proceed research 'to conduct research') while Taiwan 进行 *jinxing* prefers the negative context (e.g., 进行非法人口贩卖和贩毒活动 jinxing feifa renkou fanmai he fandu huodong proceed illegal human trafficking and drug trafficking activity 'proceed the activities of illegal human trafficking and drug trafficking'). Another significant variation difference is Taiwan 进行 jinxing shows the preference of cooccurring with other light verb (e.g., 继续进行会议 jixu jinxing huivicontinue proceed meeting 'continue the meeting') while the Mainland counterpart shows the opposite tendency.

### 3.4.2.2.2 加以 jiayi in Taiwan and Mainland

加以 jiayi shows the similar tendency as 进行 jinxing between Mainland and

Taiwan, as shown in Table 3.16.

Features		ML加以	TW加以
POS	Ν	0	0
	Deverba	+	-
	l noun		
	VP	-	+
VO		-	+
Event		0	0
complem	ent at		
subject p	osition		
Take asp	ectual	0	0
marker (	le)		
Take aspectual		0	0
marker (	guo)		
Durative	event	-	+
Formal event		0	0
Psychological		0	0
event			
Event involving		0	0
interaction of the			
agent and patient			
Accomp	ishment	+	-
event			
Attributive of the		0	0
complement			
Polarity	Positive	+	-
	Neutral	-	+
	Negativ	0	0
	e		
Occur with other		0	0
lv			

Table 3.16 Comparison between 加以 jiayi in ML and TW

As presented in Table 3.16, similar with 进行 *jinxing*, 加以 *jiayi* in two varieties also differ in that TW 加以 *jiayi* has preference in taking VP and VO complements while 加以 *jiayi* in Mainland shows the opposite preference. The examples like 加以调整实用性 *jiayi tiaozheng shiyongxing*
inflict\_adjust\_practicability 'make adjustment to the practicability' can only be found in Taiwan data. In addition, Mainland 加以 *jiayi* tends to co-occur more with accomplishment event (e.g., 加以澄清 *jiayi chengqing* inflict\_clarify 'make clarification'), while Taiwan 加以 *jiayi* prefers durative events (e.g., 加以学习/了解 *jiayi xuexi/liaojie* inflict\_learn/understand 'to learn/understand'). In terms of polarity of context, Mainland 加以 *jiayi* prefers the positive context (e.g., 加以保护 *jiayi baohu* inflict\_protect 'to protect') while Taiwan 加以 *jiayi* tends to be appeared in neutral context (e.g., 加以辅导 *jiayi fudao* inflict\_coach 'to coach').

# 3.4.2.2.3 做 zuo in Taiwan and Mainland

As displayed in Table 3.17, the variation difference of 做 *zuo* mainly lies in the selection of the taken complements. Taiwan 做 *zuo* almost has no constraints in taking the complements while the complements of Mainland 做 *zuo* are relatively limited in types. Examples show that Taiwan 做 *zuo* can take a wide range of complements compared to Mainland 做 *zuo*. LVCs as 做人身攻击 *zuo renshengongji* do\_personal\_attack 'make personal attack', 做环保 *zuo huanbao* do\_environment\_protect 'protecting the environment', 做 竞 赛 *zuo jingsai* do\_competition 'have competition', 做 上 述 表 述 *zuo shangshu biaoshi* do\_above\_statement 'to issue above statement', 做呼吁 *zuo huyu* do\_appeal 'to appeal', 做分配 *zuo fenpei* do\_allocate 'to allocate' can only be found in

Taiwan corpus. All the specific usages of Taiwan 做 *zuo* have been annotated and shown in Appendix I.

Features			TW
Complement	Event noun	+	•
Types	De-verbal	-	+
	noun		
Event complement at subject			-
position			
Take aspectual marker (le)			-
Take aspectual marker (guo)			0
Durative event			0
Formal event			+
Accomplishment ev	vent	-	+

Table 3.17 Comparison between 做 zuo in TW and ML

The table also suggests that the light verb 做 *zuo* in Taiwan has the similar preference of taking deverbal noun as the objects (same with 进行 *jinxing* and 加以 *jiayi*, e.g., 做<u>表示/考量</u> *zuo biaoshi/kaoliang* do\_statement/consideration 'issue official statement/give consideration'), while Mainland 做 *zuo* is more likely to co-occur with event noun as 工作 / 活动 / 手术 *gongzuo/huodong/shoushu* 'do work/activity/operation'.

TW 做 zuo is also likely to take formal event as the complement (政府对 上述事件做表示 zhengfu dui shangshu shijian zuo biaoshi government\_for\_above\_affair\_do\_statement 'the government issued official statement for this affairs'), while more informal complements can be found in Mainland usages (做小生意/买卖 zuo xiaoshengyi/maimai do\_small\_business 'doing small business').

# 3.4.2.2.4 搞 gao in Taiwan and Mainland

Features		ML	TW
Complement Types	Event noun	-	+
	De-verbal	+	-
	noun		
Event complement at subject po	sition	+	-
Take aspectual marker (le)		+	-
Formal event	-	+	
Event involving interaction of	+	-	
patient			
Accomplishment event	0	0	
Attributive of the complement	+	-	
Connotation of the	positive	+	-
Complement	neutral	+	-
	negative	-	+

Table 3.18 Comparison between 搞 gaoin TW and ML

As shown in Table 3.18, the taken complement of Taiwan 搞 gao has significant correlation with the feature 'negative context' (This finding is also consistent with Diao, 2012f).

Construction examples like 搞形式主义/和平演变/分裂/抹黑 gaoxingshizhuyi/hebingyanbian/fenlie/mohei do\_formalism/peaceful evolution/dispute/throw-mud 'take formalism approach/make peaceful evolution/cause state disruption/throw mud' are typical usages which are frequently appeared in Taiwan corpus. In contrast, Mainland 搞 gao can take a variety of complements, including both formal (搞国有企业经营 gao guoyouqiye jingying do\_state-owned\_enterprise\_manage 'managing stateowned enterprises') and informal event (搞批发 gao pifa do\_wholescale trade 'do wholescale trade'), interactive events (搞辩论 gao bianlun do\_debate 'engage in debate') and complements with attribute (搞了很长时间的科技承 包 gao le henchang shijian de keji chengbao do\_LE\_very\_long\_time\_DE\_technology\_manufacture 'engage in technology contract manufacturing for a very long time'). Also, the polarity of the context for Mainland 搞 gao can be positive (搞先进性学习 gao xianjinxing xuexi do\_advance nature\_learn 'learn about advanced nature'). All the specific usages of Mainland 搞 gao have been annotated and shown in Appendix II.

In general, for 做 *zuo* and 搞 *gao* in both Mainland and Taiwan, we can summarize as Taiwan 做 *zuo* can take a much wider range of complements under different context, whereas Taiwan 搞 *gao* can barely be used in other context except for political context with the negative meaning. Mainland 做 *zuo* and 搞 *gao* have shown the opposite tendency.

## 3.4.2.2.5 展开 zhankai in Taiwan and Mainland

Features		ML展开	TW展开
POS	Ν	-	+
	Deverbal	+	-
	noun		
	VP	-	+
VO		0	0
Event co	omplement	+	-
at subject position			
Take asp	oectual	+	-
marker (le)			
Formal e	event	0	0
Event in	volving	+	-

The result of Chi-square test is shown below in Table 3.19.

interaction of the		
agent and patient		
Accomplishment	0	0
event		
Attributive of the	+	-
complement		

Table 3.19 Comparison between Mainland and Taiwan 展开 zhankai

As seen in Table 3.19, 展开 zhankai in two varieties can be differentiated by five features. In terms of syntactic type of the complement, Taiwan 展开 zhankai shows the preference of taking VP (e.g., 展开调查特产店内的鱼片 zhankai diaocha techandian nei de yupian carry out investigate specialty store in DE fish filled 'to investigate the fish filled in specialty store') and nominal complement (e.g., 展开决战 zhankai juezhan carry out decisive battle 'fight a decisive battle'), while Mainland counterpart prefers the deverbal noun (e.g., 展开消费的宣传 zhankai xiaofei de xuanchuan carry out consumption DE propaganda 'carry out consumption propaganda'). 展开 zhankai in Mainland tends to be appeared in subject position (e.g., 建设工 程 全 面 开 展 jiashe gongcheng zhankai quanmian construction\_engineering overall carry out 'construction engineering is fully carried out') and aspectual marker is more likely to be co-occurred with Mainland 展开 zhankai (e.g., 展开了世界冠军争夺战 zhankai le shijie guanjun zhengduosai carry out world champion competition 'have contested the world championship'). Also, Mainland 展开 zhankai shows its preference for interactive complements (e.g., 展开交流与合作 zhankai jiaoliu yu hezuo carry out communicate and cooperate 'to communicate and cooperate') and complex complements (e.g., 展开了不懈的斗争 *zhankaile buxie de douzheng* carry out LE endless DE struggle 'has carried out endless struggle').

# 3.4.2.2.6 开展 kaizhan in Taiwan and Mainland

Variation differences in  $\mathcal{F}$   $\mathbb{R}$  *kaizhan* between two varieties can also be observed by the result of Chi-square test.

Features		ML开展	<b>TW</b> 开展
POS	Ν	-	+
	V	+	-
	VP	0	0
VO		0	0
Event		-	+
complen	nent at		
subject p	osition		
Take aspectual		+	-
marker (le)			
Formal e	event	0	0
Event in	volving	-	+
interaction	on of the		
agent and patient			
Accomplishment		0	0
event			
Attributi	ve of the	0	0
complen	nent		

Table 3.20 Comparison between Taiwan and Mainland 开展 kaizhan

As can be seen from Table 3.20, nominal object is preferred by Taiwan 开 展 kaizhan (e.g., 开展社会实践活动 kaizhan shehui shijian huodong carry out\_social\_practice\_activity 'carry out social practice activity'), while the verbal object is favored by Mainland 开展 kaizhan (e.g., 开展自我教育 kaizhan ziwo jiaoyu carry out\_self-education 'carry out self-education'). In terms of word order, the eventive complements of 开展 kaizhan in Taiwan tend to be shown before the light verb (e.g., 两岸贸易正常开展 *liang'an maoyi de zhengchang fazhan* cross-strait\_trade\_DE\_normal\_carry out 'trade cross-strait is carried out normally'), while the complements of Mainland 开展 *kaizhan* tend to follow the light verb (e.g., 开展学习的活动 *kaizhan xuexi de huodong* carry out\_learn\_DE\_activity 'carry out learning activity'). In terms of aspectual marker, 开展 *kaizhan* in Taiwan dis-prefers aspectual marker le while its Mainland counterpart has significant preference for le (e.g., 对超晶格开展了研 究 *dui chaojingge kaizhan le yanjiu* for\_super-lattice\_carry out\_LE\_research 'to conduct research for super-lattice'). Another significant difference is: compared to Mainland 开展 *kaizhan*, the Taiwan counterpart shows significant preference for interaction event (e.g., 积极展开'弹性外交' *jiji zhankai tanxing waijiao* active\_carry out\_flexible\_diplomacy 'carry out Flexible diplomacy actively').

#### 3.4.3 Summary

As presented above, the variation differences for each light verb can be effectively differentiated by the annotated features. Furthermore, dividing the light verb into near-synonym pairs can allow us to explore both light verb alternations (alternative choices between similar light verbs) and light verb variations (light verb between different varieties). The results show that the variations of light verbs are not as simple as feature or behavior difference, but in fact difference in the patterns of alternations (in their alternative choices). In addition, the light verb differences between Mainland and Taiwan shown above actually have some theoretical implications. Two points are mainly discussed here: One is the re-classification of syntactic type of the light verb taken complements, and the other is the variation differences in degree of transitivity between Mainland and Taiwan light verbs. Each point will be discussed in detail in the following section.

# **3.5 The Syntactic Type of Taken Complement**

As presented by the results of statistical tests (both logistic regression model and Chi-square test), there is one feature that is very effective in differentiating light verb pair and light verb variation -- the POS of the complement. Every light verb pair as well as every light verb variation has shown significant differences in the syntactic type of the taken complement. Based on the statistical result, a re-classification of syntactic type of the taken complement is proposed in this section.

#### **3.5.1 Analysis and Discussion**

Agreement has been reached that only a small part of noun can be taken by light verbs (e.g., 会议/赛事 *huiyi/saishi* 'conference/competition') while most of the complements taken by 进行 *jinxing* are originally verbs (e.g., 进行研究/ 改革 *jinxing yanjiu/gaige* proceed\_research/reform 'conduct research/carry on reform'). Different from common verb, most of these taken complements, being appeared in the object position, have already possessed some properties of

nominal phrase in terms of its syntactic behavior (e.g., Zhu, 1985). Evidence as shown below,

a. The eventive complement usually cannot take another object: \*进行研究这份文档 *jinxing yanjiu zhe fen wendang* proceed\_research\_this\_CL\_document 'conduct research on this document';

b. The eventive complement cannot be modified by adverbial directly: \* 进行在会议中研究 *jinxing zai ihuiyizhong yanjiu* proceed\_at\_conference\_research 'conduct research on the conference';

c. The complement can be modified by the attribute: 进行长达二十年的 研 究 *jinxing changda ershinian de yanjiu* proceed\_long\_twenty\_years\_DE\_research 'conduct a twenty-year research';

d. It can be questioned by 什么 *shenme* 'what' instead of 怎么样 *zenmeyang* 'how': 在进行**什么** *zai jinxing shenme* at\_proceed\_what 'What is being proceeded' \*在进行**怎么样** *zai jinxing zenmeyang* at\_proceed\_how '\*How is being proceeded?'

Previous studies have numbers of debates on which syntactic type do these verbal-like complements belong to. Some studies consider the taken complement as a 名动词/动名词 *mingdongci/dongmingci* 'nominal verb' which has both nominal and verbal properties (e.g., Zhu, 1985; Chen, 1987), while others hold the opinion that the complement has already been transferred from the original verb to a typical noun through the process of nominalization (e.g.,

Xiao, 1955; Li, 1990).

However, in this study based on large corpora, I find that the actual usages may be much more complex than what has been presented in the previous studies, especially when the variation differences between language varieties are taken into consideration. In Taiwan Mandarin, light verbs can be followed by a variety of NPs<sup>3</sup> (e.g., 进行议程/君子之争 *jinxing yicheng/junzizhizheng* proceed agenda/gentleman's dispute 'to proceed with the assembly's agenda/having a gentleman's dispute'). Besides, although previous studies all claim that the light verb cannot take a Verb-Object phrase as the complement (e.g., Zhu, 1985; Hu and Fan, 1994; Qiu, 1994; Du, 2010 among others), in Taiwan corpora examples like 进行处理临时提案 jinxing chuli linshi ti'an proceed\_process interim proposal 'process the interim proposal' 加以甄选优 秀学生 jiayi zhenxuan youxiu xuesheng inflict select excellent student 'to select excellent student' and 展开调查食材来源 zhankai diaocha shicai *laiyuan* carry out investigate food source 'to investigate the source of food' can be found, in which the complements taken by 进行 *jinxing*, 加以 *jiayi*, 展  $\pi$  zhankai can be a VO phrase. The VO phrase can even be odified by adverbial (3-1):

3-1. 政府目前正进行 对中非共和国提供养鱼、养虾的技术. zhengfu muqian zheng jinxing dui zhongfeigongheguo tigong yangyu yangxia de jishu Government\_now\_ZHENG\_proceed\_for\_Central\_Africa\_Republic\_pro vide\_fish\_farming\_shrimp\_farming\_DE\_technology

 $<sup>^{3}</sup>$ NPs, here refers to nouns that describe a situation (cf. nouns denoting an entity such as 苹果 'apple' and 桌子 'table'). But unlike deverbal noun (研究 'research'), a NP never has a verbal use (Huang et al., 2013).

'The government is now providing the technology of fish and shrimp farming for the Central Africa Republic.'

This may pose a challenge to the traditional classification. Therefore, in my annotation standard, three types of taken complements have been distinguished in terms of their syntactic types: event noun, de-verbal noun and VP. The comparative study shows that there are also distributional variation differences in syntactic type of the complement between different Taiwan and Mainland.

Event noun refers to a sub-type of noun which lexically encodes eventive information, including event structure and time (Wang and Huang, 2011). The proportion of taking a noun as the complement is relatively small (e.g., for 进行 *jinxing*, about 20.1% in Mainland). Taiwan can take more types of NPs (e.g., 进 行感恩之旅 *jinxing gananzhil ü* proceed\_thanksgiving\_trip 'have thanksgiving trips') with a little bit lower frequency (18.87%). Since the definition of an event noun is clear, this study mainly focuses on differentiation of the other two types (de-verbal noun and VP).

In this proposal, I argue that 研究 *yanjiu* 'research' in 进行可行性研究 *jinxing kexingxing yanjiu* proceed\_practicability\_research 'conduct research on the practicability' and 进行研究可行性 *jinxing yanjiu kexingxing* proceed\_research\_practicability 'conduct research on the practicability' belong to different syntactic types. Generally speaking, the former 研究 *yanjiu* 'research' can be considered as a de-verbal noun which has undergone the process of de-verbalization. While in 进行研究可行性 *jinxing yanjiu*  *kexingxing* proceed\_research\_practicability 'conduct research on the practicability', 研究 *yanjiu* 'research' should still be considered as typical verbs. The whole phrase should be treated as a VP. This will be illustrated in detail in the following section.

Shi (2011) differentiates two kinds of eventive phrases which can appear in the object position, by using a set of syntactic tests.

3-2. a. 我们需要注意环境保护。
women xuyao zhuyi huanjing baohu
1 PL\_need\_pay attention to\_environment\_protect
'We need to pay attention to environmental protection.'
b. 我们需要注意保护环境。
women xuyao zhuyi baohu huanjing
1 PL\_need\_pay attention to\_protect\_environment

'We need to pay attention to protect environment.'

He argues that due to its internal nominal structure and the typical nominal syntactic behavior, 保护 *baohu* 'protect' in sentence 3-2a should be considered as a Noun (can be modified by attribute). But in sentence 3-2b, 保护 *baohu* 'protect' here still has syntactic behavior of verbs (e.g., can take object and complement). The whole phrase 保护环境 *baohu huanjing* protect\_environment 'protect the environment' still has the verbal internal structure (V-O structure) and can be used as predicate (e.g., as in 我们要保护环境 *women yao baohu huanjing*1 PL\_need\_protect\_environment 'We need to protect the environment.') (Shi, 2011).

With respect to light verb constructions, the variation differences have been observed in Gigaword corpus can be well classified by this standard. For the usage of Mainland LVC, the patient can only appear before the eventive complement, either before the light verb (3-3a) or between light verb and the taken complement with DE (3-3b), which shows the typical properties of a Noun.

3-3. a. 对分配制度进行改革 dui fenpei zhidu jinxing gaige for\_distribution\_system\_proceed\_reform 'carry on reforms of distribution system' b. 进行分配制度的改革 jinxing fenpei zhidu de gaige proceed\_distribution\_system\_DE\_reform 'carry on reforms of distribution system'

But in Taiwan specific usage, the object 改革 gaige 'reform' can be followed by the patient 分配制度 fenpei zhidu 'distribution system' (3-4a), which shows that 改革 gaige 'reform' here still remains the property of a typical verb of taking another object. Moreover, 改革分配制度 gaige fenpei *zhidu* reform distribution system 'carry on reform of distribution system' as a VO phrase is considered as a VP in the sense that it has the verbal internal structure (V-O structure) and can be used as predicate (政府改革分配制度 government\_reform\_distribution\_system zhengfu gaige fenpeizhidu 'government is carrying on reforms of the distribution system'). Also, the whole VO phrase as a VP can be modified by adverbial (3-4b). Hence, I use 'VP' to distinguish it from a de-verbalized Nominal Phrase (e.g., 分配制度的改革 fenpeizhidu de gaige distribution system DE reform 'the reformation of distribution system')

3-4. a. 进行改革分配制度 *jinxing gaige fenpeizhidu* proceed\_reform\_distribution\_system 'to carry on reforms of distribution system'

b. 政府目前正进行对有关部门改革分配制度 *zhengfu muqian zheng jinxing dui youguanbumen gaige fenpeizhidu* government\_now\_ZHENG\_proceed\_for\_relevant\_department\_reform\_distribut ion\_system 'The government now is corruing on reforms of distribution system in all

'The government now is carrying on reforms of distribution system in all relevant departments.'

In this regard, I adopt Shi (2011)'s analysis and treat the 改革 gaige'reform' in mainland usage as a de-verbal noun which has been transferred from the verb through the process of de-verbalization. While in Taiwan specific usage, 改革 gaige 'reform' keeps the verbal property of taking another object (e.g., 进行改 革分配制度 jinxing gaige fenpeizhidu proceed\_reform\_distribution\_system 'make reforms of distribution system'). Moreover, examples have also shown that in Taiwan usages, when eventive complement is modified by a prepositional structure, DE-insertion is not always necessary (3-5).

3-5. 进行对大陆宣传.

*jinxing dui dalu xuanchuan* proceed\_for\_Mainland\_propaganda 'conduct propaganda to Mainland'

While in Mainland, underthis context, 'DE' has to appear to license the grammaticality of the construction (e.g., 进行对大陆的宣传 *jinxing dui dalu de xuanchuan* proceed\_for\_Mainland\_DE\_propaganda 'conduct propaganda to Mainland'). As prepositional structure cannot modify NP directly without DE (e.g., Huang and Liao, 2007), the examples found in Taiwan data is another evidence to prove that the taken complement in Taiwan usage can be a real verb

rather than a deverbal noun.

One thing should be addressed is that in Taiwan corpus, both constructions are commonly used, i.e. the patient can be the modifier of the complement (进行可行性研究 *jinxing kexingxing yanjiu* proceed\_practicability\_research 'conduct research on the practicability') as well as follow it (进行研究可行性*jinxing yanjiu kexingxing* proceed\_research\_practicability 'conduct research on the practicability'), while in Mainland only the former construction is acceptable. This displays that Taiwan light verb has less constraints in taking the complements (can take both NP and VP, with and without DE). There are more alternative patterns for LVC in Taiwan. Also, for the examples which VPs are taken as the complements, light verbs can be omitted without changing the grammatical acceptability of the sentences. The light verb itself can be totally redundant. As shown below in 3-6a:

3-6. a. 先(进行) 调整 国内有关法律以做好配合欧盟司法合作案件 xian (jinxing) tiaozheng guonei youguan falv yi peihe oumeng sifa hezuo anjian first\_(proceed)\_adjust\_domestic\_relevant\_law\_in\_orderto\_do\_goo d\_cooperate\_EU\_judicial\_cooperate\_case 'make adjustment on relevant domestic laws in order to cooperate with EU judicial cooperation cases'

In contrast, in Mainland usages, although light verb does not contribute to the eventive information, it is still compulsory in order to licenses the grammaticality of the sentences structure (3-6b). b.\*政府正对少年儿童(进行)正确消费教育

*zhengfu zhengdui shaonianertong (jinxing) zhengque xiaofei jiaoyu*government\_ZHENG\_for\_young\_children\_(proceed)\_right\_cons umption\_education

'The government is educating the young children about consumption view.'

## **3.5.2 Application and Implication**

In this section, three types of taken complements are clearly differentiated: event noun, de-verbal noun and VP. This classification can cover all the examples observed in the corpus. The re-classification proposed in this study may also help with the annotation standard for POS tagging in the corpus. As can be seen, light verb can take event noun, deverbal noun and VP as the complement, but the morphological status of these categories are typically unmarked (Lin et al., 2014), that may make the corpus identification complicated. For example, I have found several inconsistencies in the POS tagging for the taken complements in some corpora. In 国家语委现代汉语通 用平衡语料库 (CNcorpus) (Institute of Applied Linguistic Ministry of Education, 2009), when either 战斗 *zhandou* 'battle' or 斗争 *douzheng* 'fight' is used individually, it is annotated as a Verb. However, if they appear after the light verb, the annotation results are sometimes confusing, as shown below:

3-7. a. 每日/nt 都/d 在/p 进行/v 剧烈/a 的/u/连绵/a 的/u 战斗/v meiri douzai jinxing juliede lianmiande zhandou every day\_all\_at\_proceed\_fierce\_DE\_continous\_DE\_battle 'Every day they are having fierce and continuous battles.'
b. 进行/v 了/u 坚决/a 的/u 斗争/n jinxing le jianjue de douzheng proceed\_LE\_resolute\_DE\_fight 'had a resolute fight'

In 3-7 a andb, both 战斗 *zhandou* 'battle' and 斗争 *douzheng* 'fight' are modified by the attribute in DE construction, but with different tagging. The inconsistency of POS taggingmay brings a variety of inconveniences for grammatical analyses.

The differentiation and classification would be very clear according the reclassification system proposed above. Both 战斗 *zhandou* 'battle' and 斗争 *douzheng* 'fight' in example 3-7 should be treated as a de-verbal noun. But if they are followed by another object (e.g., 进行斗争不良行为 *jinxing douzheng buliang xingwei* proceed\_fight\_misconduct 'to fight with misconduct'), they should be treated as a verb.

#### 3.5.3 Summary

To make a summary, in this section, the syntactic type of the light verb taken complement is well differentiated and re-classified. Based on the distributional differences between Mainland and Taiwan light verbs, a re-classification system is proposed and the taken complements of light verbs are divided into three types: event noun, de-verbal noun and VP. Both logistic regression model and Chi-square test show that Taiwan light verbs significantly prefer to take VP as the complement while the Mainland counterparts favor the de-verbal noun. Moreover, in terms of application and implication, the re-classification of the taken complements can also help to establish the annotation standard for POS tagging in the corpus.

## **3.6 Transitivity Variations in Light Verb Constructions**

Based on their distributional differences between Taiwan and Mainland Mandarin, this study further argues that the light verb itself may differ in its degree of transitivity between these two varieties.

The concept of transitivity has always been a very important issue in linguistics (Hopper and Thompson, 1980; Lee, 2010). Transitivity is traditionally understood as a property of verbs that determines whether a verb can take objects or not (Robins, 1964; Richards, Platt and Weber, 1985). Hence verbs are often categorized into transitive verbs such as 'kick' which can take a direct object, and intransitive verbs such as 'cry' which cannot take an object. Instead of merely depending on the morphosyntactic features, later approaches define transitivity as a continuum rather than a binary category. To be particular, continuous approach takes the degree to which an action affects the objects into account (Lee, 2010). For example, the verb 'hit' is considered to have higher transitivity than the verb 'like' since only actions can be transferred from one participant to another. I will briefly discuss the most important continuous approach proposed by Hopper and Thompson (1980).

Hopper and Thompson (1980) identify transitivity as a composition notion, consisting of ten interacting but basically separate parameters. Each of the ten parameters suggests a scale according to which clauses can be ranked, as shown in Table 3.21.

Criterion	High	Low		
A. Participants	2 or more participants,	1 participant		
	A and O.			
B. Kinesis	Action	Non-action		
C. Aspect	Telic	Atelic		
D. Punctuality	Punctual	Non-punctual		
E. Volitionality	Volitional	Non-volitional		
F. Affirmation	Affirmative	Negative		
G. Mode	Realis	Irrealis		
H. Agency	A high in potency	A low in potency		
I. Affectedness of O	O totally affected	O not affected		
J. Individuation of O	O highly individuated	O non-individuated		

 Table 3.21 Transitivity Parameters (Hopper and Thompson, 1980: p.252)

The explanation of each parameter is shown below:

A. Participant: transfer can only take place with two participants being involved.

B. Kiness: actions rather than states can be transferred from one participant to another.

C. Aspect: a telic action is more effectively transferred to an atelic action.

D. Punctuality: actions carried out with no obvious transitional phase between start-point and endpoint, have a more marked effect on their patients than actions which are inherently on-going.

E. Volitionality: the effect on the patient is typically more apparent when the agent is presented as acting purposefully.

F. Affirmation: affirmative has more effect than negative.

G. Mode: This is related to the 'realis' and 'irrealis' distinction. An action which either did not occur, or which is presented as occurring in a non-real world, is obvious less effective than the one which occurs as a real event.

H. Agency: participants high in Agency can affect a transfer an action in a way that those low in Agency cannot.

I. Affectedness of object: the degree to which an action is transferred to a patient is a function of how completely that patient is affected.

J. Individuality of object: it refers both to the distinctness of the patient from the agent and to its distinctness from its own background.

As reviewed above, the transitivity degree of a verb can be measured by the ability of taking the objects (from traditional approach) and by what kind of complements can be taken (from continuous approach). Hence, transitivity will be measured in both ways in this study. To be particular, in this study for light verbs, the ability of taking the objects can be measured by the transitivity frequency and the semantic/syntactic properties of the taken complements can be illustrated by the annotated features.

For the variation differences in transitivity degree, I argue that Taiwan light verbs may have a higher degree of transitivity compared to its Mainland counterparts. This tendency is shown in two perspectives: variations in transitivity frequency and in semantic /syntactic properties of the taken complement.

In terms of variations in transitivity frequency, Taiwan light verbs have the tendency of taking more types of NP and VPs as the complements with less collocation constraints, as shown by the results of logistic regression model and Chi-square test in Section 3.4. Taiwan light verb is 'more transitive' in the sense that it is more likely to be used transitively in different contexts. It is worth noting that the versatility of argument taking refers to argument types, instead of semantic types. Higher versatility of argument types indicates higher degree

of transitivity, while the higher versatility of semantic types may imply the more lightness of light verbs. These are two different concepts and the correspondences between these two versatilities can be a very good topic for future studies.

Besides this, the variation differences in semantic and syntactic properties of the taken complements (shown by the result of logistic regression model and Chi-square test) also demonstrate the higher transitivity of Taiwan light verbs. As mentioned before, Hopper and Thompson (1980) have proposed 10 parameters to identify the degree of transitivity, as shown in Table 3.21. Among all these 10 parameters, 2 of them are directly related to the taken objects: individuality and affectedness of the object. Since for light verb construction, the taken object is the one which contains the important information, these two parameters should be effective in identifying light verb constructions. As shown by the statistical results (shown in Section 3.4), the distributional variations in semantic and syntactic properties between Taiwan and Mainland light verbs do indicate the variation differences in these two properties: Taiwan light verbs tend to have higher degree of transitivity. Basically there are two main evidences to show the tendency.

First evidence is in Section 3.4.1, the Mixed-effect logistic regression model result shows that, Taiwan light verbs can be followed by more types of event nouns (e.g., 进行议程/君子之争 *jinxing yicheng/junzizhizheng* proceed\_agenda/gentleman's\_dispute 'to carry out the assembly's agenda/having a gentleman's dispute'), this may indicate that the complements

of Taiwan light verbs tend to be more highly individuated. Since the event noun, compared to either de-verbal noun or VP, may have more 'nouny' or 'referential' property. The higher degree of individuality may refer to the higher degree of individuality of Taiwan light verbs.

Another evidence is: as shown in both mixed-effect logistic regression and Chi-square test, Taiwan light verb has the preference of taking VP as the complements while the Mainland counterparts prefer de-verbal nouns. VP like 改革分配制度 *gaige fenpei zhidu* reform\_distribution\_system 'carry on reform of distribution system', compared to the de-verbal noun 改革 *gaige* 'reform', is obviously higher in individuality as the patient is overt. Moreover, as the patient is already appeared, the affectedness of the objects in Taiwan preference is also higher. The higher individuality and higher affectedness of the object both indicate the higher degree of transitivity. In other words, the preference of VP in Taiwan actually indicates the higher transitivity of Taiwan light verbs.

To make a summary, the higher transitivity of Taiwan light verbs are basically lies in two perspectives: one is that Taiwan light verbs verb is 'more transitive' in the sense that it is more likely to be used transitively in different contexts. The other is that Taiwan light verb more prefer to take event noun and VP as the complement, which indicates the higher individuality and affectedness of object and further indicates the higher degree of transitivity of the light verbs.

## **3.7 Variation in Alternation of Light Verb Construction**

#### **3.7.1 Introduction**

As has been claimed in the introduction section (Section 3.1), for light verb construction, the contentive and eventive information mainly comes from the taken complement while the light verb itself is semantically bleached. Therefore the taken complement of the light verb is always verbal (e.g., 改进/ 研究 gaijin/yanjiu 'improve/research' in 加以改进/进行研究 jiayi gaijin/jinxing vanjiu inflict improve/proceed research 'to improve/to conduct research'). In that sense, if the verbal complement contains two arguments semantically, it is natural for it to have another theme (whether internal or external). For example, in the construction 进行研究 jinxing yanjiu proceed research 'to conduct research', 研究 yanjiu 'research'is a twoargument predicate, it is possible for it to have a theme at semantic level (e.g., 可行性 kexingxing 'practicability'), in Mainland Mandarin, the most frequently pattern is 对可行性进行研究 dui kexingxing jinxing yanjiu for practicability proceed research 'to conduct research on practicability' (Cai, 1982). According to the corpus observation, I have observed that there are different alternative patterns to introduce the theme of the verbal object. Moreover, variation differences in alternative patterns between Mainland and Taiwan Mandarin have also been detected in corpus.

For example, for the light verb 进行 *jinxing* in Mainland, there are mainly three alternations to introduce the theme of the verbal complement (3-8, 3-9 and

3-10):

3.8 对可行性进行研究 duikexingxing jinxing yanjiu for\_practicability\_proceed\_research 'to conduct research on practicability'
3.9 进行可行性研究 jinxingkexingxingyanjiu proceed\_practicability\_research 'to conduct research on practicability'
3.10 进行(对)可行性的研究 jinxing (dui) kexingxing de yanjiu proceed\_(for)\_practicability\_DE\_research 'to conduct research on practicability'

Among all the three alternations, almost every native speakers of Mainland Mandarin would agree that the first one is the most frequently used (对可行性 进行研究 duikexingxing jinxingyanjiu for\_practicability\_proceed\_research'to conduct research on practicability'). In this case, the theme is introduced by a prepositional structure and appeared before the light verb. The second one is also quite common (进行可行性研究 jinxing kexingxingyanjiu proceed\_practicability\_research'to conduct research on practicability'), the theme here is used as a modifier of the taken complement. The third one is not very natural but still can be found in Mainland corpus, the prepositional structure (对可行性 duikexingxing 'for practicability') is appeared between the light verb and the taken complement. But in this case, the taken complement has to be nominalized and marked with de (e.g., 进行(对)可行性的研究 jinxing (dui) kexingxing de yanjiu proceed\_(for)\_practicability\_DE\_research 'to conduct research on practicability\_DE\_research 'to conduct research on practicability').

In Taiwan Mandarin, there are more alternative choices to introduce the

theme. Beside these three choices shown above, another two alternative constructions have also been detected in Taiwan corpus. One is construction like 讲 行 对 可行性研究 jinxing dui kexingxingyanjiu proceed\_for\_practicability\_research 'to conduct research on practicability', the prepositional structure is appeared between the light verb and the taken complement. It should be noted that in Mainland Mandarin, the prepositional structure is also possible to be occurred in between, but the complement has to be nominalized and marked with DE (进行对可行性的研究 *jinxing dui* kexingxing de yanjiu proceed\_for\_practicability\_DE\_research 'to conduct research on practicability'). Another alternation in Taiwan is 进行研究可行性 jinxing yanjiu kexingxing proceed\_research\_practiciability 'to conduct research on practicability', the taken complement can be followed directly by the theme. Both of these two alternative constructions cannot be accepted by Mainland Mandarin speakers; neither can they be detected in Mainland corpus.

These two specific alternative constructions of Taiwan Mandarin actually may indicate the difference in syntactic type of light verbs between Mainland and Taiwan, which has already been discussed in Section 3.5.

## 3.7.2 Research Question and Methodology

Based on the observations, two research questions needed to be addressed:

Empirically and statistically, is there any distributional difference between Mainland and Taiwan in terms of their alternative choices? Which factors can influence their alternation choices and are there any variation differences?

In this study, the two most typical light verbs 加以 *jiayi* and 进行 *jinxing* are taken as an example. To answer the questions, alternation constructions of 加以 *jaiyi* and 进行 *jinxing* needed to be annotated based on their actual usages in each variant. It should be mentioned that, sometimes the theme of the event is appeared in another clause (e.g., 对投资规模, 国家要及时采取措施进行微 *dui touziguimo, guojia yao jishi caiqu cuoshi jinxing weitiao* for\_invest\_scale\_country\_need\_intime\_take\_measure\_proceed\_fine-tuning 'in terms of investment scale, country needs to take measures to fine-tuning in time'). In this study, this kind of case should be excluded since it is more complicated than the alternation has been discussed above. In other words, this study only includes the case that the verbal object and the theme are in the same construction.

The data collection is conducted based on the extracted data. After excluding the complicated examples and the ones that without the theme, 111 进行 *jinxing* examples and 65 加以 *jiayi* tokens in Mainland Mandarin, as well as 119 进行 *jinxing* examples and 78 加以 *jiayi* tokens in Taiwan Mandarin have been collected. Based on this, the alternation types for each token in each variant have been annotated. There are in total 5 types of alternative patterns to introduce the theme, as shown in Table 3.22.

	Description	Examples
1 <sup>st</sup> type	Prepositional	对可行性进行研究
	structure before	dui kexingxing jinxing yanjiu
	light verb	for_practicability_proceed_research
		'to conduct research on practicability'
2 <sup>nd</sup> type	Theme as a	进行 <b>可行性</b> 研究
	modifier after	jinxingkexingxingyanjiu
	light verb	proceed_practicability_research
		'to conduct research on practicability'
3 <sup>rd</sup> type	prepositional	进行(对)可行性的研究
	structure is	jinxing (dui) kexingxing de yanjiu
	appeared	proceed_(for)_practicability_DE_research
	between light	'to conduct research on practicability'
	verb and taken	
	complement with	
	DE	
4 <sup>th</sup> type	prepositional	进行对 <b>可行性</b> 研究
	structure is	jinxing duikexingxingyanjiu
	appeared	proceed_for_practicability_research
	betweenlight	'to conduct research on practicability'
	verb and taken	
4	complement	
5 <sup>th</sup> type	Theme can	进行研究 <b>可行性</b>
	directly follow	jinxing yanjiukexingxing
	light verb	proceed_research_practiciability
	complement	'to conduct research on practicability'

Table 3.22 The 5 types to introduce the theme of event

# 3.7.3 The First Research Question

The distribution of each alternation in each variety is shown in Figure 3.1 below.



Figure 3.1 Frequency distribution of 进行 *jinxing* and 加以 *jiayi* in two varieties

The figure shows the distribution of alternation choices in Mainland and Taiwan. For 进行 *jinxing*, Mainland Mandarin only has three alternation choices while Taiwan has at least five constructions. Besides this, their preference for each alternation is also different: for the light verb 进行 *jinxing*, Taiwan prefers both type 1 and 2, followed by type 5; While Mainland has a dominant preference for type 1. For the light verb 加以 *jiayi*, Taiwan prefers type 1 but also has some usages of type 5, while Mainland one only has type 1 usage.

Although Figure 3.1 can tell us the distributional differences between Mainland and Taiwan, it cannot tell whether these differences are significant or not. Therefore chi-square test is needed to investigate whether the distribution differences have statistical significance or not. The results (Standardized Pearson residuals  $e_{ij}$  (Agresti and Kateri, 2011)) are also transformed into signs, as shown in Table 3.8.

As can be seen from the result (shown below in Table 3.23), both 加以 jiayi and 进行 jinxing in Mainland Mandarin prefer the first alternation type: the theme is introduced by prepositional structure fronting the light verb, such 对可行性进行研究 dui kexingxing jinxing as yanjiu for\_practicability\_proceed\_research 'to conduct research on practicability'/对 优秀学生加以甄选 dui *vouxiuxuesheng jiavi* zhenxuan for excellent student inflict select 'to select excellent student'(P Value< 0.001). In Taiwan Mandarin, 加以 jiayi shows its significant preference for the type 5 (加以甄选优秀学生 jiayi zhenxuan youxiuxuesheng inflict\_select\_excellent\_student 'to select excellent student'). Taiwan 进行 jinxing is favored by type 2 (进行可行性研究 jinxing kexingxing yanjiu proceed\_practicability\_research 'to conduct research on practicability') and type 4 (进行对可行性研究 jinxing dui kexingxing yanjiu proceed\_for\_practicability\_research 'to conduct research on practicability') and 5 (进行研究可行性 *jinxing vanjiu kexingxing* proceed\_research\_practiciability 'to conduct research on practicability').

	1	2	3	4	5
Taiwan 进行	-	+	0	+	+
Mainland	+	-	0	-	-
进行					
Taiwan 加以	-	None	None	None	+
Mainland	+	None	None	None	-
加以					

Table 3.23 Result of Chi-square test for 进行 *jinxing* and 加以 *jiayi* in two varieties

The tendency is clear that for both 加以 *jiayi* and 进行 *jinxing*, the theme in Taiwan Mandarin prefers to be appeared after the light verb (either as a modifier or follow the complement), while the theme in Mainland Mandarin significantly prefers to be appeared before the light verb. In other words, variational differences have been shown in preference of word order.

## 3.7.4 The Second Research Question

As has been observed that there are significant differences in preference of word order between Mainland and Taiwan, the second question needed to be addressed is: which factors can influence their syntactic choices, and are there any variation differences?

## **3.7.4.1** Annotation Scheme

There are some potential factors which may influence the syntactic choices between different alternations. First is the complexity of the theme, whether the theme is complex or not may influence its position in the construction. The complexity of the theme is measured in two ways: the first is whether it has attributes or not (e.g., 技术 *jishu* 'technology' in 对技术进行改造 *dui jishu jinxing gaizao* for\_technology\_proceed\_improve 'to improve the technology' is a simple bare theme with no modifier, but 落后的工厂生产技术 *luohou de gongchang shengchan jishu* backward\_DE\_factory\_production\_technology 'backward factory production technology' in 对落后的工厂生产技术进行改造

dui luohou gongchangshengchanjishu de jinxing gaizao for backward DE factory production technology proceed improve 'to improve the backward factory production technology' should be considered as a complex theme because of the presence of the attribute). Another way to measure the complexity is to calculate the length of the theme (using the number of characters). For example, 落后的工厂生产技术 luohou de gongchangshengchanjishu backward DE factory production technology 'backward factory production technology'contains 9 characters while 技术 jishu 'technology' only contains 2 characters, the former one is much more complex than the latter one, in terms of the length.

Besides, the complexity of the taken complement may also influence syntactic choices: if the complement is already a complex one with modifier (e.g., 进行长期深入的学习 jinxing changgishenru de xuexi proceed longterm\_in-depth DE study'proceed with a long-term in-depth study'), the theme may prefer the position before the light verb (对这份政府报告进行长期深入) 的学习 dui zhe fen zhengfubaogao jinxing changqishenru de xuexi for\_this\_CL\_government\_report\_proceed\_long-term\_in-depth DE study 'to proceed with a long-term and in-depth study on this government report') instead of after it (进行对这份政府报告的长期深入的学习 jinxing dui zhe fen zhengfubaogaode changgishenru de xuexi proceed for this CL government report DE long-term in-depth DE study 'to proceed with a long-term and in-depth study on this government report'). For the latter construction, the information after light verb is too complicated, and this may make the whole construction very unnatural.

In addition to the complexity of both theme and complement, the features such as the animacy, pronominality of the theme, as well as whether the theme is proper noun or not also have been included (as explained in Table 3.24). To have sufficient data for simultaneous comparisons, most of these variables were simplified to binary values: pronominality was simplified to pronoun and non-pronoun, animacy to animate (+animate) and inanimate (-animate), proper noun to yes and no (Bresnan, 2010).

In addition, the features that are efficient in distinguishing light verb variations are also included in this study, to see if these features can influence the syntactic choices (shown in Table 3.6). I eliminated some variables which have a very sparse value (e.g., whether the complement is spontaneous event, whether the light verb occurs with other light verb or not).

All the 373 tokens (=111+65+119+78) in both Mainland and Taiwan Mandarin are annotated manually, with the 14 features explained above. The annotator is a trained expert in Chinese linguistics, and the reliability and validity of the annotation schema have been tested and shown in Chapter 2.

Feature	Examples of Yes	Examples of No
complex	Yes	No
ity of	对落后的工厂生产技术进行改造 dui	对技术进行改造 dui
the	luohou de gongchang shengchanjishu	jishu jinxing
theme	jinxing	gaizaofor_technology
	gaizaofor_backward_DE_factory_producti	_proceed_improve'to
	on_technology_proceed_improve 'to	improve the
	improve the backward factory production	technology'
	technology'	
complex	Yes	No

ity of the taken	进行长期深入的学习 <i>jinxing</i> <i>changqishenru de xuexi</i> proceed_long- term in-depth DE study <sup>c</sup> proceed with a	进行学习 <i>jinxing</i> <i>xuexi</i> proceed_study'to study'
comple	long-term in-depth study'	
ment		
Animac	Yes	No
y of	对陈庆辉进行调查 dui chenqinghui	对可行性加以研究
theme	jinxing	dui
	diaochafor_Chenqinghui_proceed_investig	<i>kexingxingjiayiyanjiu</i> f
	ate 'to investigate Chen Qinghui'	or_practicability_inflic
		t_research'to conduct
		research on
		practicability'
pronomi	Yes	No
nality of	为此进行协商 wei ci jinxing	对报告进行汇总 dui
the	<i>xieshang</i> for_this_proceed_negotiate 'to	baogaojinxing
theme	negotiate about this'	huizongfor_
		report_proceed_summ
		arize 'to summarize
		the report'
Whether	Yes	No
the	对陈庆辉进行调查 dui chenqinghui	对男性进行检测 dui
theme is	jinxing	nanxing jinxing
proper	diaochafor_Chenqinghui_proceed_investig	jiancefor_male_procee
noun	ate 'to investigate Chen Qinghui'	d_detect 'to detect
		male'

Table 3.24 New annotation features for alternative study

With the fully annotated data, Chi-square test is conducted to examine the significance of the co-occurrence of each feature with each alternation. The results are also transformed into signs as the Chi-square test did above (also shown in Table 3.8).

## 3.7.4.2 Data analyses and discussion

As 加以 *jiayi* in Mainland only has one syntactic choice to introduce the theme (1<sup>st</sup> type: 对可行性加以研究 *dui kexingxing jiayi yanjiu* 

for\_practicability\_inflict\_research 'to conduct research on practicability'), no test is needed in this step. Hence three Chi-square tests have been conducted for Mainland 进行 *jinxing*, Taiwan 进行 *jinxing* and Taiwan 加以 *jiayi*, respectively. The results will be illustrated in detail as shown below.

## 3.7.4.2.1 For Light Verb 进行 jinxing in Mainland

As can be seen from Table 3.25, the syntactic choices of Mainland 进行 *jinxing* is not correlated with the semantic type of the complement (i.e. whether the complement is durative/formal/psychological/interaction/accomplishment event or not does not influence the alternation choice).

In contrast, other factors show significant correlation with alternation choices to different extent. Complex theme (e.g., 政府本期的工作报告 *zhengfubenqi de gongzuobaogao* government\_current\_DE\_work\_report 'current working report of government'), animate theme (e.g., 妇女 *funü* 'women'), proper noun theme (e.g., 美国 *meiguo* 'America') and pronoun theme (他/此 *ta/ci* 'he/this') significantly prefer to be introduced by the 1<sup>st</sup> type (e.g., 对政府本期的工作报告进行汇总 *dui zhengfu benqi de gongzuobaogao jinxing huizong* for\_government\_current\_DE work\_report\_proceed\_summarize 'to summarize the current working report of government'/对妇女进行检测 *dui funü jinxing jiance* for\_women\_proceed\_detect 'to detect women'/与美国进行 对话 *yu meiguo jinxing duihua* with\_America\_proceed\_dialogue 'have a

dialogue with America'/ 为此进行协商 wei ci jinxing xieshang for this proceed negotiate 'to negotiate about this') while these features are significantly disfavored by the 2<sup>nd</sup> type (?进行对此的协商 *jinxing dui ci de xieshang* proceed for this DE negotiate 'proceed with negotiation about this'). In addition, the Part of Speech of complement may also have influence on the syntactic choices. The nominal complement significantly prefers the 2<sup>nd</sup> type (e.g., 进行人口普查 *jinxing renkoupucha* proceed census 'carry on census') while the verbal complement shows significant preference for  $1^{st}$  type (e.g.,  $\forall f$ 妇女进行检测 dui funü jinxing jiance for women proceed detect 'to detect women'). The 3<sup>rd</sup> type shows the preference of having aspectual marker le (e.g., 进行了联合开发矿产资源的合作 jinxing le lianhe kaifa kuangchanziyuan de hezuo proceed LE joint develop mineral resource DE cooperate 'cooperate in joint developing mineral resources'). Although Chi-square test shows the 4<sup>th</sup> type significantly prefers the VP and VO complement, as there is only 1 example for each feature, it is not considered to be significant.

	1	2	3	4	5
Attribute of	+	-	0	0	None
theme					detected
Attribute of	+	-	0	0	None
complement					detected
Animacy of	+	-	0	0	None
theme					detected
Proper noun	+	-	0	0	None
					detected
Pronominality of	+	-	0	0	None
theme					detected
Polarity of	0	0	0	0	None
context					detected
POS of	+V	+N	0	+VP	None

complement				(only one	detected
				example)	
VO complement	0	0	0	+ (only	None
				one	detected
				example)	
Aspectual	0	No	Le	0	None
marker					detected
Durative event	0	0	0	0	None
					detected
Formal event	0	0	0	0	None
					detected
Psychological	0	0	0	0	None
event					detected
Interaction event	0	0	0	0	None
					detected
Accomplishment	0	0	0	0	None
event					detected

Table 3.25 Chi-square test for syntactic choice of Mainland 进行 jinxing

# 3.7.4.2.2 For Light Verb 加以 jiayi inTaiwan

As shown in Table 3.26, different from the result of Mainland 进行 *jinxing*, only one factor has been found to influence the syntactic choice of Taiwan 加以 *jiayi*: complexity of theme. In particular, complex theme significantly prefers the 1<sup>st</sup> type (e.g., 对人员的身份不加以查验 *dui renyuan de shenfen bu jiayi chayan* for\_personnel\_DE\_identity\_BU\_inflict\_check 'not check the identity of personnel'), while theme without attributes may prefer the 5<sup>th</sup> type (e.g., 加以保 护 古 籍 *jiayi baohu guji* inflict\_protect\_ancient\_book 'to protect ancient books').

	1 <sup>st</sup> type	5 <sup>th</sup> type
Attribute of theme	+	-

# Table 3.26 Chi-square test for syntactic choices of 加以 jiayi
#### 3.7.4.2.3 For Light Verb 进行 jinxing in Taiwan

The result of Taiwan 进行 *jinxing* is very similar with that of Mainland 进行 *jinxing*. In Taiwan Mandarin, as shown below in Table 3.28, the alternation preference of 进行 *jinxing* (five alternations in total) is significantly associated with the feature 'complexity of theme' 'animacy of theme' 'proper noun theme' 'Part of Speech of complement' and 'durative event'. In particular, the same as Mainland 进行 *jinxing*, the complex theme with attribute also prefers 1<sup>st</sup>alternation (e.g., 对复杂景物和简单景物进行跟踪 *dui fuza jingwu he* 

jiandan jingwu jinxing genzong for complex scene and simple scene proceed trace 'to trace the complex and simple scene') in Taiwan while it disfavors the 2<sup>nd</sup> type (e.g., 进行业绩考 核 *jinxing yeji kaohe* proceed\_performance\_appraisal 'to conduct performance appraisal'). The animate and proper noun theme also prefers 1<sup>st</sup> type (e.g., 对陈 庆 讲 行 调 杳 辉 dui chenqinghui jinxing diaocha for Chenqinghui proceed investigate 'to investigate Chen Qinghui'). The nominal complement also significantly prefers the 2<sup>nd</sup> type (e.g., 进行游泳比赛 jinxing youyong bisai proceed\_swimming competition 'to have a swimming competition') while the verbal one prefers the 1<sup>st</sup> (e.g., 对最近的情况进行检讨 dui zuijin de qingkuang jinxing jiantao for\_recent\_DE\_situation\_proceed\_selfcriticism 'to make a criticism on recent situation').

	1	2	3	4	5
Attribute of	+	-	0	0	0
theme					
Attribute of	0	0	0	0	0
complement					
Animate of	+	-	0	0	0
theme					
Proper noun	+	-	0	0	0
theme					
Pronominality	0	0	0	0	0
theme					
Polarity of	0	0	0	0	0
context					
POS of	+V	+N	0	0	+VP
complement					
VO complement	0	0	0	0	0
Aspectual	0	0	0	0	0
marker					
Durative event	-	+	0	0	0
Formal event	0	0	0	0	0
Psychological	0	0	0	0	0
event					
Interaction event	0	0	0	0	0
Accomplishment	0	0	0	0	0

Table 3.27 Result of Chi-square test for Taiwan 进行 jinxing

Compared with Chi-square test results in both Mainland and Taiwan, the alternation preference of 进行 *jinxing* in both Taiwan and Mainland (five alternations in total) is significantly associated with the feature 'complexity of theme' 'animacy of theme' 'proper noun theme' 'Part of Speech of complement'. Complex theme, animate theme and pronoun theme prefer to be occurred in Type 1 alternation. Mainland and Taiwan light verbs also show differences in the factors that influence the alternation choices. On one hand, durative complement is significantly favored by the 2<sup>nd</sup> type (e.g., 进行文化交流 *jinxing wenhua jiaoliu* proceed\_cultural\_exchange 'conduct cultural exchange') in Taiwan while this factor does not influence syntactic choice in Mainland. On

the other hand, in Taiwan, the complexity of complement does not affect the alternation choice.

#### 3.7.5 Correlation between Alternation Choice and Length of Theme

As mentioned in Section 3.8.4.1, the length of theme (numbers of characters) has also been calculated to measure its complexity. Based on the data, it is necessary for us to investigate the correlation between the length of the theme and the syntactic choices. A Kruskal-Wallis test was conducted to determine if there were differences in length of theme between different alternation groups. The Kruskal-Wallis test, as the non-parametric One-Way ANOVA, is used here since it does not require for the normality of data.

For Mainland 进行 *jinxing*, the test was conducted to examine if there were differences between the four alternation groups: 1<sup>st</sup> type (n=79), 2<sup>nd</sup> type (n=27), 3<sup>rd</sup> type (n=4), 4<sup>th</sup> type (n=1). Distributions of length were similar for all groups, as assessed by visual inspection of a boxplot. As shown in Table 3.28, length of theme were statistically significantly different between different alternation groups,  $\chi^2$ =11.406, p=0.01. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values were presented. This post hoc analysis revealed statistically significant difference in median length between the 1<sup>st</sup> (median=7) and 2<sup>nd</sup> alternation (median=2) (P=0.015), but not between any other group combination. To be specific, the result revealed that the length of theme for 1<sup>st</sup> alternation is significantly higher than the one of 2<sup>nd</sup> alternation.

Column	Result				
Sig.	0.01				
Decision	Reject the nu	ll hypothesis			
Null	The distribution of character is	the same across ca	tegories of		
Hypothesis	sis alternation				
Test	Independent-Samples	Kruskal-Wallis Te	st		
Asymptotic significances are displayed. The significances are displayed and the significances are displayed at the significance at the significance are displayed at the significance at the significance at the significance at the significance are displayed at the significance at the sig		The significance le	vel is .05.		
Total N		111			
Test Statistic		11.406			
Df		3			
Asymptotic Sig. (2-sided test)		.01			

Table 3.28 Result of Kruskal Wallis test for Mainland 进行 jinxing

A similar result is presented for Taiwan 进行 *jinxing*. Kruskal-Wallis test was conducted between different alternation groups:  $1^{st}$  alternation (n=45),  $2^{nd}$  alternation (n=41), alternation  $3^{rd}$  (n=4),  $4^{th}$  alternation (n=11) and  $5^{th}$  alternation (n=18). The result in Table 3.30 shows that length of theme were statistically significantly different between alternation groups:  $\chi^2$ =26.971, p<0.001. The same as Mainland 进行 *jinxing*, the Post hoc analysis of Taiwan 进行 *jinxing* (Table 3.30) also revealed that the length median of  $1^{st}$  group (median=8) is significantly higher than the one of  $2^{nd}$  group (median=2), but no significance were found between any other group combination.

Column	Result
Sig.	2.01483118664747E-5
Decision	Reject the null hypothesis
Null	The distribution of character is the same across categories of
Hypothesis	alternation
Test	Independent-Samples Kruskal-Wallis Test

Total N	119
Test Statistic	26.971
df	4
Asymptotic Sig.	.000
(2-sided test)	

Table 3.29 Result of Kruskal Wallis test for Taiwan 进行 jinxing

Sample 1-	Test	Std. Error	Std.	Sig.	Adj.
Sample 2	Statistic		Statistic	_	Sig.
Туре 2 –	-25.112	9.628	-2.608	0.009	0.091
type 5					
Type 2-	-26.027	11.562	-2.251	.024	.244
type 4					
Type 2-	-29.015	17.836	-1.627	.104	1.000
type 3					
Туре 2 –	37.579	7.351	5.112	.000	.000
type 1					
Туре 5 –	.914	13.031	.070	.944	1.000
type 4					
Туре 5 –	3.903	18.822	.207	.836	1.000
type 3					
Туре 5 –	12.467	9.496	1.313	.189	1.000
type 1					
Туре 4 –	2.989	19.881	.150	.881	1.000
type 3					
Type 4 –	11.553	11.453	1.009	.313	1.000
type 1					

Table 3.30 The Post hoc comparisons of Taiwan 进行 jinxing

For Taiwan  $m \bowtie jiayi$ , there are 2 alternation types: 1<sup>st</sup> type (n=58) and 5<sup>th</sup> type (n=21). The result of statistical test in Table 3.31 revealed that the length median of the 1<sup>st</sup> type (median=7.5) is statistically higher than the one of 5<sup>th</sup> type (median=4.5) ( $\chi^2$ =5.918, p=0.015).

Column	Result
Sig.	0.0149892750977441
Decision	Reject the null hypothesis
Null	The distribution of character is the same across categories of
Hypothesis	alternation
Test	Independent-Samples Kruskal-Wallis Test

Total N	78
Test Statistic	5.918
df	1
Asymptotic Sig.	.015
(2-sided test)	

Table 3.31 Result of Kruskal Wallis test for Taiwan 加以 jiayi

As can be seen from these three tests, length of theme is a very important factor in terms of choosing different syntactic alternations, for both Taiwan and Mainland light verbs. Based on the result of both Chi-square test and Kruskal-Wallis test, the tendency is clear that complex theme (theme with attribute and relatively longer) tend to appear before the light verb (the 1<sup>st</sup> alternation) while the simple theme is more likely to select the after-light verb position (2<sup>nd</sup> or 5<sup>th</sup>alternation). Therefore, the complexity of theme is a crucial factor which has very important influence on syntactic choices (i.e. the length of the theme is a very important factor for both Taiwan and Mainland in choosing different alternations).

#### **3.7.6** Commonality and Difference

To make a summary, for the results presented in above section, compared to Mainland Mandarin, Taiwan Mandarin has more alternations in terms of introducing the theme of event. The two specific alternations in Taiwan: 4<sup>th</sup> (e.g., 讲 行 性 研 究 行 对 可 jinxing dui kexingxing yanjiu proceed for practicability research 'to conduct research on practicability') and  $5^{\text{th}}$ 性 (e.g., 讲 行 研 究 可 行 jinxing vanjiu kexingxing proceed\_research\_practiciability 'to conduct research on practicability') actually indicate the different syntactic types of light verb complements in Mainland and Taiwan. 研究 vanjiu 'research' in these two constructions should be considered as a verb while in Mainland 研究 yanjiu 'research' can only be

considered as a deverbal noun (对可行性进行研究 *dui kexingxing jinxing yanjiu* for\_practicability\_proceed\_research 'to conduct research on practicability').

In this study, two research questions have been answered: For the first research question, distributional differences in alternative choices have been detected between Mainland and Taiwan. The general tendency is clear that the theme in Taiwan Mandarin prefers to be appeared after the light verb (either between the light verb and the complement or follow the complement) while the theme in Mainland Mandarin significantly prefers to be appeared before the light verb.

For the factors that may influence the syntactic choices, although Mainland and Taiwan have some differences (animate, proper noun and pronoun theme prefer type 1 in Mainland while Taiwan shows the opposite tendency), they share the commonality that for both 进行 *jinxing* and 加以 *jiayi* in both varieties, the complexity of theme has significantly correlation with their syntactic choices. In particular, for both Mainland and Taiwan light verbs, if the complement is long and complex (with the modifier), then the theme is usually preferred to occur before the light verb (对这个方案的可行性进行研 究 dui zhege fangan de kexingxing jinxing yanjiu for\_this\_plan\_DE\_practicability\_proceed\_research 'to conduct research on the practicability of this plan'); otherwise, it prefers the post-verbal position (e.g., 进行可行性研究 *jinxing kexingxing yanjiu* proceed\_practicability\_research 'to conduct research on practicability').

# 3.8 Summary of Chapter 3

In this chapter, grammatical variations in Light Verb Constructions between Mainland and Taiwan Mandarin have been carefully examined, with the assistance of large-scale comparable corpus as well as statistical models. Investigating language variations from a probabilistic perspective is proved to be very effective and reliable. Comprehensive descriptive accounts of light verb variations have been provided in Section 3.4. It has been shown that most of the annotated features are very useful in differentiating light verb constructions from different language varieties. Moreover, based on the statistical result, the variation differences in syntactic type and degree of transitivity have also been studied in detail, in Section 3.5 and Section 3.6, respectively. It has also been shown that light verb itself differs in its degree of transitivity between these two varieties: Taiwan light verb tends to have higher degree of transitivity, in terms of transitivity frequency as well as semantic/syntactic properties of the taken complements. Besides variations in transitivity, differences in word order are also included. By examining the differences in alternative choices in introducing the theme of event (e.g., 对可行性进行研究 dui kexingxing jinxing) for practicability proceed research yanjiu 'to conduct research on practicability'. 进行可行性研究 jinxing kexingxing yanjiu proceed practicability research 'to conduct research on practicability', 进行 可行性的研究 jinxing (dui) kexingxing ( 対 ) de yanjiu proceed (for) practicability DE research 'to conduct research on

practicability', 进行对可行性研究 *jinxing dui kexingxing yanjiu* proceed\_for\_practicability\_research 'to conduct research on practicability' and 进行研究可行性 *jinxing yanjiu kexingxing* proceed\_research\_practiciability 'to conduct research on practicability'), I can show that variations are observed not only in differences in feature/characteristic, but also in alternative choices. With respect to the differences in introducing the theme, the general tendency is clear that the theme in Taiwan Mandarin prefers to be appeared after the light verb (either between the light verb and the complement or follow the complement), while the theme in Mainland Mandarin significantly prefers to be appeared before the light verb.

# Chapter 4 Grammatical Variations in VO compound 4.1 Introduction

The Verb-Object compound in Mandarin Chinese (here after 'VO compound'), name indicates, is composed of two constituents having as its syntactic/grammatical relation of a verb and its direct object (Li and Thompson, 1981). The vast majority of VO compounds are verbs, while only a very small number of VO compounds function as nouns (e.g., 当局 dangju 'authority', 董 事 dongshi 'executive', 枕头 zhentou 'pillow') and adverbs (e.g., 到底 daodi 'on earth', 转眼 zhuanyan 'instantly') (e.g., Li and Thompson, 1981; Li andWu, 2017). It is pointed out by some earlier studies that, one important characteristic shared by the vast majority of VO compounds is, they cannot take a direct object (e.g., Li and Thompson, 1981).

However, in fact, in recent years, more and more VO compounds have been observed to be transferred from intransitive to transitive verbs, that is, these VO compounds can take another external objects and yield the VO<sub>1</sub>+O<sub>2</sub> construction. Examples like 投资房地产 *touzi fnagdichan* invest\_real-estate 'invest in real estate', 签约唱片公司 *qianyue changpian gongsi* sign\_contract\_record\_company 'sign contract with record company' frequently appear in actual usage. The interesting thing for this phenomenon is that the verbal morpheme V in VO<sub>1</sub> has already taken an object (like 投*tou* 'invest' and 签*qian* 'sign' have already taken an object: 资*zi* 'caption' and 约 *yue* 'contract'). According to the theory in formal syntax (e.g., Li, 1985), VO typically do not take another object because the case/theta assignment by the V is fulfilled by  $O_1$ , hence the lexical unit no longer has argument taking abilities. This is also consistent with Li and Thompson (1981), in which they claim that one important characteristic shared by the vast majority of VO compound is, they cannot take a direct object. However, it has been observed that there are more and more VOs which show the tendency of being used transitively (e.g., Wang, 1997; Diao, 1998). The increasing number of transitive VO compound poses a challenge to the theoretical accounts, which has also attracted the interests of many linguists. Actually, the transitivization of VO compounds has become one of the more prominent ongoing linguistic changes in Modern Mandarin. The construction is named  $VO_1+O_2$  in the sense that it is important to differ between  $O_1$  (the internal object taken by V) and  $O_2$  (the external object taken by VO), in order to examine the transition mechanism of VO compound from intransitive to transitive.

The transitivization of VO compounds becomes much more complex if the variation differences between different varieties of Mandarin Chinese are taken into consideration. It has been observed that different Mandarin variations may differ in the transitivity of VO compounds (e.g., Wang, 1997; Diao, 1998). However, very few studies have been conducted to provide comprehensive and definitive account of syntactic variations in VO<sub>1</sub>+O<sub>2</sub> between different Mandarin varieties. The lack of systematic studies to a large extent is due to the gradient/probabilistic nature of grammatical variation. In other words, syntactic variations on  $VO_1+O_2$  construction tend to be differences in preference/tendency instead of 'grammatical' versus 'ungrammatical' distinction.

To better illustrate the tendency variation differences in  $VO_1+O_2$ construction between Mainland and Taiwan Mandarin, in this chapter, a comparable corpus-based statistical approach is still utilized. Different linguistic issues including the transitivity difference between Mainland and Taiwan, variation differences in degree of lexicalization and transition mechanism will be comprehensively investigated in this chapter.

# **4.2 Literature Review on the Transitivization of VO Compounds**

**4.2.1 Literature Review on the Study of VO**<sub>1</sub>+**O**<sub>2</sub> Construction in Mandarin Previous studies on VO<sub>1</sub>+O<sub>2</sub> construction were conducted from various perspectives. From the theoretical linguistic perspective, the grammatical constituent of VO<sub>1</sub>+O<sub>2</sub> construction (i.e. the syntactic type and the composition of VO<sub>1</sub>) has been discussed by different scholars (e.g., Xing, 1997; Wang, 1997; Rao, 1984). About the syntactic type of VO<sub>1</sub>, there are number of debates on which syntactic type does VO<sub>1</sub> belong to: whether it is a word (Xing, 1997; Wang, 1997; Liu, 1998a; Gao, 1998; Yang, 2001); a phrase (Chen, 1995; Zhang, 1999) or can be considered as both (Li, 1990; Rao, 1984; Diao, 1998).

In terms of the composition of  $VO_1$ , agreement has been reached that a vast majority of  $VO_1$  are composed by a verbal morpheme plus a nominal morpheme (e.g., Ling, 1999; Qian, 2011), but sometimes  $O_1$  can also be a

verbal morpheme. Li and Wu (2017) examines 85 transitive VO compounds and finds out that over 90% of  $O_1$  are nominal morpheme (764/845) while verbal morphemes are also possible (81/845).

Beside this, the semantic relations between different components (between V and  $O_1$ ,  $O_1$  and  $O_2$  as well as VO<sub>1</sub> and  $O_2$ ) have also been carefully examined (e.g., Liu and Li, 1998; Liu, 1993).

The semantic relations between  $VO_1$  and  $O_2$  (the semantic role of  $O_2$ ) has attracted the attention of Chinese linguistics and has become the focus of VO<sub>1</sub>+O<sub>2</sub> study. Although the categorization varies a lot among different studies, agreement has been reached that most of  $O_2$  belong to target and location; It may also include other roles like goal, result, location, causative, benefactive, quantity, patient, time and others (Rao, 1984; Liu, 1993; Gao, 1998; Luo, 1998; Liu and Li, 1998; Shi, 1999; Yang, 2001; Qian, 2011; Zhu, Guo and Chen, 2002). The semantic relation between V and  $O_1$  has also been discussed by previous scholars. O<sub>1</sub> in most cases is **patient** (负责 *fuze* 'be responsible for'/牵手 gianshou 'join hands'/提议 tivi 'propose'/扬言 yangyan 'threaten'/让利 rangli 'surrender part of the profits'/解码 jiema 'decode'/应聘 yingpin 'accept an offer of emplyment') and agent (放心 fangxin 'reassure'/垂涎 chuixian 'covet'/ 回首 huishou 'look back'/甘心 ganxin 'willing'/亮相 liangxing 'appear'/移师 yishi 'move troop to'/飘香 piaoxiang 'send off fragrance'), but sometimes other semantic roles of  $O_1$  are also found, for example, **location** (登陆 *denglu* 'land'/ 出口 chukou 'export'/出炉 chulu 'draw a charge'/出台 chutai 'policy

introduce'); **reason** (帮忙 *bangmang* 'do a favor'/备战 *beizhan* 'prepare for war') and **instrument** (过目 *guomu* 'look over'/援手 *yuanshou* 'help') (Zhu et al., 2002).

The semantic relation between  $O_1$  and  $O_2$  is often considered to be **co-referential** (e.g., 动员人民群众 *dongyuan renminqunzhong* mobilize\_masses 'to mobilize the masses') (Xu, 1988) while other studies also point that there are more relations. For example, the semantic relation between  $O_1$  and  $O_2$  could also be **benefactive** (e.g., 授权商家 *shouquan shangjia* authorize\_merchant 'authorize the merchant') or **dominance** (e.g., 执教公牛队 *zhijiao gongniudui* coach\_Bulls 'to coach the Chicago Bulls') (Liu, 1993), as well as **possessive** (e.g., 牵手女嘉宾 *qianshou nijiabin* hold\_hand\_female guest 'hold hands with female guest') and **location** (e.g., 落 户 上 海 *luohu shanghai* settle down\_Shanghai 'settle down in Shanghai') (Zhu et al., 2002).

Another interesting question which has attracted the interest of scholars is the transition requirements (i.e. what kind of VO can be transferred from intransitive to transitive). This has also been the focus of study (e.g., Luo, 1998). Agreement has been reached that the first requirement is the VO<sub>1</sub> must be bivalent at semantic level (e.g., 服务 *fuwu* 'serve', 拜师 *baishi* 'formally acknowledge one as master', 执教 *zhijiao* 'coach'). If VO<sub>1</sub> needs two arguments at semantic level, it is possible for it to become transitive. Moreover, the lexical status is also highly correlated with the transitivity of VO<sub>1</sub>. To be specific, VO as a word is more likely to become transitive compare to the VO phrase (Luo, 1998; Liu, 1998; Chen, 1995; Hu, 2003; Liu, 1998b). As the lexical status of a VO can be measure by whether it can be separated (e.g., Her, 1996; Huang, 1984), the transitivity of VO is highly related to its separation status. In other words, if more grammatical elements can be inserted into VO, the VO is less likely to take object. Besides these two basic requirements, Shan (2007) also proposes 11 requirements from both semantic and syntactic perspectives.

From the language development perspective, the formation mechanism and the motivation of transition have both been discussed (e.g., Zhang, 2010; Li and Wu, 2017).

The formation mechanism of  $VO_1+O_2$  has attracted the attention of many Chinese linguistics. Zhang (2010) attributes the formation of  $VO_1+O_2$  to the drop of the clitic-like preposition  $\pm$  following  $O_1$ , which endows the configuration  $VO_1$  with the transitive function, e.g., 驰誉于中外 *chiyu yu zhongwai* famous\_at\_home and abroad 'be famous at home and abroad' to 驰誉 中外 *chiyu zhongwai* 'be famous at home and abroad'. Yang and Zhou (2006) distinguish three types of  $VO_1+O_2$  construction (Li and Wu, 2017): (1) one type comes from the verb-complement structure with the preposition omitted, such as from 让位于年轻人 *rang wei yu nianqing ren* give\_place\_to\_young\_people 'give place to young people' to 让位年轻人 *rang wei nianqing ren* give\_place\_young\_people 'give place to young people'; (2) another type is transited from the adverbial-verb construction with the disappearance of preposition, like from 向房地产投资 xiang fangdichan touzi to\_realestate invest 'invest in real-estate' to 投资房地产 touzi fangdichan invest\_realestate 'invest in real-estate'; (3) the third type is formed based on a two-step condension and shift, i.e. verbs and heads of objects are condensed and then moved before modifiers, such as from 侵犯希望工程的权利 qinfan xiwanggongcheng de quanli infringe Project Hope DE right 'infringe the right of Project Hope' to 侵权希望工程 ginguan xiwanggongcheng infringe right Project Hope 'infringe the right of Project Hope', the second object usually serves as the modifier of the first one. Li and Wu (2017) claim that when  $O_1$  is a nominal morpheme, the  $VO_1+O_2$  construction goes through three evolutionary phases: in the first stage, V and O<sub>1</sub> are two independent constituents, and can combine to form a VO phrase; at the second stage, O<sub>1</sub> gets incorporated into verb. The transition from stage one to stage two is substantially a process of lexicalization, in the sense that the V and  $O_1$  are condensed into a disyllabic compound in the form VO<sub>1</sub>, which can be used independently to describe an event. During the third stage, O<sub>1</sub> as a bound morpheme gradually goes through a process of semantic bleaching. Accordingly, V's transitivity seems to 'resume work', which naturally makes the transitivity of the entire compound verb  $VO_1$  salient. When  $O_1$  is originally a verbal morpheme (e.g., 解聘 jiepin 'dismiss an employee', 免考 miankao 'exempt from examination'), it is very natural for  $VO_1$  to be transitive.

There are a variety of researches aiming at exploring the motivation behind the formation of the  $VO_1+O_2$  structure. Many factors that may influence 156

the formation (both linguistic internal and external) are proposed by previous researches. Linguistic external factors mainly include the influence of English language (Ding and Yuan, 1998), Hong Kong or Taiwan Mandarin (Wang, 1997; Diao, 1998), as well as social media (e.g., Yang, 2001). In terms of the linguistic internal factors, Yang (2001) points out two reasons: on one hand, the 'front-weight principle' in Chinese is responsible for the compatibility and hence the simplification of VO structure; on the other hand, in the VO structure, the morpheme meanings are compatible to the syntactic structure. Zhu et al. (2002) discuss that the formation of  $VO_1+O_2$  construction should be promoted by the basic mode 'action-result' in Modern Chinese (Li and Wu, 2017). As Modern Chinese usually emphasis on the result of the action, the  $O_2$  as the taken complement can be considered as a result. Other internal factors may include phonetic effect, semantic element, syntactic analogy, function element and functional competition (Yang and Zhou, 2006). Agreement has also been reached that the formation is also driven by the principle of economy (Yang, 2001; Hua, 1997). For example, 观光吉隆坡 guanguang jilongpo sightseeing Kuala Lumpur 'take a sightseeing tour in Kuala Lumpur' is apparently more economical than 到吉隆坡观光 dao jilunpo guanguang to Kuala Lumpur sightseeing 'take a sightseeing tour in Kuala Lumpur' (Li and Wu, 2017). Furthermore, from the psychological perspective, the  $VO_1+O_2$ construction needs less time to memorize (Yang, 2001).

In terms of pragmatic perspective, the pragmatic function/value of  $VO_1+O_2$  construction has also been examined by different studies (e.g., Zhu and

Sheng, 2008; Hua, 1997). The VO<sub>1</sub>+O<sub>2</sub> construction is often considered to have various pragmatics values. Firstly, the VO<sub>1</sub>+O<sub>2</sub> is more concise and informative (X. C. Wang and Y. F. Wang, 2006). Secondly, the O<sub>2</sub> which often refers to time/location/cause has been highlighted and emphasized. The third reason is that the VO<sub>1</sub>+O<sub>2</sub> construction is a very creative construction which meets the need of younger generation as well as social media (X. C. Wang and Y. F. Wang, 2006).

# 4.2.1.1 Research Gap of Studies on VO<sub>1</sub>+O<sub>2</sub> Construction

What should be pointed out is, almost all of the previous studies categorize VO in a dichotomy way (which ignores the gradient nature of grammar (Bresnan et al., 2007)): VO compound is considered to be either transitive or intransitive (e.g., Liu and Li, 1998). Her (1996) has made a tripartite division in terms of the transitivity: VO can be intransitive (e.g., 得意 *deyi* 'be prond'); semi-transitive (e.g., 在行 *zaihang* 'be good at') and transitive (e.g., 留意 *liuyi* 'pay attention to'). However, according to the corpus investigation, we can find that the VO compounds which are considered to be transitive actually are very different in their degree of transitivity. In that sense, I argue that the transitivity of VO is better to be viewed in a continuous way instead of a dichotomy way (which is also in accordance with the nature of gradient/probabilistic grammar). As illustrated below from 4-1 to 4-5:

4-1. 关心他人/留意他的表现 guanxin taren/liuyi ta de biaoxian concern\_someone/pay attention to\_3 SG\_DE\_performance 'be concerned with someone/pay attention to his performance'

- 4-2. 媲美澳洲的黄金海岸 *pimei aozhou de huangjian haian* rival with\_Australia\_DE \_Gold\_Coast 'rival with the Gold Coast in Australia'
- 4-3. 牵手女嘉宾 *qianshou n ijiabin* hold\_hand\_female guest 'hold hands with female guest'
- 4-4. 观光意大利 guanguang yidali sightseeing\_Italy 'take a sightseeing tour in Italy'

4-5. 出丑别人 chuchou bieren

make people feel foolish\_other people 'make other people feel foolish'

The transitivity decreases from construction 4-1 to sentence 4-5. Either  $\neq$ 心 guanxin 'concern' or 留意 liuvi 'pay attention to' can be used as transitive verb very frequently, constructions like 关心他人/留意他的表现 guanxin taren/liuvi de biaoxian concern someone/pay ta attention to 3 SG DE performance 'be concerned with someone/pay attention to his performance' are very common and can easily be detected in corpus. Compared to 关心 guanxin 'concern' and 留意 liuyi 'pay attention to', 媲美 pimei 'rival with' can also be used in a transitive way, but with a lower frequency. Native speakers of Mandarin Chinese may prefer the intransitive construction such as 与澳洲的黄金海岸媲美 vu aozhou de huangjin hai'an pimei with Australia DE Golden Coast rival with 'to rival with the Golden Coast in Australia' instead of the transitive construction (媲美澳洲的黄金海岸 pimei aozhou de huangjin haian rival with\_Australia\_DE\_Gold\_Coast 'rival with the

Gold Coast in Australia'). In construction 4-3, 牵手 qianshou 'hold hands' is recently found to be appeared in transitive constructions like 牵手女嘉宾 qianshou nijiabin hold\_hand\_female guest 'hold hands with female guest', but the context in which it can appear is very limited. In terms of 观光 guanguang 'sightseeing', it is not very natural if we say construction like 观光意大利 guanguang yidali sightseeing\_Italy 'take a sightseeing tour in Italy', but it is still comprehensible without ambiguity for native speakers. Nevertheless, it is not acceptable when the word like 出丑 chuchou 'make people feel foolish' is used transitively (e.g., \*出丑别人 chuchou bieren make people feel foolish\_other people 'make other people feel foolish').

# 4.2.2 Literature Review on $VO_1+O_2$ Variations between Mandarin Varieties

From language variation perspective, very few studies have been conducted to compare the usages of  $VO_1+O_2$  between different varieties of Mandarin Chinese. Some studies have observed that the transitivity of VO compound is likely to differ between different Mandarin Variations.

Wang (1997) lists about 50 VO compounds (e.g., 登陆 denglu 'land', 涉 嫌 shexian 'be suspected', 回笼 huilong 'withdrawal of currency', 加盟 jiameng 'join in', 入围 ruwei 'be on the finalist') which are commonly used as transitive verbs in Singapore, and also claims that the transitive usages in Mainland Mandarin to a large extent are due to the influence of Singapore Mandarin. Diao (1998) also points out that the reason behind the increasing popularity of  $VO_1+O_2$  is the influence of Taiwan Mandarin, and the transitivization of VO compound can be considered as 'revival' of ancient usages. Li (2014) has conducted a corpus-based study on the variation of  $VO_1+O_2$  construction between Mainland and Taiwan Mandarin. The corpus investigation shows that before 1980's, the transitivity frequencies of Taiwan VO compounds are obviously higher than that of Mainland counterparts, while the transition rates obviously slow down after 1980's for both Mainland and Taiwan VO compounds.

#### **4.2.2.1 Research Gap of Variations Studies**

Although some previous researches have observed the variation differences between Mandarin varieties, most of them illustrate the issue by introspective judgment. In addition, their investigations are still at the descriptive level. Hence, there lacks of systematic and comprehensive study for the VO compounds variations between different variants of Mandarin, especially based on large corpora.

Furthermore, a similar problem for previous variation studies is, the variation differences are often examined in a dichotomy way (i.e. for a certain VO, it can only be used transitively in Taiwan or Singapore Mandarin, not in Mainland Mandarin). However, in fact, with the more and more frequent communications between different regions and the influence of social media, the variation differences nowadays tend to be differences in tendency/preference instead of the dichotomy differences or absolute

differences (this is also consistent with the nature of gradient grammar (Bresnan et al, 2007)). For example, 中意 *zhongyi* 'like' has already become transitive in Mainland usage, but the frequency of being used in a transitive way is much lower than that of in Taiwan or in Hong Kong Mandarin. Moreover, Taiwan and Mainland Mandarin can also differ in the type of objects they are taking (e.g., 帮忙他 *bangmang ta* do a favor\_3 SG 'do him a favor' can only be observed in Taiwan Mandarin) or the context they can occur in (e.g., 曝光 *baoguang* 'expose' can only be used in negative context in Mainland but not necessary in Taiwan), while this kind of variations are difficult to be observed by traditional methodology.

#### 4.2.3 Summary

Based on the previous researches on both  $VO_1+O_2$  construction itself and grammatical variations between varieties, one important fact that has been neglected is the gradient nature of grammar (or grammatical variations). That is why the assistance of large-scale comparable corpus and statistical analyses are necessary for this study. In the next section of this chapter, there are mainly 4 studies are conducted in a comparable-corpus based statistical approach. For study 1, the transitivity variations between Mainland and Taiwan VO compounds will be examined. Both variation differences in transitivity degree and semantic/syntactic properties will be included. In terms of variation differences in transitivity degree, the comparison of transitivity frequency is

conducted between Mainland and Taiwan; In terms of variation differences in semantic/syntactic properties, mixed-effect logistic regression model and Chisquare test are both utilized to investigate the syntactic variation in semantic/syntactic features between these two varieties. Some theoretical implications based on the grammatical variation in both transitivity frequency and syntactic/semantic properties will also be discussed. For study 2, the correlation between transitivity and separation ability will also be explored, to see whether the grammatical variations of VO compounds are dependent on the degree of lexicalization of these compounds. Based on these two studies, the issue of language conservatism will also be discussed. For study 3, the relationship between transitivity and alternative competing pattern is also investigated to further study the transition mechanism of VO compounds, as well as to provide the basis for study of correlation between language change and language variation (based on the Lexical Diffusion Theory (e.g., Wang, 1969)). For study 4, the correlation between transitivity and word frequency is investigated to further discover the affecting reason for VO transitivization.

# **4.3 Study 1: Transitivity Variations in Mandarin VO compounds**

As has been mentioned Section 4.1, the grammatical variation differences in  $VO_1+O_2$  constructions between Mainland and Taiwan mainly lie in two perspectives: one is Mainland and Taiwan VO compounds differ in the frequency of being used transitively (e.g., 中意 *zhongyi* 'like' is more likely to

be used as a transitive verb in Taiwan than in Mainland); the other is  $VO_1+O_2$  constructions of the two varieties also differ in some semantic/syntactic properties (e.g., 帮忙 *bangmang* 'do a favor' in Taiwan can take NP as the object as in 帮忙他 *bangmang ta* do a favor\_3 SG 'do him a favor' while Mainland cannot).

In this study, to investigate variation differences from these two perspectives, different tests were included: the comparison of transitivity degree between Mainland and Taiwan Mandarin was first conducted (Z-test); based on the comparison, a mixed-effect logistic regression model and Chi-square test were also conducted to examine the grammatical variations in semantic/syntactic properties.

The VO wordlist I used in this study contains 171 VO compounds collected from previous researches (e.g., Liu and Li, 1998; Qian, 2011). As has already been presented in Chapter 2, the data was extracted from the annotated Chinese Gigaword corpus.

As has been reviewed in Section 4.2.1, there are numbers of debates on the syntactic status of VO (i.e. whether it is a word, phrase or both), to avoid the ambiguity, intersection was made between these 171 verbs and the wordlist of Gigaword and Sinica corpus (to ensure the VO is considered as a word in both Mainland and Taiwan Corpus). The procedure of intersection can also ensure the word is not unique in one variety (i.e. the word is used in both Taiwan and Mainland). After intersection, 16 Mainland unique words such as 叫板儿 *jiaobaner* 'challenge', 试水 *shishui* 'test the water' were excluded. A wordlist

of 155 VO compounds is generated for data extraction. For each VO compound in each variety, 1000 tokens were extracted from the Annotated Chinese Gigaword corpus. If the VO does not have enough 1000 tokens in corpus, I include all the tokens in the corpus. In total, more than 200000 tokens were extracted from Mainland and Taiwan corpora.

As argued in Section 4.2.1.1, the transitivity is better to be examined in a continuous way. Therefore in this study, relative frequency of transitivity usages is used to measure the transitivity degree of VO compound. The formula is shown below:

Transitivity frequency=transitive tokens/all the tokens

For example, transitivity frequency of 签约 *qianyue* 'sign contract' in Taiwan =number of transitive usages of 签约 *qianyue* 'sign contract' (13 tokens)/all the extracted tokens of 签约 *qianyue* 'sign contract' (1000 tokens) = 0.013.

The transitivity frequencies of all the 155 VO compounds in both Mainland and Taiwan varieties have been calculated based on the extracted corpus data.

#### 4.3.1 Grammatical variations in transitivity frequency of VO compounds

Based on the transitivity frequency, the first research question needed to be addressed is:

For each VO compound, does its Mainland and Taiwan usage differ in their transitivity frequencies (i.e. the frequency of being used transitively)? What kind of tendency do they appear (i.e. what is the pattern of the tendency)?

To test whether the two frequencies of Mainland and Taiwan have significant difference or not, Z-test is used. When P-value < 0.05, the difference between two varieties is considered to be significant.

The result of Z-test (shown in figure 4.1) shows that among all the 155 VO compounds, 33 words present non-significant differences between Mainland and Taiwan (e.g., 驰誉/调任/放眼 *chiyu/diaoren/fangyan* 'be famous/be transferred/look ahead'). 46 words are actually used as intransitive verbs in both varieties (i.e. the transitivity frequencies in both variants are smaller than 1‰, e.g., 参展/出丑/对话/联网 *canzhan/chuchou/duihua/lianwang* 'join an exhibition/make people feel foolish/have a dialogue with/networking'). Therefore, about half (76/155) VO compounds show statistically significant difference in transitivity frequency between these two variants.

Among these 76 words, 53 Taiwan VO compounds show significant higher transitivity frequency than their Mainland counterparts, while 23 VO compounds have significantly higher transitivity frequency in Mainland. As can be seen, Taiwan VO compounds tend to be more likely to have transitivity usages.



Figure 4.1 Result of Z-test

It should also be noted that, although all these 76 words show statistically significant frequency differences, their degrees of differences vary. In other words, P-value cannot tell us everything. For example, as shown in Table 4.1, although the variation differences of 过境 *guojing* 'transit' and 借道 *jiedao* 'channeled through' are both significant at P<0.01 level, the difference of 过境 *guojing* 'transit' between Taiwan and Mainland is much more obvious that that of 借道 *jiedao* 'channeled through'. Therefore another measurement was included in this study. It is proposed that the degree of differences can be measured by the likelihood ratio (=higher frequency/lower frequency), meaning that the higher the ratio is, the more obvious variation differences they have. For example, the transitivity frequency of 过境 *guojing* 'transit' in Taiwan is 0.341 (=341/1000) while in Mainland is 0.033 (=33/1000). The likelihood ratio

of Taiwan to Mainland is 10.33, meaning that Taiwan 过境 guojing 'transit' is about 10 times more likely to be used as a transitive verb than the Mainland counterpart. In contrast, 借道 *jiedao* 'channeled through' in Taiwan is only 1.26 (=0.87/0.69) times more likely to be used transitively than in Mainland. If the 76 words are further filtered by likelihood ratio, the transitivity tendency between Mainland and Taiwan becomes more obvious (as shown in Table 4.2).

	Transitivity in Taiwan	Transitivity in Mainland	P value for Z- test	Likelihood ratio
过境 guojing	0.341	0.033	<0.01	10.33
借道 <i>jiedao</i>	0.871383	0.689655	< 0.01	1.26

TW higher Mainland Total higher 16 14 2 Ratio>=5 Ratio>= $\overline{10}$ 8 0 8 7 0 Contrast 7

Table 4.1 Likelihood ratio of 过境 guojing and 借道 jiedao

Table 4.2 Result of likelihood ratio filter

difference

Table 4.2 presents that when likelihood ratio is larger than (or equal to) 5, 14 out of 16 words have significant higher transitivity in Taiwan. When the likelihood ratio is larger than (or equal to) 10, it is considered to have prominent significant difference in transitivity frequency between Mainland and Taiwan. Table 4.3 shows that for all the 8 words whose likelihood ratio is larger than 10, their frequencies in Taiwan are all prominently higher than their Mainland counterparts.

	TW	ML	Ratio
媲美 pimei	727/1021	28/1030	26.19
'rival with'			
中意 zhongyi	192/540	8/1337	59.42
'like'			
把关 baguan	182/743	11/1547	34.45
'guarantee'			
过境 guojing	341/1000	33/1000	10.33
'transit'			
献 计 xianji	6/84	2/1000	35.71
'offer advice'			
移 民 yimin	455/2000	1/1000	227.5
'immigrate'			
接壤 <i>jierang</i>	34/922	1/2269	83.67
'be contiguous			
to'			
撤军 chejun	23/1000	1/1000	23
'withdrawal			
troops'			

Table 4.3 Examples of words whose likelihood ratio >=10

More importantly, 7 words have been observed to have contrast transitivity difference between Mainland and Taiwan (i.e. transitivity usages only detected in one variety). To be particular, for all these 7 words (in Table 4.4), they are observed to have transitive usages only in Taiwan corpus while no transitive usages can be detected in Mainland corpus.

	TW	ML	Examples in TW
撤 兵 chebing	1/197	0/46	撤兵 <u>西岸地区</u>
'withdrawal			chebing xi'an diqu
troops'			withdrawal_troops_the West Bank
			'withdrawal troops from the West Bank'
垂爱 chuiai	5/37	0/2	老天特別垂爱 <b>钟岳岱</b>
'show tender care			Laotian tebie chuiai zhongyuedai
for'			God_special_tender care_Zhongyuedai
			'God shows special tender care for

			Zhongyuedai'
领航 linghang	76/810	0/169	有能力领航 <b>国家发展</b>
'pilot'			you nengli linghang guojia fazhan
			have_ability_pilot_country_develop
			'have ability to pilot country
			development'
观光 guanguang	4/1000	0/5224	观光 <u>意大利</u>
'sightseeing'			guanguang yidali
			sightseeing_Italy
			'take a sightseeing tour in Italy'
转行 zhuanhang	18/392	0/167	转行 <u>影视界</u>
'change career'			zhuanhang yingshiye
			change_career_filmdom
			'change to filmdom'
失望 shiwang	3/1000	0/1000	我很失望 <u>他未全力处理问题</u>
'be disappointed'			wo hen shiwang ta wei quanli chuli
			wenti
			1 SG_very_disappointed_3
			SG_not_with all effort_handle_problem
			'I am very disappointed that he did not
			handle this problem with all his effort.'
过目 guomu	22/317	0/65	过目 <u>所有的展品幻灯片</u>
'look over'			guomu suoyou de zhanpin huandingpian
			look over_all_DE_exhibit_PowerPoint
			'look over all PowerPoint slices of
			exhibits'

Table 4.4 Examples of contrast difference

With all the tendency shown above, it is fair to say that in terms of their transitivity frequency, Taiwan VO compounds are more likely to be used in a transitive way, especially when the variation difference is prominent significant.

Based on the result of Z-test and likelihood ratio, it can be summarized that generally speaking, referring to the transitivity frequency, Taiwan VO compounds are more likely to be used in a transitive way and the differences between Mainland and Taiwan are obvious.

#### 4.3.2 Grammatical Variation in Semantic/Syntactic Properties

In the above section, the result has shown that in terms of transitivity frequency, VO compounds in Taiwan have more probability to be used in a transitive way. Then the next question should be addressed is:

What kind of distributional differences in terms of semantic/syntactic features do they have between Mainland and Taiwan Mandarin?

In order to answer this question, the distributional differences in terms of syntactic and semantic properties should be examined. As the variation differences to a large extent tend to be differences in preference/tendency, two statistical methods (mixed-effect logistic regression model and Chi-square test) were carried out based on annotated data, to investigate the variation differences in preferences of collocation (follows the paradigm of Bresnan et al., 2007a).

#### 4.3.2.1 Annotation Schema

In this study, 10 potential features which may help to distinguish the differences between Mainland and Taiwan variant were selected from previous studies (e.g., Gao, 1998; Luo, 1998; some features from Bresnan et al., 2007a and DOLCE ontology). These features cover semantic (e.g., the semantic role of the taken object), syntactic (e.g., POS of the object; take aspectual marker or not), as well as discourse levels (e.g., the polarity of the context it occurs in; structural parallelism). The detailed annotated schema is shown in Table 4.5.

Feature	Values (example)
1.	Common NP (获赠一部手机 huozeng yibu shouji
syntactic	receive_one_CL_phone 'receive a phone')
type of	Event noun (投身大游行 toushen dayouxing plunge
the	into_big_parade 'join the parade')
object	Deverbal noun (插手电商的运营 chadian dianshang de yunying
taken by	intervene electronic business DE operation 'intervene the
a VO	operation of electronic business')
compou	Clause (投诉澳娱分紅不合理 tousu aoyu fenhong bu heli
na	complain_STDM_dividend_BU_fair 'complain that the dividends of
	STDM are not unfair')
	VP (插手经营家族企业 chashou jingying jiazu qiye
	intervene_manage_family_enterprise 'intervene the management of
	family enterprise')
	Quantity (驻军十六万 zhujun shiliuwan station_troop_160000
	'station 160000 troops')
	Time (执教五十年 <i>zhijiao wushinian</i> coach_fifty_years 'coach for
	fifty years')
2. proper	Yes (移民美国 yiminmeiguo immigrate_America 'immigrate to
noun or	America')
not:	No (移民发达国家 yimin fada guojia
_	immigrate_developed_country 'immigrate to developed country')
3.	Yes (帮忙他 bangmang ta do a favor_3 SG 'do him a favor')
Whether	No (帮忙家人 <i>bangmang jiaren</i> do a favor_family 'do family a
the VO	favor')
takes	
	V.a.(他们収持细关定按 强化善作 抑范损伤 配吞地调
4. Whether	Ies (他们至行北大甲核、蚀化贝住、观池保存、能长协响
the	iamen jianeni baguan snenne qiangnua zeren guijan caozuo peliao riatiao3
VO+O	Neuros PL insist guarantee check strengthen responsibility standardize
construct	operation support coordination 'They insist on guarantee the
ion has	checking, strengthen the responsibility, standardize the operation
parallel	and supporting the coordination')
structure	No (把关服务质量 <i>baguan fuwu zhiliang</i> guarantee service quality
or not	'guarantee the quality of service')
5.	Yes (海地警察头目要求出国避难太子港 haidi jingcha toumu
Whether	vaoaiu chuguo bi'nan taizigang
the	Haiti police head require go abroad seek shelter portal-prince
construct	'the head of Haiti police requires to go abroad and seek shelter at
ions are	portal-prince')
appeared	No (避难上海 bi'nan shanghai seek shelter Shanghai 'seek shelter
in	at Shanghai')
headline	

6. The	Polarity (各个部门热切希望他可以插手 gege bumen reqie xiwang
semantic	ta keyi chashou every_department_intent_hope_3
polarity	SG_can_intervene 'Every departments are intent on his intervenes')
of	Neutral (她很少不插手妹妹的事 ta henshao chashou meimei de
context	shi3 SG rarely intervene little sister DE thing 'She rarely
	intervene the business of her little sister.')
	Negative(各方势力经常插手流通环节,造成市场秩序十分混
	乱 gefang shili jingchang chashou liutong huanjie, zaocheng
	shichang zhixu shifen hunluan
	all_parties_force_often_intervene_lin_flow,
	cause_market_order_very_disruption 'The parities often intervene
	in the flow of links, resulting in the disruption of market order.')
7.Wheth	Yes (媲美这个世界一流名声大噪的新产品 pimei zhege shijie yiliu
er the	mingsheng dazao de xinchanpin rival with_this_world-
object of	class_famous_DE_new_product 'rival with this world-class famous
VO has	new product')
modifier	No (媲美中亞 pimei zhongya rival with_Central_Asia 'rival with
	Central Asia')
8.Wheth	Yes (ASP.le: 出版了一套系列丛书 chuban le yitao xilie
er a VO	congshipublish_LE_one_CL_series_books 'have published one
is	series of books'; ASP.zhe: 媲美着世界闻名的技术 pimei zhe shijie
affixed	wenming de jishu rival with_ZHE_world_famous_DE_technology
with	'rival with world-famous technology'; ASP.guo: 执教过中国队
zhe.le.gu	zhijiao guo zhongguodui coach_GUO_Chinese_National_team
0	'have coached Chinese Notional team')
	No (出版系列文集 <i>chuban xilie wenji</i>
	publish_series_collected_works 'publish serials')
9.Seman	Patient; theme; result; target; beneficient; goal; location
tic role	
of object	
10. The	Animate (中意这个徒弟 zhongyi zhege tudi like_this_apprentice
animacy	'like this apprentice')
of the	Inanimate (中意这道菜 zhongyi zhe dao cai like_this_CL_dish 'like
object	this dish')

Table 4.5 Annotation schema for VO compound

a. Feature 1: Syntactic type of the object taken by VO compound:

As VO compounds can take a variety of objects in terms of syntactic type: simple noun (e.g., 获赠一部手机 *huozeng yibu shouji* receive\_one\_CL\_phone 'receive a phone'), event noun (e.g., 投身大游行 *toushen dayouxing* plunge into\_big\_parade 'join the parade'), deverbal noun (e.g., 插手电商的运营 chadian dianshang de yunying intervene\_electronic\_business\_DE\_operation 'intervene the operation of electronic business'), clause (e.g., 投诉澳娱分紅不 合理 tousu aoyu fenhong bu heli complain\_STDM\_dividend\_BU\_fair 'complain that the dividends of STDM are not unfair'), VP (e.g., 插手经营家 族企业 chashou jingying jiazu qiye intervene\_manage\_family\_enterprise 'intervene the management of family enterprise'), quantity NP (e.g., 驻军十六 万 zhujun shiliuwan station\_troop\_160000 'station 160000 troops') and time NP (e.g., 执教五十年 zhijiao wushinian coach\_fifty\_years 'coach for fifty years') can all be appeared in the object position. Therefore it is natural to postulate that Mainland and Taiwan VO compounds may differ in the syntactic type of the object they can take.

b. Feature 2: Whether the taken object is proper noun:

It has been observed that the same VO compound can sometimes take a proper noun as the object (e.g., 移民美国 *yiminmeiguo* immigrate\_America 'immigrate to America'), while sometimes not (e.g., 移民发达国家 *yimin fada guojia* immigrate\_developed\_country 'immigrate to developed country'). Therefore this feature might be effective in distinguishing between the two varieties.

#### c. Feature 3: Proniminality of the taken object

Different nominal expression types (such as pronouns, proper nouns, common nouns) have been found to be co-occurred with VO compounds as the taken object. Following the study of Bresnan et al., (2007), pronominality in the annotation schema was also simplified to distinguish VOs followed by personal (e.g., 帮忙他 bangmang ta do a favor\_3 SG 'do him a favor'), demonstrative (e.g., 移民那里 yimin nali immigrate there 'immigrate there') and interrogative pronoun (e.g., 帮忙什么 bangmang shenme do a favor what 'help what'), from those followed by non-pronouns such as simple nouns (e.g., 帮忙家人 bangmang jiaren do a favor family 'do family a favor') and deverbal nouns 插手电商的运营 chadian dianshang (e.g., de yunying intervene electronic business DE operation 'intervene the operation of electronic business').

#### d. Feature 4: Structural parallelism:

Structure parallelism or persistence might be an important factor to influence the transitivity of VO compounds. For example, the VO is more likely to be used transitively if the constructions nearby are all transitive (e.g., 他们坚持把 关审核、强化责任、规范操作、配套协调 *tamen jianchi baguan shenhe, qianghua zeren, guifan caozuo, peitao xietiao,* 3 PL\_insist\_guarantee\_check, strengthen\_responsibility, standardize\_operation, support\_coordination 'They insist on guarantee the checking, strengthen the responsibility, standardize the operation and supporting the coordination'). In this model, this feature is measured by a variable defined as the presence of the same syntactic construction type (i.e.  $VO_1+O_2$  construction).

# e. Feature 5: Headline

The  $VO_1+O_2$  construction is frequently observed in newspaper headline (Yang, 2001), as the headline usually requires for the terseness of language. Therefore this feature is also included in the annotation schema.

#### f. Feature 6: Semantic polarity of the context

The semantic polarity of the context is also an important factor in differentiating between Mainland and Taiwan usages, as shown in the study of light verb variations. The same VO compounds can be observed to appear in different contexts: **positive** (e.g., 各个部门热切希望他可以插手 gege bumen regie xiwang kevi chashou every\_department\_intent\_hope\_3 ta SG can intervene 'Every departments are intent on his intervenes'), neutral (e.g.,她很少不插手妹妹的事 ta henshao chashou meimei de shi 3 SG rarely intervene little sister DE thing 'She rarely intervene the business of her little sister.') and negative (e.g.,各方势力经常插手流通环节,造成市 场秩序十分混乱 gefang shili jingchang chashou liutong huanjie, zaocheng shichang zhixu shifen hunluan all\_parties\_force\_often\_intervene\_lin\_flow, cause market order very disruption 'The parities often intervene in the flow of links, resulting in the disruption of market order.').
#### g. Feature 7: Complexity of the taken complement

The complexity of the taken complement is proved to be effective in distinguishing light verb variations in Chapter 3. This feature can be efficiently measured by the modifier of the taken complement. In other words, whether the taken complement has modifier (媲美这个世界一流名声大噪的科技新产品 *pimei zhege shijieyiliu mingsheng dazao de keji xinchanpin* rival with\_this\_world-class\_famous\_DE\_technology\_new\_product 'rival with this world-class famous new technology product') or not (e.g., 媲美中亞 *pimei zhongya* rival with\_Central\_Asia 'rival with Central Asia') might have significant difference between Mainland and Taiwan VO<sub>1</sub>+O<sub>2</sub> construction.

#### h. Feature 8: Aspectual marker

Since there is a prominent significant difference in the usages of aspectual marker between Taiwan and Mainland light verb constructions, therefore this feature is also included for the study VO compounds. Either of the aspectual marker 着 ZHE (e.g., 媲美着世界闻名的技术 *pimei zhe shijie wenming de jishu* rival with\_ZHE\_world\_famous\_DE\_technology 'rival with world-famous technology'), 过 GUO (e.g., 执教过中国队 *zhijiao guo zhongguodui* coach\_GUO\_Chinese\_National\_team 'have coached Chinese Notional team') and 了 LE (e.g., 出版了一套系列丛书 *chuban le yitao xilie congshi* publish\_LE\_one\_CL\_series\_books 'have published one series of books') can be appeared after VO compounds.

#### i. Feature 9: The semantic role of $O_2$

The semantic role of  $O_2$  has been discussed by numbers of scholars (e.g., Liu and Li, 1998; Liu, 1993), and agreement has been reached that various semantic roles can be taken by VO compounds. Therefore the semantic role of  $O_2$  is also included to see if this feature can help to distinguish the variation differences between Mainland and Taiwan varieties.

#### j. Feature 10: Animacy of the taken object

Animacy is an important cognitive category in differentiating various linguistic variations (e.g., Bresnan and Ford, 2009). In this model, the animacy variable was simplified into a binary category of animate vs. inanimate. The taken object of VO can be either animate (e.g., 中意这个徒弟 *zhongyi zhege tudi* like\_this\_apprentice 'like this apprentice') or inanimate (e.g., 中意这道菜 *zhongyi zhe dao cai* like\_this\_CL\_dish 'like this dish').

As can been seen from annotation schema, to have sufficient data for simultaneous comparisons, most features (7 out of 10) were simplified into binary values: yes and no (respectively denoted by 1 and 0). For example, 'AspectualMarker' was simplified to 'has' and 'not has', 'complexity of the taken complement' was also simplified to 'has modifier' or 'not has modifier'. All the transitive tokens extracted from the corps were manually annotated. The data collection and annotation procedure will be explained in the following section.

#### 4.3.2.2 Methodology

For all the 155 VO compounds in the wordlist, 46 of them are actually used as intransitive verbs in both varieties (i.e. the transitivity frequencies in both variants are smaller than 1‰, e.g., 参 展 / 出 丑 / 对 话 / 联 网 *canzhan/chuchou/duihua/lianwang* 'join an exhibition/make people feel foolish/have a dialogue with/networking'). Therefore the distributional differences are investigated among the other 109 (=155-46) transitive VO compounds.

For these 109 VO compounds, about 200 transitive tokens are randomly collected for each verb, half from Gigaword Taiwan Central News Agency subcorpus and the other half from Gigaword Mainland Xinhua News Agency. One thing should be noted is that, since some of the VO compounds may not have enough 200 transitive tokens, in that case I include all the transitive usages (as many as possible). The selection follows the principle that it could cover different uses of each VO compound. All examples collected for analysis were manually annotated based on the 10 features. The main annotator is a trained expert on Chinese linguistics. Any ambiguous cases were discussed with another two experts in order to reach an agreement. The reliability and validity of the annotation schema has been reported in Section 2.2.5 of Chapter 2.

### 4.3.2.3 Mixed-effect Logistic Regression Model

Based on the annotated corpus data, the distributional differences in semantic/syntactic properties needed to be examined through statistical

analyses. The first statistical test I refer to is the mixed-effect logistic regression model (GLMM) (Pinheiro and Bates, 2000) implemented in the R package lme4 (Bates and Maechler, 2009). This model follows the paradigm of Bresnan (2007, 2010) and is consistent with the study of Light Verb Construction in Chapter 3.

Similar to the study of Light Verb Variations, the study of VO variations also needs mixed-effect model for the same reason, as the dataset includes different verbs and the variability in the verbs should not be disregarded. The verb is considered as a random variable in this model. Feature 10 is not included in this regression model, since there are too many values for this feature and this may complicate the calculation. For the model for VO compound, the 'item' refers to different verbs, which is considered as the random-effect factor. The dependent variable is source (the source of the data: from Taiwan or Mainland), and the independent variables are the features we annotated (as shown in the annotation schema): polarity of the context ('polarity'), POS of the complement ('POS'), whether the complement is proper noun or not ('proper noun'), whether the complement is pronoun or not ('pronominality'), whether the VO compound occurs in structural parallelism ('structural parallelism'), whether the VO compound occur in headline ('headline'), whether the complement refers to animate or inanimate entity ('animacy'), whether the complement has modifier or not ('complexity'), whether the VO compound takes aspectual marker or not ('aspectual marker'). As mentioned above, to have sufficient data for simultaneous comparisons, most variables have been simplified to binary values: yes and no (respectively denoted by 1 and 0). The design of this model is consistent with the one has been conducted for light verb variations.

The specification of this model is shown below:

Formula: source ~ polarity + POS + propernoun+ pronominality + structural parallelism + headline + animacy + complexity + a spectual marker + (1 | item)

The result of logistic regression model is shown in Table 4.6. Among all the 9 features which are included in the model, 6 of them show significance: 5 features are significant at P<0.01 level while 1 feature is significant at P<0.05 level ('POS of the complement': whether the complement is deverbal noun or not). The estimates in the table represent the regression coefficients. The signs of the coefficients show the direction of the effects. To be particular, positive coefficients favor the 1 source (here refers to Taiwan) and negative coefficients favor the 0 source (here refers to Mainland).

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	1.12711	0.23234	4.851	1.23e-06 ***
Polaritynegative	-9.80745	934.17558	-0.010	0.991624
Polarityneutral	0.01529	0.13166	0.116	0.907553
Polaritypositive	1.56626	0.29038	5.394	6.90e-08 ***
POSclause	-14.66104	998.36870	-0.015	0.988283
POSdeverbal noun	-0.49266	0.19153	-2.572	0.010105 *
POSevent noun	-1.70831	0.26539	-6.437	1.22e-10 ***
POSNP	-0.61451	0.17187	-3.576	0.000350 ***
POSquantity	-1.11109	0.24406	-4.552	5.30e-06 ***
POStime	-0.73393	0.23578	-3.113	0.001853 **
POSVP	0.09721	0.19141	0.508	0.611553
propernoun1	-0.16783	0.05549	-3.025	0.002490 **
animate1	0.08194	0.07392	-1.108	0.267651

pronominality1	-0.38835	0.11730	-3.311	0.000931 ***
structualparalism1	0.02624	0.17203	0.153	0.878780
headline1	-0.64642	0.05952	10.861	< 2e-16 ***
withmodifier1	-0.02366	0.04583	-0.516	0.605678
aspectualmarker1	-1.43517	0.15300	-9.380	< 2e-16 ***
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1				

Table 4.6 The result of logistic regression model for VO compound

In the following section, each feature will be examined one by one to discuss the variation difference between Mainland and Taiwan VO compounds.

The main effect of 'polarity' is negatively significant. To be particular, the feature 'polarity positive' has a positive coefficient 1.56626, indicating the positive context favors Taiwan VO compound (e.g., 各个部门热切希望他可以 插 手 gege bumen reqie xiwang ta keyi chashou every\_department\_intent\_hope\_3 SG\_can\_intervene 'Every department is intent on his intervenes'). The negative and neutral contexts do not show significant preference for either Mainland or Taiwan Mandarin.

The feature 'POS' also shows significance in regression model. The feature 'deverbalnoun', 'eventnoun' and 'NP' all show significant negative coefficient: -0.49266, -1.70831 and -0.61451 respectively. This result implies that Mainland VO compounds are more likely to take deverbal noun, event noun and NP as the complements. Again we can take 插手 *chashou* 'intervene' as an example, constructions such as 插手电商的运营 *chashou dianshang de yunying* intervene\_electronic\_business\_DE\_operation 'intervene the operation of electronic business' (deverbal noun), 插 手比 *赛 chashou bisai* intervene competition 'intervene the competition' (event noun) and 插手中亚

问题 chashou zhongya wenti intervene\_Central\_Asia\_issue 'intervene Central Asia issue' (NP) are usages extracted from Mainland corpus. Moreover, for 'POS' of the taken complement, the feature 'quantity' and 'time' also have significant negative coefficient: -1.11109 and -0.73393 respectively. This indicates that the quantity and time complement also significantly favor Mainland VO compound. For example, constructions like 执教五十年 *zhijiao wushi nian* coach\_fifty\_years 'coach for fifty years', 投资三千万 *touzi sanqianwan* invest\_30 million 'invest 30 million' are highly frequently detected in Mainland corpus.

In terms of pronominality, the negative coefficient -0.38835 indicates that Mainland VO compounds are significantly favored by pronoun (e.g., 关心他/投 资 这 家 企 NK. guanxin ta/touzi zhejia concern 3 qiye SG/invest in this CL enterprise 'concern him/invest in this enterprise'). The feature 'proper noun' also has negative coefficient -0.16783 which indicates that proper noun as the complement, is significantly preferred by Mainland Mandarin (e.g., 入籍广东 ruji Guangdong naturalize Guangdong 'be Guangdong'; 受 聘 北 大 naturalized into shoupin beida employ Peking University 'be employed by Peking University'). The feature 'headline' also shows negative coefficient -0.64642, which implies that Mainland VO compounds have significant preference of occurring in the headline. With respect to aspectual marker (zhe, le and guo), the negative coefficient -1.90630 indicates that Mainland light verbs are more likely to occur

with aspectual marker (e.g., e.g., 媲美着世界闻名的技术 pimei zhe shijie wenming de jishu rival with\_ZHE\_world\_famous\_DE\_technology 'rival with world-famous technology'; 执教过中国队 zhijiao guo zhongguodui coach\_GUO\_Chinese\_National\_team 'have coached Chinese Notional team'; 出版了一套系列丛书 chuban le yitao xilie congshi publish\_LE\_one\_CL\_series\_books 'have published one series of books').

There are three features which do not show significance in this logistic regression model (i.e. animacy, structural parallelism and complexity). However, in Chi-square test which will be shown in the next section, two of the three features have presented significant variation differences (i.e. structural parallelism and complexity). In other words, most of the annoatated features are very effective in distinguishing between Mainland and Taiwan VO compound usages.

#### 4.3.2.4 Chi-square Test

In the previous section, the results of mixed-effect logistic regression model allow us to have a general tendency of VO variations between Mainland and Taiwan Mandarin. Furthermore, Chi-square test is used in this section, to test for the significance of the co-occurrence of each factor with each variety for each individual VO compound. Conducting Chi-square test enables us to quantify and study the variation for each individual VO compound, and is also consistent with the method adopted in the study of light verb variations in Chapter 3.

The dataset for Chi-square test contains 109 VO compounds which have transitive usages in both Taiwan and Mainland. 24 words which do not have enough collected data which violate the assumption of Chi-square test were excluded from further analysis (the cells do not have expected counts greater than or equal to five). For the other 85 VO compounds, Chi-square test was taken based on the 200 annotated transitive tokens for each verb (half from Taiwan sub-corpus and the other half from Mainland sub-corpus). The tool used here is IBM SPSS v. 22. The two variables are considered to have significant difference at P<0.05 level. The results of Chi-square tests for the 85 words are shown in Table 4.7.

	Word Number	Words show significance
No	14	入境 rujing 'enter', 驰誉 chiyu 'be famous', 留学
signifi		<i>liuxue</i> 'study abroad', 入籍 ruji 'be naturalized', 定
cance		都 dingdu 'establish a capital', 登场 dengchang
		'come on stage', 迁居 qianju 'mvoe', 中意 zhongyi
		'like', 操心 caoxin 'worry about', 满意 manyi 'be
		satisfied with', 致函 zhihan 'write to', 出土 chutu
		'be underthed', 变身 bianshen 'transform', 更名
		gengming 'rename'
POS of	33	驻军 zhujun 'station troops', 聚会 juhui 'gather', 投
compl		身 toushen 'plunge into',立足 lizu 'base upon',执教
ement		zhijiao 'coach',任职 renzhi 'hold a post', 插手
		chashou 'intervene', 动员 dongyuan 'mobilize', 起
		诉 qisu 'sue', 帮忙 bangmang 'do a favor', 提名
		timing 'nominate', 投诉 tousu 'complain', 控股
		konggu 'hold controlling interest', 投资
		touzi'invest', 入股 rugu 'buy a share', 寄语 jiyu
		'send word', 寄望 jiwang 'send hope', 签约 qianyue
		'sign a contract', 结缘 jieyuan 'become attached to',
		参选 canxuan 'participate in election', 接手 jieshou
		'take over'. 称雄 chengxiong 'bestride'. 献身

		xianshen 'devote oneself to', 赠书 zengshu 'donate
		book', 曝光 baoguang 'expose', 聚焦 jujiao 'focus',
		出兵 chubing 'dispatch troops'. 插足 chazu
		'participate' 拜师 <i>baishi</i> 'acknowledge as master'
		续约 ruvue 'renew a contract' 改名 gaining
		'rename' 拘犯 haovuan 'complain' 法储 sharian
		'he suspected'
Proper	37	登陆 danaly 'land' 今师 huishi 'join forces' 任职
noun	57	豆帽 uengiu land, 云师 nuism join loices, 江东
noun		renzni noid a post, 油足 cnazu participate, 按机
		Jiegui conform to, 获赠 huozeng receive, 信迫
		jiedao 'channel through', 过境 guojing 'trasit', 拜师
		baishi 'acknowledge as master', 致信 zhixin'write
		to', 取悦 quyue 'please someone', 改名 gaiming
		'rename', 聚焦 jujiao 'focus', 续约 xuyue 'renew a
		contract', 现身 xianshen 'appear', 立足 lizu 'base
		upon', 挑战 tiaozhan 'challenge', 寄望 jiwang 'send
		hope', 驻军 zhujun 'station troops', 命名 mingming
		'name', 出身 chushen 'originally birth from', 约会
		yuehui 'date', 对阵 duizhen 'clasli', 提名 timing
		'nominate'. 结缘 <i>iievuan</i> 'become attached to'. 执教
		zhijiao 'coach' 迁怒 ajannu 'vent on' 得罪 dezuj
		'offend' 签约 <i>ajanvue</i> 'sign a contract' 发进
		canvuan 'participate in election' $\mathbb{R} \bigtriangleup$ inhui
		(asther' 敬信 shirin'write te') 股 men (huy a
		gattlef, 我同 <i>zhuxin</i> white to, 八股 <i>rugu</i> buy a
		share, 你朝 <i>chengoa</i> seek negeniony, 帝仁
Duono	4	bangmang do a favor, 进车 jinjun march towards
PIONO	4	接手 <i>jieshou</i> 'take over', 氷沚 <i>qiuzheng</i> 'verify', 起
niniant V		诉 qisu `sue', 犹赠 huozeng `receive'
y Structu	10	立足 lizu 'hasa upon' 把土 hasuan 'suprontoo' 取
ral	10	立足 uzu base upon, 汇入 buguun guarance, 來
paralle		信 quain will the trust, 献 为 attansien devote
lism		onesell to, 成版 <i>fangyan</i> look anead, 你朝
		<i>chengba</i> 'seek hegemony', 会师 <i>huishi</i> 'join forces',
		聚焦 <i>jujiao</i> 'focus', 挑战 <i>tiaozhan</i> 'challenge', 求业
	21	qiuzheng 'verify'
Headlı	21	现身 xianshen 'appear', 登陆 denglu 'land', 聚焦
ne		jujiao 'focus', 受聘 shoupin 'be hired by', 迁怒
		qiannu 'vent on', 拜师 baishi 'acknowledge as
		master', 把关 baguan 'guarantee', 寄语 jiyu 'send
		word', 寄望 jiwang 'send hope', 结缘 jieyuan
		'become attached to', 获赠 huozeng 'receive', 改名

		gaiming 'rename' 抗衡 kangheng 'compete' 落户
		Jushu 'settle' 社宏 zhujun 'station troops' 执数
		<i>Luonu</i> seule, 在手 <i>Znujun</i> station hoops, wax
		Znijiao coacn, 大心 guanxin concern, 山市 cnuxi
		attend', 水切 quuzhu ask for neip', 头沙 xuyue
Delarit	0	
Polarit	8	称霸 chengba 'seek hegemony', 涉足 shezu 'set toot
y OI		in', 插手 chashou 'intervene', 取悦 quyue 'please
t		someone', 结缘 jieyuan 'become attached to', 取信
ι		quxin 'win the trust', 曝光 baoguang 'expose', 寄语
		<i>jiyu</i> 'send word'
Compl	27	放眼 fangyan 'look ahead', 聚焦 jujiao 'focus', 出
exity		兵 chubing 'dispatch troops', 立足 lizu 'base upon',
of		出身 chushen 'originally birth from', 调任 diaoren
object		'be transferred', 任职 renzhi 'hold a post', 关心
		guanxin 'concern', 出版 chuban 'publish', 挑战
		tiaozhan 'challenge', 致力 zhili 'devote to', 媲美
		nimei 'rival with' 把关 haguan 'guarantee'. 投资
		touzi 'invest' 命名 mingming 'name'. 讲军 jinjun
		'march towards' 呦么 chiming 'he famous' 借道
		iiadao 'channel through' 游难 hi'nan 'seek shelter'
		Jeau Chamer unough, 西方 on nun seek shener, 场王 chashou (intervene) 插卫 chazu (participate)
		山市 shuri 'attend' 但罪 dagui 'affend' 迁奴
		田府 Chuxi allend, 行非 aezul ollend, 江心
		qiannu Vent on, 到后 zhixin Write to, 奇全 jiwang
	-	'send hope', 续约 xuyue 'renew a contract'
Aspect	6	出席 chuxi 'attend', 出版 chuban 'publish', 得罪
ual		dezui 'offend', 命名 mingming 'name', 执教 zhijiao
marker		'coach', 曝光 baoguang 'expose'
Anima	17	受聘 shoupin 'be hired by', 任职 renzhi 'hold a
cy		post', 起诉 qisu 'sue', 帮忙 bangmang 'do a favor',
		提名 timing 'nominate', 得罪 dezui 'offend', 迁怒
		qiannu 'vent on', 致谢 zhixie 'express thanks to', 签
		约 qianyue 'sign a contract', 约会 yuehui 'date', 取
		信 <i>auxin</i> 'win the trust', 放眼 <i>fangyan</i> 'look ahead',
		关心 guanxin 'concern', 投诉 tousu 'complain', 把
		关 <i>baguan</i> 'guarantee'. 寄望 <i>jiwang</i> 'send hope'. 命
		又 minamina 'name'
		名 mingming 'name'

Table 4.7 Result of Chi-square test for VO compound

As can be seem from Table 4.7, among all the 85 VO compounds which are tested in the study, 14 words show no significant distributional difference between Taiwan and Mainland Mandarin in any of these 10 features: 入境 rujing 'enter', 驰誉 chiyu 'be famous', 留学 liuxue 'study abroad', 入籍 ruji 'be naturalized', 定都 dingdu 'establish a capital', 登场 dengchang 'come on stage', 迁居 qianju 'mvoe', 中意 zhongyi 'like', 操心 caoxin 'worry about', 满 意 manyi 'be satisfied with', 致函 zhihan 'write to', 出土 chutu 'be underthed', 变身 bianshen 'transform', 更名 gengming 'rename'. For the other 71 words which show significant variation differences in at least one feature, their tendency differences will be discussed in details by each feature.

# a) Feature 1: Syntactic type of object

First, about the syntactic type of the object taken by VO compound: among the 71 words, 33 words demonstrate significant variation difference in this feature. Among these 33 words, 15 words (投身 *toushen* 'plunge into', 插手 *chashou* 'intervene', 动员 *dongyuan* 'mobilize', 投诉 *tousu* 'complain', 控股 *konggu* 'hold controlling interest', 寄语 *jiyu* 'send word', 寄望 *jiwang* 'send hope', 签 约 *qianyue* 'sign a contract', 参选 *canxuan* 'participate in election', 接手 *jieshou* 'take over', 聚焦 *jujiao* 'focus', 插足 *chazu* 'participate', 献身 *xianshen* 'devote oneself to', 曝光 *baoguang* 'expose' and 涉嫌 *shexian* 'be suspected') present the tendency that the Taiwan VO compound tends to take event-denoting object (VP and clause) while the Mainland counterpart prefer to take common NP as the object. Only 4 words (立足 *lizu* 'base upon', 抱怨

baoyuan 'complain', 起诉 qisu 'sue' and 帮忙 bangmang 'do a favor') display the opposite tendency (Mainland VO compound is more likely to take the event-denoting objects). The similar tendency has already been presented in the result of logistic regression model. Several examples are shown in Table 4.8 below:

Words	TW	TW examples	ML	ML example
插 手 chash ou 'inter vene'	VP	插手经营家族企业 chashou jingying jiazu qiye intervene_manage_family_e nterprise 'intervene the management of family enterprise'	NP	插手地方事务 <i>chashou</i> <i>difangshiwu</i> intervene_local_aff airs 'intervene local affairs'
投诉 tousu 'comp lain'	Clause	投诉澳娱分紅不合理 tousu aoyu fenhong bu heli complain_STDM_dividend_ BU_fair 'complain that the dividends of STDM are not unfair'	NP	投诉有关部门 tousu youguan bumen complain relevant department 'complain about relevant department'
接 手 <i>jiesho u</i> 'take over'	VP	日本接手举办这项比赛 riben jieshou juban zhe xiang bisai Japan_take over_hold_this_CL_competi tion 'Japan took over to hold this competition.'	NP	他很愿意接手这个 职务 ta he yuanyi jieshou zhege zhiwu 3 SG_very_willing_t ake over_this_CL_post 'He is very willing to take over this post.'

Table 4.8 Variation differences in syntactic type

Therefore, in terms of the syntactic type of the taken object, it can be summarized as Taiwan VO compounds present a very strong tendency of taking event-denoting objects, while Mainland have more preference of taking common NP as the object.

Another tendency shown in this feature is that, when VO compounds can take both common NP (e.g., 撤军黎巴嫩 *chejun libanen* withdrawal troop Lebanon "troop withdrawal from Lebanon",执教中国男篮 *zhijiao zhongguo nanlan* coach\_Chinese\_Male\_Basketball 'coach Chinese National Male Basketball team') and quantity NP (e.g., 撤军三十万 *chejun sanshiwan* withdrawal\_troop\_300000 'withdrawal 300000 troops') or time NP (e.g., 执教 五十年 *zhijiao wushinian* coach\_fifty\_years 'coach for fifty years') as the object, Taiwan VO compounds more prefer the common NP (e.g., *chejun libanen* withdrawal troop Lebanon "troop withdrawal from Lebanon") while the Mainland may have more preference for the quantity and time object (e.g., 投资三千万 *touzi sanqianwan* invest\_30 million 'invest 30 million' /执教五十 年 *zhijiao wushinian* coach\_fifty\_years 'coach for fifty years'). This has also been shown in the result of logistic regression model.

#### b) Feature 2: Proper noun

Although proper noun has shown significant coefficient in logistic regression model, it does not present preference difference in Chi-squares test. Among all the 37 words which show significant difference in this feature, 18 VO compounds in Taiwan show the preference of choosing a proper noun as the object, while the other 19 words present the opposite preference.

#### c) Feature 3: Pronominality

In the result of logistic regression model, it has been shown that Mainland VO compounds are more likely to take pronoun as the complement. This tendency is also shown in the result of Chi-square test. 4 words (接手 *jieshou* 'take over', 起诉 *qisu* 'sue', 求证 *qiuzheng* 'verify' and 获赠 *huozeng* 'receive') display significant difference in this feature, and 3 of them (接手 *jieshou* 'take over', 求证 *qiuzheng* 'verify' and 获赠 *huozeng* 'receive') show the preference of taking pronoun in Mainland, while only 起诉 *qisu* 'sue' is significantly favored by pronouns in Taiwan.

#### d) Feature 4: Structural parallelism

Referring to the feature of structural parallelism, although this feature does not show significance in logistic regression test, the variation tendency presented by Chi-square test is quite clear. Among all the 10 words which present significant difference in this feature, 7 words (把关 *baguan* 'guarantee', 取信 *quxin* 'win the trust', 献身 *xianshen* 'devote oneself to', 称霸 *chengba* 'seek hegemony', 聚焦 *jujiao* 'focus', 求证 *qiuzheng* 'verify' and 挑战 *tiaozhan* 'challenge') demonstrate the tendency of favoring structural parallelism strategy in Mainland, while only 3 words (立足 *lizu* 'base upon', 会师 *huishi* 'join forces' and 放眼 *fangyan* 'look ahead') display the opposite tendency. In other words, Mainland VO<sub>1</sub>+O<sub>2</sub> construction is more likely to be appeared in the parallelism structures. The transitivization of VO in Mainland is more likely to be appeared in transitive parallel structure in the context (e.g., 他们坚持把关 审核、强化责任、规范操作、配套协调 *tamen jianchi baguan shenhe*, *qianghua zeren*, *guifan caozuo*, *peitao xietiao*, 3 PL\_insist\_guarantee\_check\_strengthen\_responsibility\_standardize\_operation\_s upport\_coordination 'They insist on guarantee the checking, strengthen the responsibility, standardize the operation and supporting the coordination').

#### e) Feature 5: Headline

For this feature, the tendency is very obvious that Mainland VO<sub>1</sub>+O<sub>2</sub> construction is more likely to be appeared in the headline, while the Taiwan ones prefer the normal texts (15 words). Example as 登陆 *denglu* 'land', 15.38% (16/104) of VO<sub>1</sub>+O<sub>2</sub> constructions (e.g., 登陆上海 *denglu shanghai* 'land on Shanghai') are shown in the headline in Taiwan, while in Mainland the frequency is obviously higher (34.62%=36/104). The same tendency has already been shown in the result of logistic regression model.

#### f) Feature 6: Polarity of context

About the polarity of the context, 8 words show variation differences in this feature. The difference between Taiwan and Mainland is not very obvious. For Taiwan variety, 2 words prefer neutral context, 3 words have preference for negative context and 4 words favor the positive context (words can have significance in more than one polarity). While in Mainland, only 3 words prefer the negative context, the other 5 all prefer the neutral context, and no word

presents preference for the positive context. We can take 插手 chashou 'intervene' as an example. In Mainland, most of the constructions are shown in negative context (e.g., 各方势力经常插手流通环节,造成市场秩序十分混 乱 gefang shili jingchang chashou liutong huanjie, zaocheng shichang zhixu all parties force often intervene lin flow, shifen hunluan cause market order very disruption 'The parities often intervene in the flow of links, resulting in the disruption of market order.'). In Taiwan, 插手 chashou 'intervene' are frequently used in both neutral (e.g., 她很少不插手妹妹的事 ta henshao chashou meimei de shi 3 SG\_rarely\_intervene\_little sister\_DE\_thing 'She rarely intervene the business of her little sister.') and positive context (e.g., 各个部门热切希望他可以插手 gege bumen reqie xiwang ta keyi chashou every department intent hope 3 SG can intervene 'Every departments are intent on his intervenes'). The same tendency that Taiwan  $VO_1+O_2$  construction favors positive context has already been shown in the result of logistic regression model.

# g) Feature 7: Complexity of object

As discussed in the Section 4.3.2.1, the complexity of object is measured by 'whether the object has modifier or not'. Although this feature does not show significance in logistic regression model, the significance difference presented by Chi-squarest test is also quite clear. Among 27 words which show significant difference in this feature, 18 words shows preference over this feature in Taiwan (e.g., 媲美 pimei 'rival with', 放眼 fangyan 'look ahead', 致力 zhili 'devote to'). Only 9 words show the opposite tendency. Complicated examples as 任职国界委员会与大陆礁层委员会所辖海洋事务部 renzhi guojie weiyuanhui yu dalu jiaoceng weiyuanhui suoxia haiyang shiwubu hold a post\_boarder\_committee\_and\_Continental\_reef\_committee\_jurisdiction\_ocean \_affairs\_Ministry 'hold a post in Boarders Committee and Ministry of Ocean Affairs under the jurisdiction of the Continental Reef Commission' /媲美这个 世界一流名声大噪的科技新产品 pimei zhege shijieyiliu mingsheng dazao de keji xinchanpin rival with\_this\_world-class\_famous\_DE technology new product 'rival with this world-class famous new technology product' are not very commonly found in Mainland Mandarin. Therefore, Mainland VO more prefers the bare object while the object of Taiwan VO is usually complicated, with modifier in front.

#### h) Feature 8: Aspectual marker

The aspectual marker taken by VO compounds is also considered. Although only 6 words display significance variation difference in this feature, the difference is absolute. All of the 6 words (出席 *chuxi* 'attend', 出版 *chuban* 'publish', 得罪 *dezui* 'offend', 命名 *mingming* 'name' and 执教 *zhijiao* 'coach') show significant favor of aspectual marker (either 了 LE or 过 GUO) in Mainland, while their Taiwan counterparts show dis-favor over aspectual marker. This tendency is consistent with the findings presented in logistic regression model as well as in previous light verb variation study (Huang et al., 2014).

For the other 2 features (Feature 10: animacy of the object; Feature 9: semantic role of the object), Chi-square test results do not show very strong difference between Mainland and Taiwan. Therefore they are not discussed in this section.

#### 4.3.3 Transitivity Degree in VO Variations

In the previous section, the variation differences in both transitivity frequency and semantic/syntactic properties have been investigated by various statistical tests (Z-test, likelihood ratio test, mixed-effect logistic regression model and Chi-square test). Based on their distributional differences between Taiwan and Mainland Mandarin (differences in transitivity frequency as well as in semantic/syntactic properties), it is possible to argue that the VO compound itself may differ in its degree of transitivity between these two varieties. Taiwan VO compound may have a higher degree of transitivity compared to its Mainland counterpart.

The concept of transitivity has already been reviewed in Section 3.6 of Chapter 3. In short, the transitivity degree of a verb can be measured by the ability of taking the objects, and by collocation preference (i.e. what kinds of complements are more likely to be taken). Same as the study of light verbs, the transitivity degree of VO compound is also measured by both transitivity frequency and the semantic/syntactic properties of the taken objects. In terms of transitivity frequency, Taiwan VO compounds have higher degree of transitivity, compared to the Mainland ones. There are mainly two evidences to support this argument. The first evidence is that Taiwan VO compounds tend to have higher transitivity frequency, as the result of Z-test and likelihood ratio test showed (in Section 4.3.1). Taiwan VO compounds are more likely to be used as transitive verbs, and this tendency becomes more obvious when the transitivity differences are prominent. The second evidence is that Taiwan VO compounds are found to have the tendency of taking more types of NPs and VPs as the complements with less collocation constraints, as has been shown in the result of logistic regression and Chi-square test. Based on the two evidences, it can be summarized as, in terms of the ability of taking objects, the VO in Taiwan is more transitive, in the sense that it is more likely to be used transitively in different contexts.

Besides these, the variation differences in semantic/syntactic properties in taken complements also illustrate the higher transitivity of Taiwan VO compounds. Similar to the study of light verb variations, here I also refer to the 10 parameters proposed by Hopper and Thompson (1980), to identify the degree of transitivity. The differences in semantic/syntactic properties actually are related to two of the parameters (i.e. individuality and affectedness of object). The individuality refers to both the distinctness of the patient from the agent and to its distinctness from its own background, and the degree to which an action is transferred to a patient is a function of how completely that patient is affected. There are mainly three evidences related to the two parameters.

The first evidence is, as has been presented in the result of logistic regression model and Chi-square test, Taiwan VO compound has the preference of taking VP or clause as the complement (e.g., 投诉澳娱分红不公平 tousu aoyu fenhong bu heli complain STDM dividend BU fair 'complain that the dividends of STDM are not unfair'; 插手经营家族企业 chashou jingying jiazu *give* intervene manage family enterprise 'intervene the management of family enterprise'), while the Mainland one may prefer the deverbal noun (e.g., 插手 电 商 的 经 营 运 营 chashou de dianshang yunying intervene electronic business DE operation 'intervene the operation of electronic business') or common NP (e.g., 插手中亚问题 chashou zhongya wenti intervene Central Asia issue 'intervene Central Asia issue'). VP, compared to the deverbal noun, is obviously higher in individuality, as the patient is overt. Also, as the patient is already in the construction, the affectedness of the objects in Taiwan is also higher than that of in Mainland. The higher individuality and affectedness of object both indicate the higher transitivity of Taiwan VO compounds.

The second evidence is, as shown in Chi-square test, when VO can take both common NP and time/quantity object, Taiwan VO compounds prefer the common NP (e.g., 撤军黎巴嫩 *chejun libanen* withdrawal troop Lebanon "troop withdrawal from Lebanon") while the Mainland may have more preference for the quantity and time object (e.g., 投资三千万 *touzi sanqianwan* invest\_30 million 'invest 30 million'; 执教五十年 *zhijiao wushinian*  coach\_fifty\_years 'coach for fifty years'). The common NP (often refers to a concrete object), compared to the object denoting time and quantity, is obviously higher in concreteness, which indicates higher individuality, and further, higher transitivity.

The third evidence is that the Chi-square tests show that Taiwan tend to prefer the complicated object (object has modifier with DE) while Mainland may prefer the bare object. This contrast actually corresponds to the distinction of endurant/perdurant variations (Huang, 2016). An endurant is an entity that has spatial component but does not depend on a specific time of occurrence. A perdurant is an entity which has a time element crucially associated with its meaning (Huang, 2016). According to Huang (2016), DE-insertion is allowed only when the object has perdurant properties. In contrast, the bare object without DE, which is preferred by Mainland Mandarin, may refer to a generic concept, which is endurant. Therefore the referentiality of the bare object is less, which demonstrates a lower individuality, and also a lower transitivity of VO compound in Mainland. In summary, for both transitivity frequency and properties of the taken complements, Taiwan VO compounds have presented a higher degree of transitivity, compared to their Mainland counterparts.

#### 4.3.4 Summary of Study 1

To make a summary, this study has shown that VO compounds in Mainland and Taiwan Mandarin can be differentiated effectively by most of the annotated features. The statistical models used in this study (both logistic regression model and Chi-square test) are very helpful in providing comprehensive descriptive account of grammatical variations. The results indicate that Taiwan VO compounds tend to have higher degree of transitivity, in terms of transitivity frequency as well as semantic/syntactic properties of the taken complement (according to Hopper and Thompson, 1980).

# 4.4 Study 2: Lexicalization and Transitivity

# **4.4.1 Introduction**

The popularity of  $VO_1+O_2$  construction has attracted the interests of numbers of scholars in Chinese linguistics. As mentioned in Section 4.2.1, one research question that often be addressed by previous scholars is the transition requirement for VO compound (i.e. what kind of VO is easier to be transferred from intransitive to transitive). The importance of lexicalization has often been mentioned by previous researches. Previous studies (e.g., Li and Wu, 2017) have agreed that the transition process mainly includes noun incorporation (i.e.  $O_1$  gets incorporated into V), which is substantially a process of lexicalization in the sense that the V and  $O_1$  are condensed into a disyllabic compound in the form of  $VO_1$ , which can be used independently to describe an event. Agreement has been reached that for a VO compound, the ability of taking the object is closely related to its lexical status, i.e. the higher degree of lexicalization, the more possibility it can take the object and be used transitively. Thus it has also become a common belief that for a VO compound, the degree of lexicalization.

can be used to measure its transitivity.

If this correlation between transitivity and lexical status is true for VO compound, it would be able to account for the variation differences between Mainland and Taiwan Mandarin. As has been observed in Section 4.3.1, there is a significant difference in transitivity frequency between Mainland and Taiwan Mandarin (i.e. Taiwan VO compounds have a significantly higher transitivity than the Mainland counterparts). Then based on the correlation between transitivity and lexical status, it is able to investigate whether the differences in transitivity between Taiwan and Mainland Mandarin indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum/process of lexicalization (i.e. whether the grammatical variations of VOs are dependent on the degree of lexicalization).

# **4.4.2 Literature Review**

Numbers of researches claim that for a VO compound, the ability of taking the object is closely related to its lexical status, i.e. the higher degree of lexical status, the more possibility it can take the object and be used transitively.

Liu (1998a) argues that whether a VO can take object or not is closely depend on its lexical status: the higher degree of lexicalization may lead to less influence to the syntactic structure, thus taking the object can be possible. Once the VO has been lexicalized, it can only be used as one unit. Luo (1998) has the similar argument that whether VO can take object or not usually closely related to its internal constituencies. Gao (1998) classifies the VO sequences into three types: the ones cannot take object (intransitive); the ones can take the object freely; and the ones can take the object with many restrictions. He also argues that the VO that can take object without constraints is the one who has been highly lexicalized (the semantic meaning of the VO is not predictable from its internal constituents).

Actually their definition of lexicalization is in accordance with that proposed by Brinton and Traugott (2005, p. 96), which claims that lexicalization is to use a syntactic construction or word formation as a new form that cannot be completely derivable or predictable from the constituents of the construction or the word formation pattern. Over time there may be further loss of internal constituency and the item may become more lexicalized.

In addition, the degree of lexicalization can be tested through separation test (e.g., Her, 1996; Liu, 1998a). Liu (1998a) argues that the degree of lexicalization of a VO can be measured by its separation ability. The easier it can be separated, the higher degree of its lexicalization. Her (1996) also used the separation test to differentiate three different types of VO sequence (VO word, phrase or dual-status). To be particular, a VO as a word cannot be inserted into any grammatical elements; if it is a phrase, since the structure is syntactically transparent, grammatical elements such as aspectual marker, possessive NP, object, classifiers and various adjunctive modifiers can be inserted between V and O. According to Her (1996), this classification criteria is actually based on the 'Lexical Integrity Hypothesis', proposed by C-T. J. Huang (1984, p.60): no phrase-level rule may affect a proper subpart of a word. Since a VO compound as a word is thus a lexical unit whose internal structure is of a V+O (Her, 1996), an important feature that distinguishes a lexical units from a phrase is the lexical integrity. Brinton and Traugott (2005, p.94) have also proposed the productivity of lexicalization, which is related to the separation ability. Productivity is understood as the 'ability of word-forming elements to be used to form new linguistic expressions' and is a gradient concept. As presented in Figure 4.2, lexical items from a cline defined primarily with respect to degree of fusion in internal structure. Lexical items range from fully transparent to less transparent, then to idiosyncratic (from middle to the left):



Figure 4.2 Synchronic clines of lexicality and grammaticality

L1 = partially fixed phrases, e.g., lose sight of, agree with,

L2 = complex semi-idiosyncratic forms, e.g., desktop,

L3 = simplexes and maximally unanalyzable idiosyncratic forms, e.g., desk

Lexicalization is a process to non-productive, therefore the degree of lexicalization can be reflected in the degree of fusion in internal structure.

As discussed above, previous researches all claim that the lexical status of a VO sequence can be measured by the separation ability, and its lexical status usually has a close correlation with the ability of taking the object. Therefore, based on the previous studies, it has become a common belief among linguistic researchers that there is a strong correlation between the transitivity of VO and whether the VO is separable (the lexical status), i.e. the VO which cannot be separated is much more likely to be used as a transitive verb, and vice versa.

Gao (1998) has classified VO into three types according to their separation ability: VO can be separated without constraints (e.g., 着急 zhaoji 'worry', 放 心 fangxin 'reassure', 发愁 fachou 'be anxious'), VO can be separated with constraints (e.g., 毕业 bive 'graduate', 担心 danxin 'anxious', 留心 liuxin 'be careful', 害怕 haipa 'be scare') and VO cannot be separated (e.g., 出版 chuban'publish', 当心 dangxin 'take care', 动员 dongyuan 'mobilize'). After investigating some of the VO in the corpus, he then concludes that all the VOs that cannot be separated are used as a transitive verb (e.g., 动员群众 dongyuan *qunzhong* mobilize masses 'mobilize the masses'), while the VOs which can be separated without constraints are usually cannot be used transitively (e.g., \*放 心他的能力 fangxin ta de nengli reassure 3 SG DE ability 'rest assured his ability'). For the VOs that can be separated with constraints, they usually have transitive usages in the corpus (e.g., 担心工程的进度 danxin gongcheng de jindu worry about project DE progress 'worry about the progress of the project'), but some of the words are still under the process of changing (e.g., ? 过目这份文件 guomu zhe fen wenjian look over this CL document 'look over this document').

#### 4.4.2.1 Research Gap

Although the correlation between transitivity and separation of VO compound has been well recognized by linguists, in literature I can barely find empirical studies using real world data to verify this common belief. For the very few studies (e.g., the study of Gao (1998) mentioned above) that are conducted based on empirical data, the data size is relatively small and the statistical methods are also quite simple (often just percentage or raw numbers). Although the numbers and percentages can reveal the difference, they cannot tell whether there is significance correlation or not.

Therefore it is important to investigate the correlation between transitivity and lexical status in a more objective way, with the assistance of large-scale corpus as well as the statistical tool.

From the perspective of language variation, to the best of my knowledge, no work has been conducted to investigate the variation differences in lexical status of VO compounds in different Mandarin varieties. As the variation differences in transitivity frequency has already been reported in Study 1, the variation differences in lexical status also need to be examined to study whether the transitivity variation differences are dependent on the degree of lexicalization.

Therefore in this study, a comparable-corpus based statistical approach is still adopted, first to verify whether the correlation between VO transitivity and its lexical status is empirically true. Secondly, the variation differences in lexical status of VO compound will be explored. Based on the comparison, the theoretical issue of whether the variation difference in transitivity frequency is dependent on degree of lexicalization will also be discussed.

#### 4.4.3 Methodology

### **4.4.3.1 Data Collection and Calculation**

# 4.4.3.1.1 Measurement of Separation Ability

The first aim of this study is to examine the relationship between transitivity frequency and lexical status of VO sequences. It should be noted that previous studies usually examine both separation status and transitivity issue in a dichotomy way. In other words, the VO is classified as separable vs. inseparable, transitive vs. intransitive (e.g., Gao, 1998; Her, 1996). But I argue that the issues of both separation and transitivity are not simply binary dichotomy, it should be considered as tendency/frequency. For separation ability, for example, both 把关 baguan 'guard a pass' and 插手 chashou 'intervene' are separable (e.g., 把了关 ba le guan 'guaranteed'; 插过手 cha guo shou 'have intervened'), but the frequency of separation usages are very different (i.e. 把关 baguan 'guarantee' is much more frequently to be used separately than 插手 *chashou* 'intervene'). In addition, the grammatical elements which can be inserted also vary across different verbs (e.g., plenty of elements can be inserted into 把关 baguan 'guarantee' (e.g., 把产品质量关 ba chanpin zhiliang guan guard produce quality pass 'guarantee the quality of products'/把好几道关 ba hao ji dao guan guard\_many\_several\_CL\_pass 'carefully check for several times'/把了关 ba le guan 'guaranteed'), while only aspectual marker can insert into 插手 chashou 'intervene' (e.g., 插了手 cha le shou 'intervened'/ 插 过手 cha guo shou 'have intervened'). Therefore frequency/percentage of separation usages (Ren and Wang, 2005) is used to measure the separation degree in this study.

RF=separated usages/all the usages (Ren andWang, 2005)

Example: separation degree for 操心 *caoxin* 'worry about' = the number of 操….心 usages /all the usages of 操心 *caoxin* 'worry about' + separation usages

#### 4.4.3.1.2 Measurement of Transitivity Degree

For transitivity of VO comound, it is observed that the transitivity degree also varies among transitive VO compounds. For example, although both 驰名 *chiming* 'famous' and 约会 *yuehui* 'date' can be used transitively, the frequency of being used as a transitive verb for 驰名 *chiming* 'famous' (e.g., 驰名中外 *chiming zhongwai* famous\_at home\_and\_abroad 'achieve renown both inside and outside the country') is much higher than that of 约会 *yuehui* (e.g., 约会拜 金女 *yuehui baijinn ü* date\_material girl 'date material girl').

The transitivity of VOs is measured by frequency also, as has been shown Section 4.3: Transitivity frequency=transitive tokens/all the tokens

For example, transitivity frequency of 签约 *qianyue* 'sign a contract' =number of transitive usages of 签约 *qianyue* 'sign a contract' /all the tokens of 签约 *qianyue* 'sign a contract'.

The VO word list used in this study is the same as that has been used in the previous study: the 109 transitive VO compounds which have been collected from previous researches (e.g., Liu and Li, 1998; Qian, 2011). In this study, 13 words that do not show significant variation difference in both transitivity frequency and Chi-square test were excluded for further analyses:

登场 dengchang 'come on stage', 操心 caoxin 'worry about', 致信 zhixin 'write to', 出土 chutu 'be unearthed', 参演 canyan 'act in', 更 名 gengming 'rename', 涉嫌 shexian 'be suspected', 领军 lingjun 'play a leading role', 揭秘 jiemi 'expose', 解码 jiema 'decoding', 启 航 qihang 'set sail', 失信 shixin 'break promise', 移情 yiqing 'love someone else'.

These words are ruled out because they are used identically/very similarly in two varieties (both in frequency and in semantic/syntactic properties), hence will not contribute to the examination of language variations. For all the 96 words in the wordlist, their separation and transitivity frequency in both Mainland and Taiwan were calculated.

#### 4.4.3.2 Statistical Method and Tool

In order to examine the relationship between the degree of transitivity of VO and the separation ability, two tests were conducted in this section. For the first test, Mann-Whitney U test was used to compare the transitivity frequency between words that have separation usages (hereafter 'separable VO compound') and words that have no separation usages in corpora (hereafter 'inseparable VO compound'<sup>4</sup>). For the second test, the separation frequency was taken into consideration, and Spearman's rank-order correlation was used to access the correlation between separation frequency and transitivity frequency. As both Mann-Whitney U test and Spearman's rank-order correlation are non-parametric (i.e. does not require the normality of the data), they are very suitable for the current analyses. The statistical tool I use is IBM SPSS V.22.

#### 4.4.4 Data Analysis and Result

# 4.4.4.1 Test 1: Comparison between Separable and Inseparable VO Compounds

The 96 words in the list were divided into two categories according to their separation ability: the VOs that have separable usages in the corpus (hereafter 'separable VO compound') and the VOs that do not have any separate usages in

<sup>&</sup>lt;sup>4</sup>Here I call the words which have no separation usages in corpora as 'inseparable VO compounds'. It does not mean that these words cannot be separated under any context (i.e. some 'inseparable words' may be observed to have separation usages in Google search). But as the Gigaword Corpus contains more than 1.1 billion characters, if no separation usages were found in the corpus, the separation frequency for these words should be very low.

the corpus (hereafter 'inseparable VO compounds'). Based on their separation status, one research question is addressed: Is there a significant difference in transitivity frequency between the two groups (i.e. Empirically and statistically, is it true that the inseparable VOs are more likely to be used in a transitive way, and vice versa?)

The separation status is annotated for both Taiwan and Mainland VO compounds data. The overall distribution of these two groups in two varieties is shown in Table 4.9.

	Taiwan Mandarin	Mainland Mandarin
Separable VO	39	37
Inseparable VO	57	59

# Table 4.9 Number of separable/inseparable VO in both Taiwan and Mainland Mandarin

As can be seen from Table 4.9, the numbers of separable and inseparable VO are very close between the two varieties, which indicate the comparability of the data.

A Mann-Whitney U test was run for the data in each variety to determine if there were differences in transitivity frequency between separable and inseparable VO compounds. Mann-Whitney U test is often presented as the non-parametric alternative to independent-sample t-test.

First for Taiwan data, distributions of the transitivity frequency for these two groups of words were similar, as assessed by visual inspection.

The result of Mann Whitney U test for Taiwan data is shown in Table 4.10. It displays that the median value of transitivity frequency for inseparable VO compounds (0.3607142857) is significantly higher than that for separable VO compounds (median value is 0.1378091873), U=801.000, Z=-2.316, P=0.021.

<b>7</b> 1		2	
Null hypothesis	Test	Sig.	Decision
The distribution of Taiwan	Independent	0.021	Reject the null
transitivity is the same	Samples		hypothesis
across categories of	Mann		
separation type	Whitney U		
	Test		

Hypothesis Test Summary

Asymptotic significances are displayed. The significance level is .05

Total N	96
Mann-Whitney U	801.000
Wilcoxon W	1581.000
Test Statistic	801.000
Standard Error	134.049
Standardized Test Statistic	-2.316
Asymptotic Sig. (2-sided test)	.021

Table 4.10 The result of Mann Whitney U test for TW data

Mainland data presents a similar result, Table 4.11 shows that median value of transitivity frequency for inseparable compounds in Mainland is 0.278, which is statistically significant higher than that of separable VO compounds (0.076), U=761.500, Z=-2.485, P=0.013.

	Null hypothesis	Test	Sig.	Decision	
1	The distribution of Mainland transitivity is the same across categories of separation type	Independent Samples Mann Whitney U Test	0.013	Reject the null hypothesis	

Hypothesis Test Summary

Asymptotic significances are displayed. The significance level is .05

Total N	96
Mann-Whitney U	761.500
Wilcoxon W	1464.500
Test Statistic	761.500
Standard Error	132.811
Standardized Test Statistic	-2.485
Asymptotic Sig. (2-sided test)	.013

Table 4.11 The result of Mann Whitney U test for Mainland data

Summary for test 1: The results of Mann Whitney U test in both Mainland and Taiwan Mandarin show that the transitivity frequencies between separable and inseparable VO compounds are significantly different. In particular, empirically and statistically, inseparable VOs are much more likely to be used in a transitive way.

#### 4.4.4.2 Test 2: Separation Frequency and Transitivity

The first test has shown that compared to separable VO compounds, the inseparable ones are more likely to be used as a transitive verb. But as has been mentioned in Section 4.4.3.1.1, the separation frequencies vary among separable VO compounds. Therefore, it is necessary to take the separation frequency into consideration. The research question addressed here is, whether the separation frequency and the transitivity frequency of the VOs are significantly correlated? (i.e. Is it empirically true that the more frequently it is used separately, the less frequently it can be used transitively?)

In this test, the separation frequency (separation frequency=separated usages/all the usages) is included as a variable for statistical analysis. The separation frequencies of all the separable VO compounds in both Mainland and Taiwan Mandarin have been calculated based on their actual usages in the corpora.

To answer the research question, Spearman's rank-order correlation was selected to assess the relationship between transitivity frequency and separation frequency in both Taiwan and Mainland Mandarin. The result of Taiwan data is shown in Table 4.12.

The result indicates that there exist a negative correlation between transitivity frequency and separation frequency in Taiwan Mandarin, the correlation is statistically significant, r=-0.221, P=0.03. That is, for a VO compound, the tendency is the higher its transitivity frequency, the lower its separation frequency.

		Correlation		
			TW separation	TW
			fre	transitivity
Spearman'	TW separation	Correlation	1.000	221*
rho	fre	coefficient		
		Sig. (2-tailed)		0.030
		Ν	96	96
	TW	Correlation	221*	1.000
	transitivity	coefficient		
		Sig. (2-tailed)	0.030	
		Ν	96	96

\*. Correlation is significant at the 0.05 level (2-tailed).

Table 4.12 The result of Spearman's rank-order correlation for Taiwan data Mainland data has presened the similar result, as shown below (Table 4.13). The negative correlation between transitivity frequency and separation frequency in Taiwan Mandarin is also statistically significant, r=-0.237; P=0.02.

1 ...

		Correlation		
			TW separation	TW
			fre	transitivity
Spearman'	TW separation	Correlation	1.000	237*
rho	fre	coefficient		
		Sig. (2-tailed)		0.020
		Ν	96	96
	TW	Correlation	237*	1.000
	transitivity	coefficient		
		Sig. (2-tailed)	0.020	
		Ν	96	96

\*. Correlation is significant at the 0.05 level (2-tailed). Table 4.13 The result of Spearman's rank-order correlation for Mainland data
Summary for test 2: The results of spearman's correlation in both Taiwan and Mainland show that for a VO compound, the more frequently it is used separately, the less likely it can be used as a transitive verb. In other words, for a VO compound, the more it is lexicalized, the more likely it is used in a transitive way.

### 4.4.4 Summary for the Two Tests

Based on the result of the two empirical tests, the tendency can be observed is that, compared to separable VO compounds, inseparable ones are more likely to be used in a transitive way. In addition, for a VO compound, the less frequently it is used separately, the more likely it is used as a transitive verb. In other words, if a VO sequence is less lexicalized, it has higher probability of taking an object. The tendency is in accordance with what has been presented in the previous papers (e.g., Gao, 1998; Liu, 1998a; Luo, 1998). It should also be noted that, although the result of the test 2 shows there is a significant negative correlation between transitivity and separation, the correlation coefficients in both Mainland and Taiwan are to some extent small, which indicates that the negative correlation is relatively weak. It is not hard to be explained. Although the lexical status of a VO compound does affect its transitivity, it is not the only factor. In the actual/real language, there are varieties of factors which can influence the transitivity apart from the lexical status. These may include both internal linguistic factors (e.g., word frequency; event type of the verb), and external social factors (e.g., the influence of social media or other

languages/dialects).

### 4.4.5 Grammatical Variation and Lexicalization

As shown in the above section, the transitivity of a VO is statistically significant correlated with its lexical status (e.g., measured by separation usages). Since the variation difference in transitivity between Mainland and Taiwan VO compounds has already been observed and reported in Section 4.3, then it is natural to address the following question: Are there any variation differences in lexical status (measured by separation ability) of VO compounds between Mainland and Taiwan? If the answer is yes, then do the differences in transitivity between Taiwan and Mainland Mandarin indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum/process of lexicalization (i.e. whether the grammatical variations of VOs are dependent on the degree of lexicalization of these VOs)?

As has been shown in Section 4.3, the transitivity of VO in Taiwan Mandarin is obviously higher than its Mainland counterparts (both in transitivity frequency and semantic/syntactic properties). To have a general picture of the separation comparison between two varieties, we can start from the comparison of averages of separation frequency, as shown below in Table 4.14.

	Separation frequency	Transitivity frequency
Taiwan	0.007068073	0.3538
Mainland	0.019565008	0.2919

 Table 4.14 Comparison of average separation/transitivity frequency

As can be observed in Table 4.14, the average separation frequency of Mainland VO compounds (0.019565008) is obviously higher than that of Taiwan VO compounds (0.007068073). Hence the general tendency between Mainland and Taiwan is that, while Taiwan VO compounds have higher transitivity frequency, VO compounds in Mainland have obviously higher separation frequency. In the next section, more specific statistical analyses for variety comparison will be presented.

In terms of transitivity frequency, Z-test has been used in Section 4.3.1 to test for the significance of the difference between two independent proportions, which further illustrate the differences in transitivity frequency between Mainland and Taiwan VO componds. All the 96 words can be categorized into three groups according to the result of Z-test: 1) the VO whose transitivity frequency in Taiwan is significantly higher than that in Mainland (Taiwan transitivity higher: 53 words); 2) the VO whose transitivity frequency in Mainland is significantly higher than that in Taiwan (Mainland transitivity higher: 23 words); 3) there is no significant difference in transitivity frequency between Mainland and Taiwan (no transitivity difference: 20 words) (all significant at P < 0.05 level). Based on this classification, it is obviously to see that for the compounds in 'Taiwan transitivity higher' group, the separation frequencies in Taiwan are much lower. For the other two groups ('Mainland transitivity higher' and 'no transitivity difference'), although the differences in separation frequency between varieties are not very obvious, Mainland is still observed to have higher separation probability (shown in Table 4.15).

Transitivity frequency	ML separation	TW separation		
TW transitivity higher	0.0579	0.0154		
ML transitivity higher	0.0251	0.0188		
No significant difference	0.0076	0.0020		

Table 4.15 Separation differences based on Z-test

Furthermore, it should also be noted that the differences in separation frequency become more obvious when the transitivity differences between the two variants become bigger. As has been presented in Section 4.3.1, likelihood ratio test (likelihood ratio = higher transitivity frequency/lower transitivity frequency) is also conducted to describe the degree of differences. For example, although both 过境 *guojing* 'transit' and 借道 *jiedao* 'channeled through' have shown significant differences in Z-test at P<0.01 level, the likelihood ratio of 过境 *guojing* 'transit' is 10.33, meaning that it is about 10 times more likely for 过境 *guojing* 'transit' to be used transitively in Taiwan than in Mainland Mandarin. While for 借道 *jiedao* 'channeled through', the likelihood ratio is only 1.26.

If the likelihood ratio is larger than 10, the words are considered to have prominent significant difference in transitivity between Taiwan and Mainland Mandarin. For the 8 words which belong to this group, their differences in separation frequency are also prominent.

	TW	ML
Separation usages	43 types, 59 tokens	<b>906</b> types, <b>1808</b> tokens
Transitivity frequency	1.19%	45.75%
Separation frequency	24.5%	0.71%
Separation examples	把好质量关	天津市严 <b>把</b> 进津企业资质审验 <b>关</b>
	ba hao zhiliang guan	tianjinshi yan ba jin jin qiye zizhi
	guard_good_quality_pass	shenyan guan

'guarantee the quality'	Tianjin_strict_guard_enter_Tianjin_e
	nterprise_qualification_verification_p
	ass
	'Tianjin strictly guarantee the
	enterprise qualification'
	把好建设前期工作质量关
	bahao jianshe qianqi gongzuo
	zhiliang guan
	guard_good_construction_preparator
	y_work_quality_pass
	'guarantee the quality of preparatory
	work of construction'
	<b>把</b> 了/过 <b>关</b>
	ba le/guo guan
	guard_LE/GUO_pass
	'guaranteed/have guaranteed'
	把水利建设质量效益关
	ba shuili jianshe zhiliang xiaoyi guan
	guard_water
	conservancy_construction_quality_be
	nefit_pass
	'guarantee the quality and benefit of
	water conservancy construction'
	ba hao ji dao guan
	guard_many_several_CL_pass
	'carefully check for several times'

Table 4.16 Separation comparison of 把关 bagua

One example of 把关 *baguan* 'guarantee' is shown in Table 4.16. The word 把 关 *baguan* 'guarantee' has prominent significant difference in transitivity between Mainland and Taiwan (likelihood ratio: 38.45). The differences in separation usages between Mainland and Taiwan are also very obvious (not only in separation frequency, but also in the grammatical elements can be inserted). The separation frequency of Mainland 把 关 *baguan* 'guarantee' (45.74%) is much higher than that of Taiwan counterpart (1.19%). In addition, very few grammatical elements can be inserted into Taiwan 把关

baguan 'guarantee' (examples like 把好质量关 ba hao zhiliang guan guard\_good\_quality\_pass'guarantee the quality' is frequently appeared in Taiwan corpus), while varieties of elements can be inserted into 把关 baguan 'guarantee' in Mainland Mandarin (e.g., **aspectual marker** 把了/过关 ba *le/guo guan* guard\_LE/GUO\_pass 'guaranteed/have guaranteed'; **classifier** 把 好几道关 ba hao ji dao guan guard\_many\_several\_CL\_pass 'carefully check for several times'; the **object** 把质量关 ba zhiliang guan guard\_quality\_pass'guarantee the quality'; and even the **object with modifier** 把好进津企业资质审验关 ba jin jin qiye zizhi shenyan guan guard\_enter\_Tianjin\_enterprise\_qualification\_verification\_pass 'guarantee the enterprise qualification', etc.).

The words which have prominent significant differences in transitivity frequency between Mainland and Taiwan are also observed to have contrast differences in separation frequency (i.e. separation usages can only be detected in Mainland corpus, while no separation usage can be observed in Taiwan corpus<sup>5</sup>). Several examples are shown in Table 4.17:

	ML examples	TW
撤军	从约旦河撤了军	None
	cong yuedan he che le jun	detected
	from_Jordan_River_Withdrawal_LE_troops	
	'withdrawal troops from the River Jordan'	
联手	需要香港和内地联起手	None
	xuyao xianggang he neidi lianqishou	detected
	need_Hong Kong_and_Mainland_join_up	

<sup>&</sup>lt;sup>5</sup>Again, here is not claiming that for these words, no separation can be found in any context in Taiwan. But as the Gigaword corpus is very large, if no separation usages are found in Taiwan corpus, the relative frequency for separation usages should be very low.

	_hand	
	'need the alliance between Hong Kong and	
	Mainland'	
献计	为改革发展献一计	None
	wei gaige fazhan xianyiji	detected
	for_reformation_development_offer_one_ad	
	vice	
	'offer advice to reformation development'	
移民	移了民	Not detected
	yi le min	
	move_LE_nationality	
	'have immigrated'	

Table 4.17 Examples of contrast difference between Taiwan and Mainland

Generally speaking, variation differences in lexical status/separation ability have been found between Mainland and Taiwan VO compounds. The separation frequencies of Mainland VO compounds are obviously higher than that of Taiwan VO compounds. For the VOs whose transitivity frequencies in Taiwan are significant higher than in Mainland, the separation frequencies in Taiwan are much lower, especially when two varieties have prominent significant differences in transitivity frequency, their differences in separation frequency are also prominent, sometimes even contrast (as shown in Table 4.15, 4.16 and 4.17). Therefore it is possible to argue that the differences in transitivity frequency and separation frequency between Mainland and Taiwan Mandarin actually indicate the different stages that Mainland and Taiwan VO compounds are located in the continuum/process of lexicalization. If the lexicalization status/process is considered as a continuum from phrase to word, then compared to Mainland VO compounds (more likely to be used separately as phrase), Taiwan ones behave more like words instead of phrases, therefore it is more likely for the Taiwan VO sequences to be used in a transitive way.

It should be noted that, the lexical status is not the only factor which can affect the transitivity differences. A variety of factors (both internal linguistic factor and external social factor) may also account for the variations. For example, the word 感恩 *gan'en* 'be thankful' has very large difference in separation frequency between Mainland and Taiwan (in Taiwan: 0.02%; in Mainland: 6.7%). Particularly, 感恩 *gan'en* 'be thankful' is 335 times more likely to be used separately in Mainland than in Taiwan. But the transitivity frequency between the two varieties for 感恩 *gan'en* 'be thankful' does not have significant difference (in Taiwan: 4.2%; in Mainland: 7.5% with a P value = 0.245). That may imply there are some other reasons that are affecting its separation ability or transitivity degree.

## 4.4.6 Summary of Study 2

In this study, a comparable large corpus-based approach is taken to empirically prove that there is a significant correlation between the transitivity of a VO compound and its lexical status (separation ability): compared to separable VOs, inseparable ones are more likely to be used as transitive verbs; moreover, there is a significant negative correlation between transitivity frequency and separation frequency. Based on this, further comparison between the separation frequencies of Mainland Mandarin and Taiwan Mandarin has been conducted. Mainland VO compounds are observed to have higher separation frequency than Taiwan VO compounds, especially when the transitivity frequency of Mainland VO is prominently lower than that of Taiwan VO compound. This makes it possible for us to argue that Mainland and Taiwan VOs are located in different stages in the continuum/process of lexicalization. If the lexicalization is considered as a continuum from phrase to word, Taiwan VO is more close to 'word' compared to Mainland VOs.

Furthermore, in terms of the complexity of inserted elements, Taiwan VO tends to be inserted into more complicate elements. That may account for the separation frequency differences between Mainland Taiwan. Since complicated grammatical elements are not as flexible as the simpler ones, Taiwan VOs are less likely to be used separately.

# 4.5 Evolutionary Mechanism of VO Transitivization and Language Variations – Correlation among Language Alternation, Language Variation and Language Change

#### **4.5.1 Introduction**

As has been mentioned in Section 4.1, there is an increasing number of VO compounds which show the tendency of being transitive (taking another object), and the transitivization of VO compounds has become one of the more prominent ongoing linguistic changes in Modern Mandarin. One of a critical question for VO transitivization is about the transition mechanism. There are many researches focusing on examining the evolutionary mechanism of VO transition, and some of the studies have already been mentioned in Section

4.2.1.

The formation mechanism of  $VO_1+O_2$  has attracted the attention of many Chinese linguistics. The main claims are summarized as following. Zhang (2010) attributes the formation of  $VO_1+O_2$  to the drop of the clitic-like preposition  $\pm yu$  'at/to' following O<sub>1</sub>, which endows the configuration VO<sub>1</sub> with the transitive function, e.g., 驰 誉于中外 chiyu yu zhongwai famous\_at\_home\_and\_abroad 驰 誉 中 外 chiyu zhongwai to famous home and abroad 'be famous at home and abroad'. Yang and Zhou (2006) distinguish three types of  $VO_1+O_2$  construction (Li and Wu, 2017) which I will illustrate in detail later. Li and Wu (2017) have proposed two mechanisms in terms of the syntactic formation of VO: when  $O_1$  is a nominal morpheme, the  $VO_1+O_2$  construction goes through three evolutional phases: in the first stage, V and  $O_1$  are two independent constituents, and can combine to form a VO phrase; at the second stage,  $O_1$  gest incorporated into verb. The transition from stage one to stage two is substantially a process of lexicalization, in the sense that the V and  $O_1$  are condensed into a disyllabic compound in the form VO<sub>1</sub>, which can be used independently to describe an event. During the third stage,  $O_1$  as a bound morpheme gradually goes through a process of semantic bleaching. Accordingly, V's transitivity seems to 'resume work', which naturally makes the transitivity of the entire compound verb  $VO_1$  salient. When  $O_1$  is originally a transitive verbal morpheme (e.g., 解聘 jiepin 'dismiss an employee', 免考 miankao 'exempt from examination'), it is very natural for VO1 to become transitive.

Although numbers of studies have been conducted to examine the evolutionary mechanism of VO transitivization, some crucial issues still remain to be discovered. One of the most crucial issues is that most of the studies categorize VO in a dichotomy way (i.e. VO compound is either transitive or intransitive). Her (1991) has made a tripartite division in terms of the transitivity: VO can be intransitive (得意 *deyi* 'be prond'); semi-transitive (在 行 *zaihang* 'be good at') and transitive (留意 *liuyi* 'pay attention to'). In this study, I argue that the VOs which are considered to be transitive actually are very different in their degree of transitivity. It is more appropriate to view the transitivity of VO in a continuous way (which is also in accordance with the nature of gradient/probabilistic grammar).

In Section 4.1, it has been presented that the rates of transitivity vary across different VO compounds. For example, in Mainland Mandarin, the transitivity degree decreases from 关心 guanxin 'concern'/留意 liuyi 'pay attention to' to 观光 guanguang 'sightseeing'.

4-6.关心他人/留意他的表现 guanxin taren/liuyi ta de biaoxian concern\_someone/pay attention to\_3 SG\_DE\_performance 'be concerned with someone/pay attention to his performance'
4-7.媲美澳洲的黄金海岸 pimei aozhou de huangjian haian rival with\_Australia\_DE \_Gold\_Coast 'rival with the Gold Coast in Australia'
4-8.牵手女嘉宾 qianshou n iijiabin hold\_hand\_female guest 'hold hands with female guest'
4-9. ???观光意大利 guanguang yidali sightseeing\_Italy 'take a sightseeing tour in Italy'

The differences of transitivity are often neglected by previous scholars, while my study takes this as a very important evidence for the investigation of evolutionary mechanism. Because to a large extent (and in most cases), the differences in transitivity actually indicate their degree of transitivity, which further imply their evolutionary stages.

Moreover, from the perspective of language variation, to the best of my knowledge, no study has been conducted to examine the variation differences in transition mechanism across regions. In Study 1 of this chapter, the significant variation differences in degree of transitivity has already been reported (i.e. it has been observed that the rate of transitivity is different between different varieties). For some verbs, they have very prominent different transitivity frequencies in different regions. For example, 移民 yimin 'immigrate' has frequently been used as a transitive verb in Taiwan Mandarin (e.g., 移民澳洲 yimin aozhou immigrate\_Australia 'immigrate to Australia'), with a relatively high transitivity frequency (22.75%), while in Mainland, the transitive usages are rare and the transitivity frequency is very low (0.1%). It is necessary to ask the question, do their transitivity differences actually indicate the variation differences in terms of transition process or transition stages? In other words, can the correlation between language variation and language change be discovered?

Previous papers remain to be largely at the level of description, while

examination from theoretical perspective is still needed. To explore the interaction between language variation and change from theoretical perspective, an important theory should be reffered to is the Lexical Diffusion Theory. Actually, the variations in transitivity rates (within a variety or between different varieties) which have been observed in Study 1, are in accordance with the prediction of Lexical Diffusion Theory (e.g., Wang, 1969, 1977), but no previous account has explained why they differ. Therefore in this study, the evolutionary mechanism of VO transitivization is going to be investigated in both varieties (Mainland and Taiwan Mandarin), based on the theoretical framework of 'Lexical Diffusion Theory'. Based on the theory, the correlation among language variation, language alternation and language change can also be discovered.

## 4.5.2 Lexical Diffusion Theory

#### **4.5.2.1 Introduction to Lexical Diffusion Theory**

The basic idea of Lexical Diffusion Theory has already been reviewed in Section 2.1.2 of Chapter 2. Originally, the lexical diffusion theory is to describe the process of sound change. The essence of lexical diffusion theory is that for sound change, it is phonetically abrupt but lexically gradual. We can understand as each individual word undergoes the change at different rates or at different times. In particular, the process of change is like the following:

During the early phrase of the change, only a small sector of the relevant

morphemes is affected. Some of the affected morphemes may change from X to Y directly. Other morphemes, however, will at first have both the X-pronunciation and Y-pronunciation. But the X-pronunciation will gradually be suppressed in favor of Y-pronunciation (Wang, 1969).

In that sense, Wang (1969) has pointed out the importance of synchronic variation in the study of language change. In sharp contrast with the case of phonetic gradualness, where intermediate forms are all but easy to come by, intermediate forms for lexical diffusion are everywhere, provided we recognize them for what they are. One of the clearest statements of this issue is that by Vogt (1954, p. 367):

"At any moment, between the initiation and conclusion of these changes, we have a state characterized by presence of more or less variants, so that the speakers have the choice between alternative expressions. In each case the choice will be determined by an interplay of factors, some linguistic, some esthetic and social, an interplay so complex that most often the choice will appears as a being due to pure chance... what therefore in a history of linguistic system appears as a change will in a synchronic description appear as a more or less free variation between forms of expression, equally admissible within the system."

To be particular, synchronic variation can serve as a mechanism in the implementation of sound change, whether the actuation was from inside or outside the linguistic system undergoing change. In other words, the synchronic alternation can help us to investigate the transition mechanism and route. Once we recognize the critical role that synchronic variation plays in lexical diffusion, then such evidence is indeed easily found in every living languages (Wang, 1979). The study of language change might more much easier to be conducted. Although it is difficult to collect sizeable historical documentation, at least the synchronic data do give a time-slice of the change in progress.

In Table 4.18, a schematized version of lexical diffusion via synchronic variation is given (originally inWang, 1979). For any  $W_i$ , the notation  $\overline{W}_i$  represents the changed counterpart of that particular word. So a typical change involves three stages: u (unchanged), a (synchronic alternation) and c (changed). In the table,  $W_1$  is the most advanced word, having already reached the c stage.  $W_2$  and  $W_3$  are in the a stage, while  $W_4$  and  $W_5$  are still in the u stage. Although it is obviously important to know whether a change is actuated internally or externally, phonetically or conceptually, the implementation by such a process of lexical diffusion should be the same.

$\backslash$	U	А	С
Stages			
Words			
$W_1$			$\overline{W}_1$
$W_2$		$W_2 \sim \overline{W}_2$	
<b>W</b> <sub>3</sub>		$W_3 \sim \overline{W}_3$	
$W_4$	$\mathbf{W}_4$		
$W_5$	$W_5$		

Table 4.18 A schematized version of lexical diffusion via synchronic variation

As presented above, synchronic alternation (i.e. syntactic variations as well as alternative expressions as mentioned in Vogt, 1954) can serve as a mechanism in the implementation of sound change. Language alternations serve as a kind of psychological bridge between the two end points of a sound change (Wang, Ke and Minett, 2004).

#### **4.5.2.2 Lexical Diffusion Theory in Syntactic Change**

Although the term 'lexical diffusion' is frequently employed in the context of phonology, there has been an increasing awareness in recent researches that the same concept can be applicable to syntactic changes as well (Iyeiri , 2010). There are several papers which examine the syntactic change from the perspective of lexical diffusion (e.g., Yue-Hashimoto, 1993; Zhang, 2000; Cheng, 1990, 1998; Tottie, 1991; Nevalaine, 2006; Ogura, 1993).

Yue-Hashimoto (1993) examines the syntactic change of V-not-V pattern in Pekinese, the Southern Min dialects and the Yue dialects, and finds that the emergence of the V-neg-V(P) form question form follows the route of first appearing with high-frequency verbs, then gradually to other types of verbs. New and old forms are competing with each other and therefore usually coexist for a very long time. Nevalainen (2006) points out in the context of syntactic developments, the fact is that the new syntactic form does not spread to all contexts at once, but some acquire it earlier than others. Tottie (1991) supports the view that (morpho) syntactic change proceeds gradually across the lexicon, and he further argues that the frequency is the determinant of linguistic conservatism, i.e. that the more frequent a construction is, the more likely it is to be retained in its older form for a longer period of time. Ogura (1993) examines the validity of the simultaneous equal activation scenario that Kroch propose (changes occur simultaneously and spread at the same rate in all contexts) based on the development of periphrastic do. The paper claims that changes in the different contexts initiate at different times and the later a change begins, the greater the rate of change becomes. Furthermore, within each context, there is an effect of word frequency.

Based on the previous works, there are several points can be summarized: a) The lexical diffusion theory can be applied to the study of syntactic change, and agreement has been reached by previous scholars that the new syntactic form proceeds gradually across the lexicon. But the number of syntactic studies based on lexical diffusion theory is relatively small. b) Syntactic variation plays a very crucial role in language change. Particularly, syntactic alternation should be considered as a key concept for syntactic change, that is, before the replacement of new syntactic rule, we should expect to observe several sets of constructions with alternations. In other words, the synchronic alternation pattern can provide evidence for the transition route/rate of syntactic change.

#### **4.5.2.3 Lexical Diffusion in Transitivization of VO Compound**

With respect to VO compound, their transition from intransitive to transitive needed to be examined from the perspective of lexical diffusion theory. There are at least two assumptions can be referred to according to the lexical diffusion theory. The first one is the transition from intransitive to transitive verb should apply idiosyncratically to different verbs. The second assumption is before a VO has totally become a transitive verb, we should expect to observe a stage where different alternations exist (new and old syntactic form co-exist). Different alternations are competing with each other for a long time before one of them wins out. The so-called 'competing change' is comprehensive described in Yue-Hashimoto (1993). According to Yue-Hashimoto (1993), on one hand, given the lexicon of a language, the new form will first attack a certain kind of lexicon, competing with the old form and gradually widening its scope of application over the rest of the lexicon. On the other hand, the old form also tries to wage a battle of resistance, competing all along with the newcomer and trying to hold on to its old territory in the lexicon. There is evidently a long period of coexistence of both the old and the new from over the same lexical item; so that both forms are acceptable.

Based on the first assumption, we should start from different words with different transitivity frequencies (i.e. the frequency of being used transitively as in VO<sub>1</sub> + O<sub>2</sub>). The transitivity frequency can to a large extent reflect the transition rate of VO compound. Based on the second assumption, we should also focus on different alternation patterns (i.e. also understand as 'competing patterns') of VO<sub>1</sub> + O<sub>2</sub> construction (i.e. the distribution of different related constructions that a transitive VO compound can occur in). To be specific, according to lexical diffusion theory, the alternation pattern of VO<sub>1</sub> + O<sub>2</sub> (usually the old form before VO becomes transitive) is competing with the VO<sub>1</sub> + O<sub>2</sub> form.

#### 4.5.3 Transition Patterns of VO Compound

In order to investigate the competing mechanism, the alternative pattern (also can be understood as the competing pattern) of  $VO_1 + O_2$  needed to be discovered. Previous studies (e.g., Yang and Zhou, 2006; Cui, 2008; Zhang, 2010) usually classify the transition from intransitive to transitive into three types

types.

From VO<sub>1</sub>+Prep+O<sub>2</sub>to VO<sub>1</sub> O<sub>2</sub>
 The competing alternation: the verb-complement construction with the post-verbal preposition (VC construction),
 e.g., 致力科研 VS.致力于科研
 *zhili keyan* VS.*zhili yu keyan* devote\_research VS. devote\_to\_research
 'devote to research'
 Substant State
 'devote to research'
 'devote to research'

# 2) From **Prep** + $O_2$ + $VO_1$ to $VO_1 O_2$

The competing alternation: the adverbial-verb construction with the preverbal preposition (**AV construction**),

- e.g., 把脉国家经济 VS. <u>为</u>国家经济把脉 bamai guojia jingji VS. wei guojia jingji bamai
- feel the pulse\_country\_economic VS. for\_country\_economic\_feel the pulse 'feel the pulse of economic of the country'

## 3) From $VO_2$ (的) $O_1$ to $VO_1O_2$

The competing alternation: the separation construction e.g., 耗资一百万 VS. 耗一百万的资 hao zi yi baiwan VS. hao yi baiwan de zi cost\_fund\_one\_million VS. cost\_one\_million\_DE\_fund 'cost one million funds'

However, according to my corpus observation, the actual usages are much

more complex than what has been reported by previous scholars. Sometimes

the  $VO_1 + O_2$  pattern has more than one competing alternation. Examples as

sown below:

1) Can have both VC and AV, e.g.,

VC: 媲美于澳洲的黄金海岸 pimei yu aozhou de huangjin haian rival with\_to\_Australia\_DE\_Gold\_Coast AV: 与澳洲的黄金海岸媲美 yu aozhou de huangjin hai'an pimei with\_Australia\_DE\_Golden\_Coast\_rival with 'to rival with the Golden Coast in Australia'

- Can have both AV and separation pattern, e.g., AV: 从广东过境 cong Guangdong guojing from\_Guangdong\_transit Separation: 过广东境 guo Guangdong jing pass\_Guangdong\_boarder 'transit through Gungdong'
- Can have both VC and separation pattern, e.g.,
   VC: 入籍到美国 *ruji dao meiguo* naturalize\_to\_America
   Separation: 入美国籍 *ru meiguo ji* in\_America\_nationality
   'be naturalized into America'
- 4) Can have both AV, VC and separation pattern e.g., VC: 感恩于家人 gan'en yu jiaren be thankful\_to\_family AV: 对家人感恩 dui jiaren gan'en for\_family\_thankful Separation: 感家人的恩 thank\_family\_DE\_gratitude 'be thankful to family'
- 5) No alternative pattern \*获赠于一套书 huozeng yu yi tao shu receive\_to\_one\_series\_book \*对/向/与一套书获赠 dui/xiang/yu yi tao shuhuozeng for/for/with\_one\_series\_book\_receive 'receive one series of book'

## 4.5.4 Research Question

According to the lexical diffusion theory, synchronic alternations serve as an important mechanism in the implantation of language change. The synchronic variations do give us a time-slice of the change in progress and different choices for alternations also help us to examine the evolutionary route/stage of syntactic development. Based on this hypothesis, two research questions needed to be addressed:

1) Statistically and empirically, is there any correlation between the

transitivity rate of VO compound and the alternative patterns (competing patterns)? Are there any variation differences between Mainland and Taiwan?

2) If correlation between transitivity and alternation is significant, then does the observation of synchronic competing alternations indicate the transition route of different VO compounds? What kind of transition route does the data indicate? Any variation differences can be observed?

# 4.5.5 Methodology

#### 4.5.5.1 Data Collection and Annotation

This study focuses on 109 VO compounds collected from previous studies (e.g., Qian, 2011; Luo, 1998; Liu and Li, 1998), which already began to have transitive usages but with different transitivity frequencies. The data for this study was extracted from Annotated Chinese Gigaword corpus (Huang, 2009).

Based on all the extracted data, the relative frequency of transitive usage for each VO has been calculated, as shown in Section 4.3.

In addition to this, for each VO compound, the competing pattern of  $VO_1+O_2$  is also annotated according to the actual usage in corpus (as shown in Table 4.19, there are 8 groups in total and each VO is annotated as the group number it belongs to).

	Alternation		VO example	Alternation example
Group	Both	VC	e.g., 进军 <i>jinjun</i>	AV:向电影界进军 xiang dianyingjie
1	and	AV	'march towards'	<i>jinjun</i> towards_filmdom_march
	alternation			VC:" 进军到电影界 jinjun dao
				<i>dianyingye</i> march_towards_filmdom

			'march towards filmdom'		
Group	Only AV	e.g., 把 关	AV:对产品质量把关 dui chanpin		
2	alternation	baguan	zhiliang		
		'guarantee'	<i>baguan</i> for_product_quality_guarantee		
		-	'guarantee quality of product'		
Group	Only VC	e.g., 驰誉 chiyu	VC:驰誉于中外 chiyu yu zhongwai		
3	alternation	'be famous'	famous_at_home and abroad		
			'be famous at home and abroad'		
Group	No	e.g., 获 赠	获赠一套书 huozeng yi tao shu		
4	alternation	huozeng	receive_one_series_book		
		'receive'	'receive one series of book'		
Group5	Only sep	None detected	None detected		
Group	AV and	e.g., 过 境	AV:从广东过境 cong Guangdong		
6	separation	guojing 'transit'	guojing from_Guangdong_transit		
			Separation: 过广东境 guo		
			Guangdong jing		
			pass_Guangdong_boarder		
			'transit Gungdong'		
Group	VC and	e.g., 入籍 <i>ruji</i>	VC:入籍到美国 ruji dao		
7	separation	'be naturalized'	meiguonaturalize_to_America		
			<b>Separation</b> :入美国籍 ru meiguo ji		
			in_America_nationality		
			'be naturalized into America'		
Group	All	e.g., 致信 zhixin	AV:给青少年致信 gei qingshaonian		
8	alternations	'write to'	<i>zhixin</i> to_teenager_write		
			Separation: 致 青 少 年 信 zh		
			qingshaonian xin to_teenager_letter		
			<b>VC</b> : 致信给青少年 zhixin gei		
			qingshaonian write_to_teenager		
			'write to teenager'		

Table 4.19 Type of alternations

The annotation result shows that no VO belongs to group 5 and there is only 1 word (i.e. 入籍 *ruji* 'be naturalized') in group 7 (does not meet the requirement for statistical analysis), therefore these two groups were excluded from further analyses.

## **4.5.5.2 Statistical Tests and Tool**

Kruskal- Wallis test is used to study the correlation between transitivity frequency and the alternative patterns and the statistical tool used here is IBM SPSS V.22.

## 4.5.6 Results and Discussions

To answer the 1<sup>st</sup> question, statistically and empirically, is there any correlation between the transitivity rate of VO compounds and the alternative patterns, Kruskal – Wallis test is used in our study. Kruskal-Wallis test is often presented as non-parametric alternative to One-way ANOVA. As it does require the normality of the data, it is very suitable for my current study. The statistical tool used here is IBM SPSS V.22. Particularly, in this study, Kruskal-Wallis test is conducted to compare whether there are any statistical significant differences between the transitivity frequencies of these 6 groups in each variety.

# 4.6.6.1 Result of Taiwan Data

Let us discuss the result of Taiwan data first. The comparison in Taiwan is conducted between group 1 to 8: Group 1 (n=33), Group 2 (n=32), Group 3 (n=21), Group 4 (n=9), Group 6 (n=6), Group 8 (n=7). Distributions of transitivity frequency of VO compounds are not very similar for all groups, as assessed by visual inspection of a boxplot.

According to the result of Kruskal – Wallis test shown in Table 4.20, there

are statistical significant differences in transitivity frequency between groups that differed in their alternative patterns,  $\chi^2(3) = 36.236$ , p<0.001. Subsequently, pairwise comparisons were performed using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons. Adjusted P-values are presented. The post hoc analysis (shown in Table 4.21) indicated that the transitivity frequency of Group 1 (both AV and VC construction) (mean rank=56.61) is significantly higher than that of Group 2 (only AV construction) (mean rank=31.72) (p=0.02). But the transitivity of Group 3 (only VC construction) (mean rank=82.29) is significantly higher than Group 1 (p=0.05). Correspondingly, the relative transitive frequency of Group 2 is significant lower than that of Group 3 (p<0.005). The result also reveals there is no significant difference between Group 1 and the other three groups: Group 4 (no alternative pattern) (mean rank=59.33) (p=1.000), Group 6 (mean rank=42.67) (p=1.000) as well as Group 8 (mean rank=69.29) (p=1.000). There is also no significant difference between the frequency of Group 3 and that of Group 4 (p=0.988), and Group 8 (p=1.000). It is noted that the transitivity frequency of Group 4, 6, 8 has no significant difference with all the other 5 groups, respectively.

column	Result
Sig.	8.5178093303358E-7
Decision	Reject the null hypothesis
Null	The distribution of Taiwan transitivity is the same across
hypothesis	categories of alternation
Test	Independent-Sample Kruskal-Wallis Test

Total N	108
Test Statistic	36.236
Df	5
Asymptotic Sig.	.000
(2-sided test)	

Table 4.20 The result of Kruskal-Wallis test for 8 groups in Taiwan

Sample 1-	Test	Std. Error	Std.	Sig.	Adj. Sig
Sample 2	Statistic		Statistic		
type2-type 6	-10.948	13.934	786	.432	1.000
type 2-type1	24.887	7.771	3.203	.001	.020
type 2-type 4	-27.615	11.818	-2.337	.019	.292
type 2-type 8	-37.567	13.069	-2.875	.004	.061
type 2-type 3	-50.567	8.796	-5.749	.000	.000
type 6-type 1	13.939	13.901	1.003	.316	1.000
type 6-type 4	16.667	16.507	1.010	.313	1.000
type 6-type 8	-26.619	17.425	-1.528	.127	1.000
type 6-type 3	39.619	14.400	2.733	.006	.094
type 1-type 4	-2.727	11.778	232	.817	1.000
type 1-type 8	-12.680	13.033	973	.331	1.000
type 1-type 3	-25.680	8.743	-2.937	.003	.050
type 4-type 8	-9.952	15.784	631	.528	1.000
type 4-type 3	22.952	12.478	1.839	.066	.988
type 8-type 3	13.000	13.670	.951	.342	1.000

Table 4.21 The post hoc comparison for 8 groups in Taiwan

In summary, transitivity frequency is observed to have the following ranking: Group 3>Group 1> Group 2. There is no significance in any other group combination. Therefore it seems that the main difference does not lie in these three groups (Group 1, 2 and 3).

To verify this result, another test was conducted based on Group 4, 6 and 8. The result (shown in Table 4.22) reveals that there is no significant difference in transitivity frequency between these three groups,  $\chi^2(3) = 2.319$ , p=0.314. In addition, the numbers of words for these three groups are also small. Since Group 4, 6 and 8 do not contribute to the significant correlation coefficient, they are excluded from further analysis.

Column	Result			
Sig.	0.313667964687823			
Decision	Accept Null Hypothesis			
Null	The distribution of Taiwan transitivity is the same			
hypothesis	across categories of alternation			
Test	Independent-Sample Kruskal-Wallis Test			

Total N	22
Test Statistic	2.319
Df	2
Asymptotic Sig.	.314
(2-sided test)	

Table 4.22 The result of Kruskal-Wallis test for 468 groups in Taiwan

As shown in the Kruskal-Wallis result of Group 1, 2 and 3 (in Table 4.23), the transitivity frequency is statistically different between different alternation pattern groups,  $\chi^2(3) = 32.886$ , p<0.001. Post hoc analyses (in Table 4.24) using Dunn's (1964) procedure with a Bonferroni correction for multiple comparisons indicate that the transitivity frequency of Group 2 is significantly lower than both Group 1 (p=0.004) and Group 3 (p<0.001), which is consistent with the result shown above. Futhermore, the transitivity frequency of Group 3 is significantly higher than Group 1 (p=0.011).

column	Result			
Sig.	7.22565119959384E-8			
Decision	Reject the null hypothesis			
Null	The distribution of Taiwan transitivity is the same across			
hypothesis	categories of alternation			
Test	Independent-Sample Kruskal-Wallis Test			

Total N	86
Test Statistic	32.886
Df	2
Asymptotic Sig.	.000
(2-sided test)	

Table 4.23 The result of Kruskal-Wallis test for 123 groups in Taiwan

Sample 1-	Test	Std. Error	Std.	Sig.	Adj. Sig
Sample 2	Statistic		Statistic		
type 2-type	19.722	6.195	3.183	.001	.004
1					
type 2-	-39.908	7.012	-5.691	.000	.000
type3					
type 1-type	-20.186	6.970	-2.896	.004	.011
3					

Table 4.24 The post hoc comparison for Group123 in Taiwan

# 4.5.6.2 Result of Mainland Data

The similar result has been found for Mainland data. As for Group 4, 6 and 8, there is also no significance in any group combination (shown in Table 4.25), therefore I only report the Kruskal-Wallis result of Group 1, 2 and 3. As shown in Table 4.26, the transitivity frequency of Group 2 (mean rank=28.70) is significantly lower than that of Group 1 (mean rank=47.7) (p=0.006) and Group 3 (mean rank=69.45) (p<0.001). But different from Taiwan data, there is no significance difference in transitivity frequency between Group 1 and Group 3 (p=0.275). The transitivity frequency in Mainland ranks as: Group 1> Group 2; Group 3> Group 2.

Column	Result		
Sig.	0.356060097189569		
Decision	Accept Null Hypothesis		
Null	The distribution of Mainland transitivity is the same		
hypothesis	across categories of alternation		
Test	Independent-Sample Kruskal-Wallis Test		

Total N	22
Test Statistic	2.065
Df	2
Asymptotic Sig.	.356
(2-sided test)	

Table 4.25 The result of Kruskal-Wallis test for 468 groups in Mainland

Sample 1-	Test	Std. Error	Std.	Sig.	Adj. Sig
Sample 2	Statistic		Statistic		
type 2-type	18.994	6.193	3.067	.002	.006
1					
type 2-	-30.749	7.011	-4.386	.000	.000
type3					
type 1-type	-11.755	6.968	-1.687	.092	.275
3					

Table 4.26 Post hoc comparison for 123 groups in Mainland

# **4.5.6.3** Line Chart to Present Tendency – Variation Difference

The line chart is also used to present the transition tendency. As both Group 2 and Group 3 only contain 1 alternation pattern, and the tendency difference between these two groups are significant in both varieties, the trend line of Group 2 and 3 are first shown. As shown in Figure 4.3, the X-axis is the transitivity frequency and the Y-axis stands for the frequency/probability of occurrence for each alternation group.





Figure 4.3 Line chart with log for Group 2 and 3 in both varieties

According to the figure, it is shown very obviously that for both Mainland and Taiwan, the VC construction presents a positive correlation with the transitivity frequency of VO compounds, meaning that the VO with high transitivity frequency prefers to have VC construction as the alternation, while VO with low transitivity frequency dis-prefers VC construction as the alternation. In contrast, the AV structure presents the opposite tendency. When a VO has low transitivity frequency, it is more likely that the AV construction is the alternation. It seems that there is no obvious difference between these two varieties, especially when we compare the trend line ( 対数 logarithmic), the tendencies of Group 2 and 3 in both varieties are almost the same.

It is not surprising that no variation difference between Mainland and Taiwan is detected. In contrast, this result actually confirms the status of Mainland and Taiwan Mandarin as two variables of the same language, instead of two independent languages. This also confirms our hypothesis to be reliable and stable, since the same tendency has been shown in different variables.

However, the variation differences still can be observed, mainly in Group 1. According to the statistical result, the transitivity frequency in Taiwan ranks as: Group 3>Group 1> Group 2, while in Mainland there is no transitivity difference found between Group 1 and 3. Furthermore, as the line chart of Group 1 (in Figure 4.4) reveals that although the general tendencies (presented by trend line presented logarithmic) of Group 1 in Mainland and Taiwan are very similar (a weak negative correlation, the probability decreases when the transitivity gets higher), some subtle variation differences can still be detected. To be particular, in Taiwan, the peak of the trend line is in the middle (when the transitivity frequency is around 20%), which indicates that in Taiwan when the transitivity frequency is around 20%, the VO is more likely to have both AV and VC alternations (belong to Group 1). While in Mainland, the VO which has a very high transitivity (more than 50%) tends to have both AV and VC alternations.





Figure 4.4 Line chart with log for Group 1 in both varieties

The first question people may ask is why the tendencies of Group 2 and 3 are shared by different varieties, while variation differences are observed in Group 1. This can be explained/be accounted for by the fact that compared to Group 2 and 3, Group 1 is more complex in terms of containing more than 2 alternative patterns. The preference for each alternative pattern would vary across different VOs, or across different variants. The variation differences in Group 1 may be related to the preference for each alternative pattern (whether the VO prefers AV or VC). Details will be discussed in Section 4.5.8.

# 4.5.7 The Evolutionary Route for VO Compound

Based on the result of statistical analysis, it is possible to answer the 2<sup>nd</sup> research question: what does the observation of synchronic alternation indicate? As mentioned above, both Mainland and Taiwan Mandarin share the same evolutionary pattern for both Group 1 and 2, meaning that the evolutionary route for these two groups are quite convicing and reliable.

## **4.5.7.1 For VC Alternation:**

It is not difficult to explain why high transitivity VO is highly correlated with VC construction. Zhang (2010) has already proposed that the transfer from intransitive to transitive VO is the result of dropping of the post-verbal preposition  $\mp yu$  'to'. Other previous studies also have similar arguments (e.g., Yang and Zhou, 2006). According to them, examples like 致力科研 zhili keyan devote research 'devote to research' is transited from 致力于科研 zhili yu keyan devote to research 'devote to research' which dropping the post-verbal preposition 于 yu 'to'. This type also includes other VO compounds like 取悦 quvue 'please someone', 投身 toushen 'plunge into', 取信 quxin 'win the trust', 插足 chazu 'participate', 驰名 chiming 'be famous', 放眼 fangyan 'look ahead', 迁怒 qiannu 'vent on', 移情 yiqing 'love someone else', 移民 yimin 'immigrate', 转行 zhuanhang 'change career', 启航 gihang 'set sail' and so on. Because of the grammaticalization and meaning bleaching of the post-verbal preposition  $\pm yu$ , the dropping of  $\pm yu$  'to' does not affect the meaning of the whole construction, thus makes the transition possible. The motivation of transition is the rhythmic constraints as well as the economic principle.

The transition process proposed by Zhang (2010) is shown in the following examples (examples are cited from Zhang (2010)).

4-10. 尝有/恩/于/晟 (cited from Jiu Tangshu Volume 6, 133: 旧唐书卷 133, 列 传第 83: collected biographies, No. 83) *chang you en yu cheng* once\_have\_kindness\_to\_Cheng 'once be kind to Cheng' 4-11. 子敬有恩/于/玄德 (cited from The Romance of the Three Kingdoms Chapter. 58 三国演义 58 回)

Zijing youen yu Xuande Zijing\_have\_kindness\_to\_Xuande 'Xuande feels indebted to Zijing'

4-12. 农民有恩于/国家, 有恩于/中华民族 nongmin youen yu guojia, youen yu zhonghua minzu farmer\_have\_kindness\_to\_country, have\_kindness\_to\_the\_Chinese\_Nation 'The country and the Chinese Nation feel indebted to farmers.'
4-13. 老金有恩/中国

> Laojin youen zhongguo Laojin\_have\_kindness\_China 'China feels indebted to Laojin'

These examples show the transition path of 有恩 you'en 'be kind' from intransitive to transitive compound. At the first stage (4-10), 有恩于晟 you en yu cheng have\_kindness\_to\_Cheng 'be kind to Cheng', each character should be segmented as different units (有/恩/于/晟). At the second stage (4-11), 子敬 有恩/于/玄德 Zijing youen yu Xuande Zijing\_have\_kindness\_to\_Xuande "Xuande feels indebted to Zijing', 有恩 you'en 'be kind' has already been lexicalized as a word and followed by the whole complement 于玄德 yu Xuande 'to Xuande'. At the third stage (4-12), although  $\mp$  yu 'to' is still served as a preposition and remain its semantic meaning, 有恩于 you'en yu 'be kind to' has already been combined as one segmentation unit. At the fourth stage (4-13), the preposition  $\mp$  yu 'to' has become semantically bleached and finally dropped. The VO compound 有恩 you'en 'be kind' has the potential to transit from the intransitive to transitive verbs.

The transition from 1<sup>st</sup> to 2<sup>nd</sup> stage is mainly motivated by the

disyllabification tendency in Chinese. As the tendency is dominant in Chinese language change, lexicalization process (i.e. from phrase 有/恩 you en 'have kindness' to word 有恩 you'en 'be kind') has very strong motivation. The transition from the  $2^{nd}$  to the  $3^{rd}$  stage may mainly be motivated by the psychological and cognitive reason (Zhang, 2010). The preposition  $\pm vu$  'to' and the host 有恩 you'en 'be kind' originally belong to two constituents. Because of analogy effect, cognitively forming as a chuck is much easier for memorizing and learning process. The transition from the third stage (有恩于国 家 youen yu guojia have kindness to country 'the country feels indebted to') to the fourth stage (有恩中国 youen zhongguo have kindness China 'China feels indebted to') may be motivatied by rhythmic constraints as well as the economic principle. In terms of the rhythmic constrains, as the disyllabic  $VO_1$ and the monosyllabic  $\mp yu$  'to' form a hypermetrical syllable. It is known that Modern Mandarin words have a strong tendency of being disyllabic. 有恩于 *you'en yu* 'be kind to' as a tri-syllabic word is not very stable. Once  $\mp$  yu 'to' has become semantically bleached, it has a very high probability of being omitted (to become disyllabic). In addition, the economic principle is also affecting the transition at this stage, as the expression 有恩国家 youen guojia have kindness country 'the country feels indebted to' is obviously more economical than 有恩于国家 youen yu guojia have kindness to country 'the country feels indebted to', the post-verbal preposition is easily to be dropped when it became highly grammaticalized. The transition from stage 2 to stage 4 can be summarized in Figure 4.5.



Figure 4.5 The transition process for Group 3

Therefore we can see that the transition from  $VO_1+\mp+O_2$  to  $VO_1+O_2$  is motivated by several different factors, including some dominant factors in Chinese languages. Furthermore, there is only one step in the transition process (i.e. dropping of preposition). In other words, the dominant motivation and the simple step/procedure may account for the high transitivity rates that VO compounds of Group 3 obtain.

# 4.5.7.2 For AV Alternation:

Another alternation has shown an exactly opposite tendency: the VOs which have low transitivity frequency are highly correlated with AV construction. In other words, the AV construction shows a negative correlation with transitivity frequency. Example like 把脉国家经济 *bamai guojia jingji* feel the pulse\_country\_economic 'feel the pulse of economic of the country' has the AV alternation as in 为国家经济把脉 *wei guojia jingji bamai* for\_country\_economic\_feel the pulse 'feel the pulse of economic of the country'. Other VOs such as 动员 *dongyuan* 'mobilize', 中意 *zhongyi* 'like', 挑 战 *tiaozhan* 'challenge', 把关 *baguan* 'guarantee' and so on. The opposite tendency may be because, compared to the transition from VC construction to  $VO_1O_2$  construction, changing form AV construction to  $VO_1O_2$  are much more complex. On one hand, it requires more steps for transition (at least two steps: dropping and shifting). On the other hand more effort are also required (shifting takes much more effort than dropping). For example, from 与唱片公司签约 *yu changpian gongsi qianyue* with\_ record\_company\_sign\_contract 'sign contract with record company' to 签约唱片公司 *qianyue changpian gongsi* sign\_contract\_record\_company 'sign contract with record company', at least two steps are involved: one is the dropping of the pre-verbal preposition 与 *yu* 'with' and then the whole NP 唱片公司 *changpian gongsi* company' are transferred to follow the verb (shown in Figure 4.6). Therefore, the transition rate might be much lower compared with that from VC construction to  $VO_1O_2$ .



In summary, as the transition is much more complex and there is no strong motivation for the change, the transition rate for this group is significantly lower than that of Group 3.

## 4.5.7.3 No Alternative Pattern

In addition, there are several VO<sub>1</sub>+O<sub>2</sub> constructions which are obserived to have
no alternative pattern. These constructions are usually formed by two types of VOs. The first type contains words like 获赠 huozeng 'receive', 参选 canxuan 'participate in election', 抱怨 baoyuan 'complain', 投近 tousu 'complain', 参 演 canyan 'participate acting'. Different from most of VO compounds, O in these VO is a verbal morpheme. 赠 zeng 'give away', 选 xuan 'elect', 诉 su 'tell' and 演 yan 'act' are already transitive, therefore it is very natural for them to take another external object. Another type includes words like 出席 chuxi 'attend' and 接手 jieshou 'take over'. For these words, the internal objects 席 xi 'seat' and 手 shou 'hand' are used metaphorically, meaning that the O<sub>1</sub> is highly incorporated into the verb. Therefore it is easy for the VO to become transitive.

#### 4.5.7.4 Both AV and VC Alternation:

As presented above, variation differences have been observed mainly in Group 1. According to the Kruskal –Wallis result, the transitivity frequency in Taiwan ranks as: Group 3>Group 1> Group 2, while no transitivity difference between Group 1 and 3 has been found in Mainland. One thing should be noted is that, although there is no statistical significant difference between these two groups in Mainland (based on the Kruskal-Wallis test), the transitivity frequency of Group 3 is still observed to be obviously higher than that of Group 1 (according to the comparison of mean, mean rank and median), which is shown in Table 4.27.

Mainland	Mean	Mean rank	Median
Group 1	0.351706108	47.7	0.298246
Group 3	0.466447201	69.45	0.5238

Table 4.27 Comparison of mean, mean rank and median between Group 1 and 3

In other words, for both varieties, the transitivity of these three groups ranks as: Group 3>Group 1>Group 2. The transitivity of VO which has both AV and VC constructions (Group 1) is lower than of VO which only has VC (Group 3) constructions, but higher than the transitivity of VO which only has AV constructions (Group 2). It can be explained as, since the VO of Group 1 has both AV and VC constructions, the middle transitivity is the result of neutralization of the high transition rate of VC construction and the low transitivity rate of AV construction. The transition route of words in Group 1 is more complex than that of in Group 2 and 3. VO in this group can be transited from either AV or VC constructions, or probably both. This is related to the 'Multilevel Competition', which will be proposed and discussed in detail in the following Section 4.5.8.

#### **4.5.8** Variation Differences in Group 1 – Multilevel Competition Model

As presented by both Kruskal-Wallis test and Line chart, variation differences exist in Group 1. According to the line chart (Figure 4.4a), the peak of the Taiwan trend line of Group 1 is in the middle (when the transitivity frequency is around 20%). It means that in Taiwan when the transitivity frequency is around 20%, the VO is more likely to have both AV and VC alternations (belong to Group 1). In Mainland, the highest point of probability of Group 1 is when the transitivity is very high (more than 50%), meaning that when a VO has a very high transitivity, it is more likely to have both AV and VC alternations. In other words, the VO with mid-transitivity frequency in Taiwan prefers to have both AV and VC construction as its alternation. While in Mainland, the VO with very high transitivity frequency may prefer to have both alternations. The other difference is that the trend line of Taiwan shows an increase first and then decreases after it reaches the peak, while in Mainland, the trend line falls first and then increases slightly.

My basic assumption is that, the variation differences presented above may be related to the 'Multilevel Competition Model'. The lexical diffusion theory emphasizes that in the process of language change, there always exists competition between the new forms and the old forms. On one hand, given the lexicon of a language, the new form will first attack a certain kind of lexicon, competing with the old form and gradually widening its scope of application over the rest of the lexicon. On the other hand, the old pattern also tries to wage a battle of resistance, competing all along with the newcomer and trying to hold on to its old territory in the lexicon. This is called 'bi-directional competition' (Yue-Hashimoto, 1993). For the situation in Group 2 and 3, there is only 1 level of competition. Only the new form  $(VO_1+O_2 \text{ construction})$  and the old form (its alternation: either AV or VC construction) are involved in the competition, therefore there is no obvious variational difference in evolutionary mechanism for these two groups. While for Group 1, since there are two old forms (both AV and VC constructions) for each word, another level of competition might exist

in the process of transition – the competition between different old forms (i.e. AV and VC). The 'Multilevel Competition Model' is shown in Figure 4.7. As can be seen, besides the external competition between old forms and new forms, there is also another level of internal competition within old forms (between the two different old forms: AV and VC construction). The two levels of competition makes the evolution of words in Group 1 much more complicated, hence variation difference also emerges.



Figure 4.7 Multilevel Competition Model

Based on the 'Multilevel Competition Model', I further propose that the variation differences in Group 1 may due to the internal competition. That is, the variation differences may due to the differences in preference of alternative patterns (the competition between the two old forms). In other words, since each VO compound in Group 1 has both VC and AV constructions, different VOs or VOs across different regions may reveal differences in their preference of alternation. Some VO in Group 1 may prefer to have AV rather than VC (e.g., 媲美 *pimei* 'rival with' prefers 与..... 媲美 *yu* ... *pimei* with\_rival 'rival with'

rather than 媲美于 *pimei yu* rival\_at 'rival with'), while some VO may have the opposite preference (e.g., 失信 *shixin* 'break promise' prefers 失信于人民 *shixin yu renmin* break\_promise\_to\_people 'lose credibility to people' rather than 对人民失信 *dui renmin shixin* to\_people\_break\_promise 'lose credibility to people'). Moreover, for each VO compound, their preference may also vary across varieties, i.e., the same VO prefers AV rather than VC in Mainland, while it shows the opposite preference in Taiwan (e.g., 求助 *qiuzhu* 'ask for help' in Taiwan prefers AV 向政府求助 *xiang zhengshu qiuzhu* to\_government\_ask for help 'turn to government' while in Mainland it shows preference for VC 求助 于政府 *qiuzhu yu zhengfu* ask for help\_to\_government 'turn to government'). In that sense, the preference of alternation should be included to account for the variational differences in Group 1.

To measure the preference of alternation, relative frequency is used: Relative Frequency = usages for each alternation/all usages. For example, 媲美 *pimei* 'rival with' in Taiwan Gigaword Corpus has 204 usages in total. Among them, 55 tokes are used in AV construction, therefore the relative frequency for AV construction is 26.96% (=55/204) while only 1 媲美 *pimei* 'rival with' in Taiwan is used in VC construction, which leads relative frequency for VC construction to 0.49% (=1/204). It is obviously that 媲美 *pimei* 'rival with' in Taiwan prefers AV construction rather than VC construction.

In this section, all the 33 words of Group 1 in both Mainland and Taiwan are annotated with calculated relative frequency. Taiwan data is first here.

	>20%	5-20%	<5%
AV	0.133393287	0.287854938	0.063994526
preference			
VC	0.0627	0.074373545	0.128011169
preference			

Table 4.28 Alternation preference for words of Group 1 in Taiwan

As shown in Table 4.28, in general, in Taiwan Mandarin, VO with high transitivity (larger than 50%) tends to prefer AV rather than VC, while VO with low (smaller than 5%) transitivity tends to prefer VC rather than AV. This can explain why VO in Group 1 has such a tendency presented in line chart (shown in Figure 4.8): as high transitivity VOs (i.e. which have transitivity larger than 20%) prefer AV, therefore for these VO compounds, AV instead of VC construction is playing a more important role in the process of transition, hence it presents a same tendency with Group 2: AV alternation (i.e. a negative correlation between transitivity and alternation); when VO transitivity is smaller 5%, VC instead of AV construction is playing a critical role in the process of transition, the tendency line is the same with Group 3 (i.e. a positive correlation between transitivity and alternation). To sum up, the tendency line of Group 1 fluctuates due to the competition between AV and VC: When the VO in Group 1 prefers AV over VC, the evolutionary route of AV will be leading the transition; while when VO prefers VC, the evolutionary route of VC will play a more important role in the process of transition.



Figure 4.8 Line chart with log for Group 1 in Taiwan

In contrast, VO compounds of Group 1 in Mainland show the opposite tendency in preference. As shown in Table 4.29, in Mainland, , low transitivity VO (i.e. when transitivity is smaller than 5%) tends to prefer AV construction (instead of preferring VC as Taiwan VO compounds do). In other words, AV is leading the transition process. Hence, when the transitivity is smaller than 5%, Figure 4.9 shows a negative correlation between transitivity and alternation (as the same with Group 2).

Alternation	Preference frequency	
	when transitivity <5%	
AV	0.36142	
VC	0.0041	

Table 4.29 Alternation preference for words of Group 1 in Mainland (transitivity <5%)

When transitivity is larger than 5%, more VO compounds obviously prefer VC in Mainland, with relatively higher frequency, compared to Taiwan Mandarin. For example, 立足 *lizu* 'base upon': 0.2549; 定都 *dingdu* 'establish a capital': 0.184; 出身 *chushen* 'originally birth from': 0.1636; 受聘 *shoupin* 'be hired by': 0.17149; 迁居 *qianju* 'mvoe': 0.175; 求助 *qiuzhu* 'ask for help': 0.145488 all have relatively high VC frequency. In that sense, the VC construction is playing a more important role in the process of transition. Therefore, when the transitivity is larger than 5%, it presents a positive correlation between alternation and transitivity (the same as Group 3).



Figure 4.9 Line chart with log for Group 1 in Mainland

As can be seen, the variation difference in Group 1 mainly lies in the differences in alternation preferences. The internal competition between different old forms can also affect the transition process. To be particular, for the VO compounds in Group 1, low transitivity words (smaller than 5%) tend to prefer VC in Taiwan Mandarin, and show preference for AV alternation in Mainland Mandarin. Furthermore, higher transitivity words (larger than 5%) tend to prefer AV in Taiwan Mandarin, and are more likely to be appeared in VC construction in Mainland. Again, the variation differences between different varieties for Group 1 also confirm our hypothesis that it is the competing alternation that is actually affecting the transition rate of VO compounds from intransitive to transitive. The variation differences are also due to the different preference of alternations (competition between old forms).

#### **4.5.9** Other Factors that May Affect the Process of Change

Although this study has shown that the competing alternation is a very important factor, it is not the only factor that can influence the transition rate of a VO compound. Other factors (including both linguistic internal and external factors) may also motivate the transition process. In terms of internal factors, one factor that should be mentioned is word frequency. Word frequency is considered as a determinant linguistic factor in language development (e.g., Tottie, 1991). The impact of word frequency will be discussed in detail in Study 4.

Besides the word frequency, some VO compounds are observed to be frequently occurred in a fixed context such as 驰誉中外 *chiyu zhongwai* 'be famous at home and abroad'/驰名海内外 *chiming haineiwai* 'be famous at home and abroad'. The formation of a chunk/fixed phrases actually allow the transitivization of VO compounds to progress at a much faster rate. Therefore the transitivity frequencies of these VO tend to be very high in both varieties (e.g., 100% for 驰誉 *chiyu* 'be famous' and 77.12% for 驰名 in Taiwan).

In addition to the linguistic internal factors, some linguistic external factors can also affect the transition rate of VO compound. For example, the influence of other languages or dialects may increase the transition rate. The lexical word 中意 *zhongyi* 'like' in Mainland Mandarin is actually borrowed from Cantonese, therefore the usages between Mainland Mandarin and Cantonese are very similar (e.g., 中意这个男生 *zhongyi zhege nansheng* like\_this\_boy 'like this boy'). The influence of social media is also very

important. One typical word is 牵手 *qianshou* 'join hands', constructions like 牵手女嘉宾 *qianshou nijiabin* hold\_hand\_female guest 'hold hands with female guest' are frequently used in a popular Chinese TV dating show and this may to a large extent increase the productivity and popularity of the transitive usage of 牵手 *qianshou* 'hold hands' in Mainland Mandarin.

#### **4.5.10 Summary**

In this study, one important idea has been proposed is that the most important factor which influences the transition rate of a VO compound is its competing alternation.

Based on lexical diffusion theory, the correlation between transitivity frequency and competing alternation has been investigated in this section. The observation of synchronic competing alternations also shed light on the transition mechanism of VO compounds from intransitive to transitive words. To be particular, the VOs which have VC constructions as their competing alternations tend to have high transitivity frequency, due to the strong motivation as well as the simple step; the VOs with AV competing constructions as their alternations tend to have low transitivity frequency, probably because the motivation is weak while the effort for changing is much. For VOs that have both AV and VC construction, another level of internal competition is also involved in the process of change ('Multilevel Competition Model'), the competition between AV and VC construction may to a large extent influence the transition rate: When the VO prefers AV over VC, the evolutionary route of AV will be leading the transition; while when VO prefers VC, the evolutionary route of VC will play a more important role in the process of transition.

In addition, the correlation among language alternation, language variation and language change has also been discussed. According to Lexical Diffusion Theory, language alternation can help us to investigate the mechanism in the implantation of language change. As shown in Section 4.5.8, language variation mainly lies in the differences in alternation preferences. Therefore, in this study, by exploring the differences in alternation preferences between Mainland and Taiwan, their variation differences in transition route are also examined.

# 4.6 Study 4: The Correlation between Transitivity and Word Frequency

#### **4.6.1 Introduction**

In study 3, I have proposed that the most important factor which influences the transition rate of a VO compound is its competing alternation. Another important linguistic internal factor should not be neglected is the word frequency. As studied in most of the literatures in Lexical Diffusion Theory, word frequency usually plays a critical role in determining the changing order in the process of language change (e.g., Phillips, 1984; Tottie, 1991).

#### **4.6.2 Literature Review**

Schuchardt (1885) has claimed that the frequency of words have a prominent influence on phonetic transformation, not within small differences, but within significant ones (Phillips, 1984). Most studies under Lexical Diffusion Model have indeed supported this hypothesis. Two directions for lexical diffusion with regard to frequency have been observed by previous studies, in terms of both sound and syntactic change: changes affecting high-frequency words earlier and changes affectinng low-frequency words earlier.

#### **4.6.2.1 For Sound Change**

#### **4.6.2.1.1** Changes Affecting High-Frequency Words Earlier

The notion that more frequent words will change earlier has been around for a while (S. Y. Wang, 1979). Schuchardt (1885) has already pointed out that high-frequency words are affected by sound change earlier than low-frequency words.

Such tendency has been confirmed in a broad range of changes, in various researches. For example, In terms of the /t/ and /d/ deletion in American English, Losiewicz (1992) and Bybee (2000) has shown that the deletion occurred more in high-frequency words (Bybee, 2002). Similar results were obtained by Jurafsky, Bell, Gregory, and Raymond (2001), Gregory, Raymond, Bell, Fosler-Lussier and Jurafsky (1999). For Spanish [ð] deletion, Bybee (2002) has shown that the rate of deletion is higher among high-frequency words. In terms of

vowel shifts in American English, Labov (1994: 506-507) noted that, when word-initial short [x] occurs before a voiceless fricative, only the more common, monosyllabic words are tensed. Moonwoman (1992) shows some evidence that high-frequency words undergo vowel shifts before low-frequency words. In addition to consonant reduction, another type of change that shows robust word frequency effects is vowel reduction and deletion (Bybee, 2002). Fidelholtz (1975:200-201) pointed out the essential difference between words that reduce a pre stress vowel and phonetically similar words that do not, is word frequency. Van Bergem (1995) claimed the reduction of a pre stress vowel in Dutch is also highly conditioned by frequency: the high-frequency words are more likely to have a schwa in the first syllable than the phonetically similar low-frequency words. Deletion of reduced vowels may also be conditioned by word frequency. Hopper and Bybee (1976) said the deletion is less likely to have in the low frequency word. Phillips (1980) has observed that word frequency is clearly one of the crucial determinants in the raising of OE /a/ to /o/ before nasals: the most frequent words are more likely to be affected. In addition, the formation of new diphthongs in early Middle English through the assimilation of the glide to the vocalicity of the preceding vowels has also been shown to have affected the most frequent words first (Phillips, 1983).

#### 4.6.2.1.2 Changes Affecting Low-Frequency Words Earlier

While many studies have demonstrated the tendency that changes tend to occur earlier in words and phrases of high frequency, some researches have observed the opposite tendency. Hopper and Bybee (1976) have noted a lexical diffusion paradox: sound change seems to affect high-frequency words first, but analogical change affects low-frequency words first. Phillips (1984) showed that some sound changes can proceed from low frequency to high-frequency words. He then compares the two directions for lexical diffusion with regard to frequency (affecting high-frequency words earlier and low-frequency words earlier) and pointed out that none of the sound changes that have affected the least frequent words first (e.g., Glide deletion in Southern American English, unrounding in Middle English, and diatone formation in Modern English) was physiologically motivated, which leads to the proposal of Frequency Actuation Hypothesis by Phillips (1984): Physiologically motivated sound changes affect the most frequent words first; other sound changes affect the least frequent word first.

#### **4.6.2.2 For Syntactic Change**

From a syntactic change perspective, these two directions have also been revealed by scholars. Tottie (1991) has shown that for the development of negation in English, the more frequent a construction is, the more likely it is to be retained in its older form for a longer period of time. This finding is consistent with Ogura (1993), reported that there is a significant tendency for the high frequency words to change late. In Yue-Hashimoto (1993), a revised two-dimensional model was proposed, which argues that a new form enters the language and runs its own course of gradual victory and at the same time the old form independently runs its own course of gradual disappearance. Given the example of V-neg-V(P) question, Yue-Hashimoto (1993) has observed that the emergence of the V-neg-V(P) question form follows the tendency of first appearing with high-frequency verbs such as the copula, and/or the existential/possessive verb, and then gradually spread to the optative verbs and finally to other types of verbs. While new form first attacks high-frequency items, the old form at the same time meets the challenge and tries to wage a battle of resistance. It may recede first from the familiar, high-frequency items or from the unfamiliar, low-frequency items. The former is seen in the YILAN dialect of Taiwan and the latter is probably the case of Peking Mandarin. Therefore it turns out that for Chinese V-neg-V(P) question, the high frequency words may be more likely to be affected or be easily to be retained.

It can be summarized from the previous studies that word frequency is considered as the most crucial factor that is influencing the changing rate, for both phonological changes and syntactic changes. But to the best of my knowledge, no study has been conducted to discuss the role of word frequency in the process of VO transitivization in Mandarin Chinese.

#### 4.6.3 Research Question

Therefore in this study of VO compound transitivization, I would also like to investigate the correlation between word frequency and transitivity. The following research questions are addressed in this section:

1) Does the word frequency play a role in the process of VO

transitivization? If it does, what kind of role does it play?

- 2) Based on the previous studies, it seems that for the transition of VO compound, both directions are possible. If the word frequency does play a role, what direction can be observed (change affect high-frequency words earlier or change affect low-frequency words earlier)?
- 3) From the perspective of language variation, are there any variation differences in terms of the correlation between word frequency and transitivity, between Mainland and Taiwan Mandarin?

#### **4.6.4 Transitivity and Word Frequency**

This study focuses on the same 109 VO compounds that have been examined in the previous sections. The word frequencies of all the 109 VO compounds in both Mainland and Taiwan Mandarin have been calculated based on their frequency of occurrence in Gigaword Corpus (Mainland and Taiwan sub-corpus, respectively).

First for all the 109 words, correlation test (Spearman's correlation) is conducted to determine the correlation between transitivity and word frequency in both Mainland and Taiwan.

For all the 109 words, correlation test (Spearman's correlation) shows no significant correlation between transitivity frequency and word frequency, in both Taiwan (in Table 4.30) and Mainland (in Table 4.31);

		Correlation		
			TW	TW
			separation fre	transitivity
Spearman' rho	TW transitivity	Correlation coefficient	1.000	084
		Sig. (2-tailed)		.386
		Ν	109	109
	TW frequency	Correlation coefficient	084	1.000
		Sig. (2-tailed)	.386	
		Ν	109	109

\*. Correlation is significant at the 0.05 level (2-tailed). Table 4.30 The result of Spearman's correlation in Taiwan

		Correlation		
			TW	TW
			separation fre	transitivity
Spearman'	Mainland	Correlation	1.000	038
rho	transitivity	coefficient		
		Sig. (2-tailed)		.693
		Ν	109	109
	Mainland	Correlation	038	1.000
	frequency	coefficient		
		Sig. (2-tailed)	.693	
		Ν	109	109

\*. Correlation is significant at the 0.05 level (2-tailed). Table 4.31 The result of Spearman's correlation in Mainland

If there is no significant correlation between transitivity and word frequency, then what would be the case if the transitivity between highfrequency words and low-frequency words are compared?

Then the 109 words were divided into two groups in term of word frequency: high-frequency words and low-frequency words. There are several methods to determine the cut-off point between high and low frequency, e.g., TOP-N method, WF>=M method, %WF=P method, T method and h-point (e.g., An, 2014; Cheng, 2016). The most frequently used ones are TOP–N and WF $\geq$ = M. TOP-N method is ranking the words according to their frequency from high to low, and the first N words are considered as high frequency words (N usually to be 10, 20, 30, 50). WF>= M is very similar to TOP-N method, and the only difference is that for  $WF \ge M$  method, the threshold is frequency instead of ranking (M usually is 3, 5, 10, 15, 20). Another method is when % WF = P, the cut-off point is equal to a certain percentage (P is usually 20%, 50% and 70%). Although the first three methods are frequently used by many studies, the cut-off point determined by these methods seems to be subjective and random. Another two more complex methods are also adopted by previous scholar: T-method and H-point. T method is proposed by Donohue, J. C. (1973), which is using the formulation  $T = \frac{1}{2} (-1 + \sqrt{(1+8*I_1)})$  to determine the cut-off point ( $I_1$  refers to the numbers of words that only occur once in the dataset). H-point is also used as the threshold of high frequency words, it is an extension of the mathematical fixed point to the actual discrete rank-frequency distribution, f(h) = h (Macutek, Popescu and Altmann, 2007). Although the T method and h-point are more objective, they have requirements for the size of dataset. Since the dataset in my study is relatively small (109 words in total), T method and H-point may not be very efficient in determining the threshold.

Therefore in this study, I follow the method adopted in Bybee's studies (2000, 2002) and sort the VO compound by word frequency from high to low, the cut-off point between high and low frequency was arbitrarily chosen to make the number of tokens in the high and low groups approximately equal. This method is adopted for mainly two reasons: the first is this method is

already shown to be effective and workable in previous studies on lexical diffusion (e.g., Bybee, 2000, 2002). The second reason is due to the relatively small size of my dataset, this method is more appropriate and seems more likely to be efficient in determining the cut-off point.

The average of transitivity frequencies for both high-frequency and lowfrequency words in both varieties are shown in Table 4.32 and Table 4.33, respectively.

	High frequency	Low frequency
Numbers of words	55	54
Average transitivity	0.312317837	0.365271832
frequency		

 Table 4.32 Comparison of Taiwan transitivity between high and low frequency words

		High frequency	Low frequency
Numbers of words		55	54
Average	transitivity	0.270884515	0.296380724
frequency			

 Table 4.33 Comparison of Mainland transitivity between high and low frequency words

As can be seen from the above table, the differences in transitivity between high and low frequency words are not obvious, for both Mainland and Taiwan VO compounds. The average frequencies between these two groups are very close. Therefore, based on the statistics, it can be summarized as: in general, there is no obvious correlation between word frequency and transitivity rate. That is, the important effect of word frequency in the process of VO transitivization has not been observed in either Mainland or Taiwan Mandarin.

#### **4.6.5 Word Frequency and Alternation Type**

Based on the observations shown above, can we conclude as word frequency does not have any influence in the process of VO compound transition? If we make the conclusion like this, the theories of lexical diffusion theory will be challenged. However, we should also note that there is one more factor needs to be involved. In Study 3, it has been observed that there is a very close correlation between transitivity and its alternation pattern. That is, the alternative pattern of VO compound plays a very important role in the process of changing (i.e. the competing pattern is the most important factor which influences the transition rate of a VO compound). Therefore, what would be the case if both word frequency and alternation type are considered? That is, within the same alternation type, would the word frequency play a role in transition process?

To include the alternation type of VO compound as a variable, the 109 words were classified into 8 groups according to their alternation patterns (shown in Table 4.34). Since it has already been shown that there is no significant correlation between transitivity and type 4, 6, 7 and 8 groups (in Study 3), my study will only focus on the first 3 alternation patterns: only AV structure (hereafter 'Group 2'), only VC structure as its alternation (hereafter 'Group 3'), as well as both AV and VC structure as alternations (hereafter 'Group 1').

	Alternation	VO example	Alternation example
Group	Both VC	e.g., 进军 jinjun 'march	向电影界进军 xiang
1	and AV	towards'	dianyingjie jinjun
	alternation		towards_filmdom_march 进军
			到电影界 jinjun dao
			dianyingye
			march_towards_filmdom
			'march towards filmdom'
Group	Only AV	e.g., 把关 baguan	对产品质量把关 dui chanpin
2	alternation	'guarantee'	zhiliang baguan
			for_product_quality_guarantee
			'guarantee quality of product'
Group	Only VC	e.g., 驰誉 chiyu 'be	驰 誉 于 中 外 chiyu yu
3	alternation	famous'	<i>zhongwai</i> famous_at_home
			and abroad 'be famous at
			home and abroad'

Table 4.34 Alternation type 1, 2 and 3

Mainland data is first shown as an example. The data for each alternation is presented below. I still follow the studies of Bybee (2002, 2002) in determining the cut-off point between high and low frequency words. The principle is also to produce two groups of tokens of approximately equal size.

It should also be pointed out that the average frequency is usually used to make the comparison in most of studies on Lexical Diffusion Theory. According to Phillips (1984), considering the average frequency of each group proved more insightful than looking at individual items within each group. Therefore this study also compares the average of transitivity frequency between each group.

It is obvious to see from Table 4.35 that for Group 2 and 3, average transitivities of high frequency words tend to be higher. While for alternation 1 group, the transitivity frequencies of high and low frequency words are very close (low frequency words even have a slightly higher transitivity frequency).

Group	Description	Number	For each	High frequency	Low
		of	sub-		frequency
		words	group		
Group 1	VO has	33	Numbers	16	17
	both VC		Average	0.301580889	0.345275515
	and AV		frequency		
Group 2	VO has	31	Numbers	16	16
	only AV		Average	0.140627033	0.059937378
	alternation		frequency		
Group 3	VO has	21	Numbers	10	11
	only VC		Average	0.5027	0.4335
	alternation		frequency		

Table 4.35 Comparison of transitivity for each group in Mainland

As can be seen from Table 4.36, Taiwan VO compound is showing the exactly same tendency with Mainland: for alternation group 2 and 3, high frequency words tend to have higher transitivity. While for alternation group 1, the transitivity frequency of low frequency words is slightly higher (but the transitivity of high and low frequency words are still close).

Group	Description	Number	For each	High frequency	Low
		of	sub-		frequency
		words	group		
Group 1	VO has	33	Numbers	16	17
	both VC		Average	0.280242099	0.385658103
	and AV		frequency		
Group 2	VO has	31	Numbers	16	16
	only AV		Average	0.160174269	0.091113434
	alternation		frequency		
Group 3	VO has	21	Numbers	10	11
	only VC		Average	0.7073	0.5853
	alternation		frequency		

Table 4.36 Comparison of transitivity for each group in Taiwan

In terms of the role of word frequency, both Mainland and Taiwan Mandarin indicate that for the words with either only AV or VC construction as the competing alternation, the word frequency may be one of the important factors which are influencing the transition: high frequency words tend to have higher transitivity.

#### **4.6.5.1** Alternation Preference for Alternation Group 1

For the words from alternation Group 1 (have both AV and VC constructions), the transitivity frequency of low frequency words is higher, does it mean that for this group of words, the words with low frequency change first? Why the transition rate for this group is different from the other two? Are there any other factors influencing the transitivity?

There is one thing needed to be re-mentioned is the 'Multiple Competence Model' I have proposed in Study 3. That is, there also exists internal competition between different competing alternations. In other words, since VO compounds in Group 1 allow two alternations co-exist (both AV and VC constructions), the preferences for each alternation choice might be different. The different preferences may also influence their transition rate. In general, in study 3, I have found that if a VO prefers VC rather than AV, it is the VC that is leading the transition; if a VO prefers AV rather than VC, then AV will be playing a more important role in the process of transition. Therefore the internal competition between alternation types should also be considered as a factor in this study 4, to determine if this can influence the degree of transitivity.

The frequency of each alternation is used to measure the preference, i.e. the frequency of AV/VC usages in the corpus. The formulation is shown as

below.

Frequency of AV/VC=number of AV/VC constructions/all tokens of a VO

The ratio of AV to VC constructions is also used to measure whether there is an obvious differences in preferences between these two alternations, i.e. if the ratio is smaller than 2, this VO is not considered to have obvious preference for either of the two alternation constructions. For example, in Taiwan corpus, 媲美 *pimei* 'rival with' is used in an AV construction (such as 与美国媲美 *yu meiguo pimei* with\_America\_rival 'rival with America') for 55 times among all the 204 tokens of 媲美 *pimei* 'rival with' usages; and is used in a VC construction (such as 媲美于德国 *pimei yu deguo* rival with\_at\_Germany 'rival with Germany') for only 1 time: the frequency of AV alternation usage is 0.2696 (=55/204) and the frequency of VC alternation usage is 0.049 (=1/204). The ratio of AV to VC is 55, meaning that 媲美 *pimei* 'rival with' is 55 times more often to be used in AV construction than in VC construction, and this clearly displays that 媲美 *pimei* 'rival with' in Taiwan has a significant preference of AV constructions.

Both frequencies of AV and VC construction were calculated for each VO compound, and in both varieties (33 words in this group for each variety). Based on that, the ratio of VC to AV was also calculated. All the data was extracted from the Chinese Gigaword corpus (both Taiwan Central News Agency and Mainland Xinhua News Agency).

Based on the result of calculation, the words whose frequencies of AV

and VC alternations are both smaller than 0.01 were first excluded (the data size is too small for analysis). For example, 插手 *chashou* 'intervene' in Mainland, the frequency of its VC construction usages is 0.001, while the frequency of its AV construction usage is 0.004914. The words which do not show significant preference for neither of the alternation (ratio of AV to VC is smaller than 2) were also excluded for further analysis.

Then the words were divided into two groups according to their preference: whether it significantly prefers VC construction (ratio of VC/AV >2) or AV construction (ratio of AV/VC >2). The data of both Taiwan and Mainland can be seen in Table 4.37 and Table 4.38, respectively.

	Prefer VC	Prefer AV	
Number of words	5	18	
Average transitivity	0.301628055	0.269727864	
Table 4.27 Alternation and frances for Tairren date			

 Table 4.37 Alternation preference for Taiwan data

According to the criteria mentioned above, 5 words (立足 *lizu* 'base upon', 聚会 *juhui* 'gather', 献身 *xianshen* 'devote oneself to', 命名 *mingming* 'name' and 出土 *chutu* 'be underthed') in Taiwan which do not have obvious preference (ratio of AV to VC is smaller than 2) were excluded. In addition, the other five words (中意 *zhongyi* 'like', 寄语 *jiyu* 'send word', 借道 *jiedao* 'channeled through', 插手 *chashou* 'intervene' and 垂爱 *chuiai* 'show tender care for') whose frequencies of AV and VC alternations are both smaller than 0.01 were also removed. The other 23 words were divided into two groups according to their preference.

As can be seen from Table 4.37, there are only 5 VO compounds in

Taiwan show significant preference of VC construction: 迁居 *qianju* 'mvoe', 定 都 dingdu 'establish a capital', 聚焦 jujiao 'focus', 受聘 shoupin 'be hired by' and 失信 shixin 'break promise'. The average of their transitivity frequency is 0.301628055. There are 18 VO compounds which show obvious preference of AV construction: 媲美 pimei 'rival with', 称霸 chengba 'seek hegemony', 出身 chushen 'originally birth from', 称雄 chengxiong 'bestride', 现身 xianshen 'appear', 留学 liuxue 'study abroad', 会师 huishi 'join forces', 求证 qiuzheng 'verify', 满意 manyi 'be satisfied with', 献计 xianji 'offer advice', 驻军 zhujun 'station troops', 启航 qihang 'set sail', 避难 bi'nan 'seek shelter', 登场 dengchang 'come on stage', 进军 jinjun 'march towards', 出兵 chubing 'dispatch troops', 求助 giuzhu 'ask for help' and 执教 zhijiao 'coach'. The average transitivity for this group is 0.269727864. Although the frequencies of two groups are quite close, we can still observe a higher transitivity in the former group (i.e. words that prefer VC construction tend to have higher transitivity).

As shown in Table 4.37, the Mainland data shows a very similar result compared with Taiwan data.

	Prefer VC	Prefer AV
Number of words	6	20
Average transitivity	0.4448	0.261540288

#### Table 4.38 Alternation preference for Taiwan data

The 5 VO compounds in Mainland corpus which do not show significant preference for either of the alternation were removed: 中意 *zhongyi* 'like', 借道

*jiedao* 'channeled through', 命名 *mingming* 'name', 聚会 *juhui* 'gather' and 出 土 *chutu* 'be underthed'. Furthermore, there are 2 words (插手 *chashou* 'intervene' and 垂爱 *chuiai* 'show tender care for') whose frequencies of AV and VC alternations are both smaller than 0.01 have also been removed. The other 26 words are divided in terms of their preferences.

As can be seen from Table 4.38, the transitivity frequency of the 6 VO compounds which prefers VC construction (迁居 qianju 'mvoe', 定都 dingdu 'establish a capital', 立足 lizu 'base upon', 求助 qiuzhu 'ask for help', 受聘 shoupin 'be hired by' and 聚焦 jujiao 'focus') is 44.48%. There are 20 VO compounds which show obvious preference for AV construction: 进军 *jinjun* 'march towards', 媲美 pimei 'rival with', 称霸 chengba 'seek hegemony', 出身 chushen 'originally birth from', 称雄 chengxiong 'bestride', 现身 xianshen 'appear', 留学 liuxue 'study abroad', 会师 huishi 'join forces', 求证 qiuzheng 'verify', 满意 manyi 'be satisfied with', 献计 xianji 'offer advice', 驻军 zhujun 'station troops', 启航 qihang 'set sail', 避难 bi'nan 'seek shelter', 登场 dengchang 'come on stage', 进军 jinjun 'march towards', 出兵 chubing 'dispatch troops', 求助 *qiuzhu* 'ask for help', 失信 shixin 'break promise' and 执教 *zhijiao* 'coach', their transitivity frequency average is 0.261540288, which is obviously lower than that of words prefer VC construction. The same tendency is presented here: words prefer VC construction tend to have higher transitivity.

Based on the result of both Mainland and Taiwan Mandarin, for words have both AV and VC as alternations, the ones that prefer the VC construction tend to have higher transitivity frequency. This tendency actually is consistent with our analysis in study 3: VC construction is much easier to be transited into transitive usages.

#### 4.6.5.2 Why Alternation 1 has Such Tendency?

It should be noted that for the 5/6 words which prefer VC as the alternation, their word frequencies tend to be very low (迁居 *qianju* 'mvoe': 1.5; 定都 *dingdu* 'establish a capital': 0.28; 聚焦 *jujiao* 'focus': 0.55; 受聘 *shoupin* 'be hired by': 1.33 and 失信 *shixin* 'break promise': 0.79), while their transitivity frequencies tend to be very high, in both Mainland and Taiwan. This can explain the tendency we observed in Gourp 1: the transitivity of lower frequency words is slightly higher. The words with lower word frequency prefer VC construction, and the transition rate is mainly leaded by the VC construction.

#### 4.6.6 Summary

To make a summary, in this study, the correlation between transitivity and word frequency has been examined. The results show that the alternative pattern is the most important factor which influences the transition rate. The statistics indicate that in both Mainland and Taiwan Mandarin, for the VOs which prefer either only AV or VC alternation, their transitivity is higher among high frequency words. For the ones that have both AV and VC alternations, the ones prefer VC tend to have higher transitivity.

In that sense, I argue that word frequency only matters when it does not involve internal competition between alternations. If a word is still in the transition stage (that is when it allows both alternations in this case), then the frequency does not have much effect. The alternation choice plays a more important role in this stage.

In this case, to be particular, the VOs with only AV or VC alternation have finished their internal competition already, their syntactic constructions cannot be used to measure their transitivity, because it is the frequency predicting that. For the ones allow both AV and VC alternations, there is still competition between the two alternations, hence their nature of transitivity is corresponding to their syntactic structures. If they prefer VC construction more, they tend to have higher transitivity.

Therefore it is not surprising that for all the words have different alternation patterns, they do not show an obvious correlation between their transitivity frequency and word frequency. This is because word frequency is not the most determinant factor which is influencing their transition from intransitive to transitive. Their alternation type is more important. As has been illustrated in Study 3, there is a very close relationship between transitivity and alternation type (high transitivity VO is highly correlated with VC alternation while low transitivity VO is correlated with AV alternation). However, under each alternation type, word frequency does play a role in the process of changing.

In a word, according to my study, the alternative form (whether it is AV or VO construction) is the dominant factor which can influence the transition rate, while the word frequency is a lower-order impact factor.

#### 4.7 Summary of Chapter 4

In this chapter, grammatical variations in transitivity of VO compound between Mainland and Taiwan Mandarin have been carefully examined, with the assistance of large-scale comparable corpus as well as varieties of statistical models. Comprehensive descriptive accounts of VO compound variations have been provided in Section 4.3. It has been shown that most of the annotated features are very useful in differentiating VO usages from different language varieties. Four related studies have been conducted based on the statistical results. For study 1, the transitivity variations between Mainland and Taiwan VO compounds have been carefully investigated. Both variation differences in transitivity frequency and semantic/syntactic properties of the taken complement were considered. The comparison shows that Taiwan VO compounds tend to have higher degree of transitivity, and the same tendency has been presented in the study of light verb constructions (Chapter 3). Theoretical implications based on the grammatical variation in both transitivity frequency and syntactic/semantic properties have also been discussed. For study 2, the correlation between transitivity and separation ability is also examined, to see whether the grammatical variations of VO compounds are dependent on the degree of lexicalization of these compounds. For study 3, the relationship between transitivity and alternative competing pattern is also examined to further study the transition mechanism of VO compounds, as well as to provide the basis for study of correlation between language change and language variation (based on the Lexical Diffusion Theory (e.g., Wang, 1969)). A Multilevel Competition Model has also been proposed to account for the variation differences observed in corpus data. For study 4, the correlation between transitivity and word frequency is investigated to further discover the factors which affect VO transitivization. The analyses indicate that alternative pattern is the most important factor which can influence the transition rate while word frequency only matters when there does not involve internal competition between alternations.

## **Chapter 5 Conclusion**

The major goal of this thesis is to investigate the grammatical variations between Mainland and Taiwan Mandarin speaking varieties. For this goal, two grammatical constructions (i.e. light verb constructions and  $VO_1+O_2$ constructions) were carefully examined. Since the grammatical variations tend be differences in preference/tendency, probabilistic to syntactic model/framework was adopted. For each grammatical construction, mixedeffect logistic regression model and Chi-square test were carried out, to provide comprehensive descriptive account of syntactic variations. Based on the statistical results, various studied were conducted to discuss theoretical linguistic issues, including variation differences in degree of transitivity, syntactic type, lexical status, transition mechanism, alternative choices and word order. The correlation among language variation, language alternation and language change has also been discovered under the theoretical framework of Lexical Diffusion Theory.

### 5.1 Summary

There are several novel points need to be summarized:

#### 5.1.1 Probabilistic Nature of Grammatical Variations

The thesis has shown that grammatical variations between different language varieties tend to be differences in preference/tendency instead of un/grammatical dichotomy differences. This has been presented clearly in both

light verb and VO compound studies. Absolute differences are seldom been detected in corpora (i.e. one specific usage is only found in one variety), while in most cases, Mainland and Taiwan constructions are more likely to be differed in frequency or percentage of usages.

#### 5.1.2 Probabilistic Syntactic Model

The lack of sutides on grammatical variations is to a large extent due to its gradient nature. The probabilistic nature of grammatical variations actually is in accordance with the gradient grammar, and is also related to the probabilistic syntactic model (Bresnan et al., 2007, 2008, 2010). To be particular, the gradient nature makes the identification of detection of grammatical variations much more complicated. Hence, the probabilistic syntactic model needs to be adopted. With the assistance of large-scale comparable corpora, statistical models and linguistic features, solid empirical basis for the description of grammatical variations between Mainland and Taiwan Mandarin can be provided.

# 5.1.3 The Interaction between/of Language Alternation and Language Variation

In this thesis, I have shown that there is a close interaction between alternations and variations. Grammatical variations are not as simple as feature or behavior differences, but in fact differences in the patterns of alternations. Evidences have been shown in both light verbs and VO compounds studies.

For light verb variations, as shown in Section 3.4 of Chapter 3 that, Mainland and Taiwan have displayed significant differences in light verb alternations (i.e. differences in alternative choices between similar light verbs). Besides that, in Section 3.8 of Chapter 3, variations in patterns of alternations have also been observed between Mainland and Taiwan. Since there are different alternative patterns to introduce the theme of the verbal object for light verb constructions (e.g., 对可行性进行研究 dui kexingxing jinxing yanjiu for practicability proceed research 'to conduct research on practicability'; 进 行可行性研究 *jinxing kexingxing yanjiu* proceed\_practicability\_research 'to conduct research on practicability'; 进行(对)可行性的研究 *jinxing (dui)* kexingxing de yanjiu proceed (for) practicability DE research 'to conduct research on practicability'; 进行对可行性研究 jinxing dui kexingxing yanjiu proceed\_for\_practicability\_research 'to conduct research on practicability' and 进行研究**可行性** *jinxing vanjiu kexingxing* proceed research practiciability 'to conduct research on practicability'), Mainland and Taiwan are observed to differ in their syntactic choices. The theme in Taiwan Mandarin prefers to be appeared after the light verb (e.g., 进行可行性研究 jinxing kexingxing yanjiu proceed\_practicability\_research 'to conduct research on practicability'; 进行对 可行性研究 jinxing dui kexingxing yanjiu proceed for practicability research 'to conduct research on practicability'; 进行研究可行性 jinxing yanjiu kexingxing proceed research practiciability 'to conduct research on practicability'), while the theme in Mainland Mandarin significantly prefers to be appeared before the light verb (e.g., 对可行性进行研究 dui kexingxing jinxing yanjiu for\_practicability\_proceed\_research 'to conduct research on practicability'). In terms of VO transitivization, VO compounds have different alternative patterns to introduce the external objects: Adverbial-Verb construction (e.g., 与澳洲媲美 yu aozhou pimei with\_Australia\_rival 'rival with Australia'), Verb-complement construction (e.g., 媲美于澳洲 pimei yu aozhou rival with\_at\_Australia 'rival with Australia'), separation construction (e.g., 过广东境 guo Guangdong jing pass\_Guangdong\_boarder 'transit Guangdong') or transitive VO<sub>1</sub>+O<sub>2</sub> construction (e.g., 媲美澳洲 pimei aozhou rival with\_Australia 'rival with Australia'). In the study 3 of Chapter 4, the results have shown that Mainland and Taiwan actually differ in the syntactic choices between these alternations.

Furthermore, because of the close interaction, alternations can be used to serve as a linking bridge between language variation and language change. Therefore it is also possible to explore the relationship between language variation and language alternation (which has been shown in the study 3 of Chapter 4).

#### **5.1.4 Higher Transitivity of Taiwan Mandarin**

In Section 3.6 of Chapter 3 and Section 4.3 of Chapter 4, I have shown that for both light verb and VO compound, Taiwan Mandarin tends to have a higher transitivity. The higher degree of transitivity reflects in two aspects: in verb's ability of taking the objects and in collocation preference (i.e. what kinds of complements are more likely to be taken). In terms of the ability of taking objects, Taiwan light verb and VO compound appear to have higher transitivity in the sense that they are more frequently to be used in a transitive context, with less collocation constrains. In terms of collocation preference, the complements of light verb and VO compound in Taiwan show a higher degree in individuality and affectedness of the object. According to the parameters proposed by Hopper and Thompson (1980), higher individuality and affectedness of the object actually indicate a higher degree of transitivity.

#### 5.1.5 Language Conservatism Issue

Language conservatism issue is a challenging but important topic without definite answer. The innovative/conservative statuses of the two varieties remain uncertain and the direction of change also requires further studies.

For the light verb constructions, the higher degree of transitivity in Taiwan Mandarin in fact shows that Taiwan light verbs are less grammaticalized, hence more 'verbal' (remains more verbal properties) and therefore more conservative in terms of bleaching of verbal properties. The observation that Taiwan Mandarin are being more conservative actually is consistent with the null hypothesis in historical linguistic that the peripheral varieties away from the main speaking community tend to be more conservative. Taiwan Mandarin, as a peripheral variety away from the main Mandarin speaking community (i.e. the
Mainland), tends to show the characteristic of conservatism in language development and use. However, we can also argue that Taiwan light verbs are more close to the English light verbs like 'make' and allow for more types of complements. If the common direction for light verb change is from not taking argument to taking more types of arguments, then compared with Mainland light verbs, the Taiwan counterparts are more innovative.

For the VO compounds, the conservatism issue also remains uncertain. In study 1, it has been clearly observed that Taiwan VO compounds tend to have higher degree of transitivity. On one hand, the higher transitivity of Taiwan VO compounds may indicate the conservatism of Mandarin in Taiwan, in the sense that Taiwan VO compound remain more verbal properties, especially the ability of taking the object. Remaining the transitivity is an evidence of being conservative. In other word, the higher degree of transitivity is the consequence of attempting to preserve the transitive nature of the verbal construction during incorporation. In contrast, Mainland VO is introducing a new verb class that can take the object, which is transferring from the intransitive to transitive verbs. The evidence of Mainland usages being innovative is that we have found for the  $VO_1+O_2$  construction, the frequency of being appeared in headline is much higher in Mainland than in Taiwan Mandarin. The headline, as the summaries of a normal text, has length limit and also has more flexibility. It is usually forced to be innovative and new constructions usually emerge from the headline. On the other hand, the higher transitivity of Taiwan VO compounds may also indicate the innovationess of Taiwan Mandarin. Since the VO<sub>1</sub>+O<sub>2</sub> is a newly

emerged/popularized construction, having more unique transitive usages in Taiwan may be an evidence of being more creative in Taiwan Mandarin.

In study 2, Mainland VO compounds are observed to have higher separation frequency (i.e. they are more likely to be used separately). According to previous studies (e.g., Li and Wu, 2017), the noun incorporation and lexicalization are considered as an important common direction of VO change. To be specific, for the transitivization of VO compounds, O gets incorporated into the V and the process of noun incorporation is substantially a process of lexicalization, in the sense that the V and O are condensed into a disyllabic compound in the form VO, which can be used independently to describe an event. Based on this assumption, which is well-attested cross-linguistically, the steps of changes should be the following:

Step 1 originally: V and O are two independent constituents, V+ O is a verbal phrase, and there is no restriction (other than by other grammatical rules) on what can occur between V and O.

Step 2: VP incorporation, O becomes a part of the compound verb, in the process, several things can happen:

- a. There could be residue cases where O can still be separated from V (i.e. separation)
- b. The compound verb VO should be intransitive at first (as there is no other arguments), additional argument should be introduced in oblique positions (prepositional phrase, possessors, or topic in Chinese) remaining at their original position in VP

c. Eventually, the transitivity of VO can become increase, which allows VO<sub>1</sub>+O<sub>2</sub>; that is some of the original oblique argument in b will move to direct object position

Hence from the point of view of incorporation and lexicalization, the correlation between facts and indication of being innovative/conservative when comparing the variations are illustrated as follows:

i. For VO compound, the variety with the following characteristics is more innovative (i.e. incorporation is more complete)

-More lexical types of VO

-Less flexibility among the attested VO compounds

-higher frequency of VO uses vs. separable uses

-higher transitivity (i.e. more O<sub>2</sub>'s etc)

ii. For  $O_1$  in  $VO_1+O_2$  construction, the variety with the following characteristics is more conservative

-More varieties/complexity of separable constructions

-More concrete/referential O<sub>1</sub>'s in separable construction (allowing individual classifiers, modified by physical attributes etc.)

iii. For  $O_2$  in  $VO_1+O_2$  construction, the variety with the following characteristics is more innovative: (Based on the assumption that after incorporation, it would be more conservative to keep the original argument forms from VP, and it would be more innovative to move those oblique argument to direct object position)

-higher number of VO's taking O<sub>2</sub>

-More varieties of O<sub>2</sub>,

-More varieties/complexity of the constructions for O<sub>2</sub> to appear in In this perspective, it seems that Taiwan is more innovative. Having higher transitivity of VO compounds, higher number of VO's taking O<sub>2</sub>, more varieties of O<sub>2</sub> and more varieties/complexity of the constructions for O<sub>2</sub> to be appeared in, are all evidences for Taiwan's innovativeness. This seems to contradict prediction by its language status, which is already supported by the adaption of orthographical change (Taiwan, Hong Kong, Singapore, and Macau were all conservative, Mainland innovative in adopting Pinyin and simplified characters first).

Overall, with respect to both light verb construction and VO transitivization, it seems that there is a contradictory description of the conservative/innovative status of the two varieties, and this surely requires further studies to resolve. More evidences, especially the historical evidences need to be involved.

#### **5.1.6 Variation Differences in Word Order**

Another tendency needs to be mentioned is Taiwan tends to prefer heavier objects. This finding actually is related to the preference of word order.

There are mainly two evidences to show the tendency: in Section 3.8 of Chapter 3, I have shown that for the light verb construction, the theme of eventive complements in Taiwan Mandarin prefers to be appeared after the light verb (e.g., 进行可行性研究 *jinxing kexingxing yanjiu*  proceed\_practicability\_research 'to conduct research on practicability'; 进行研 究可行性 *jinxing yanjiu kexingxing* proceed\_research\_practiciability 'to conduct research on practicability'), while the theme in Mainland Mandarin significantly prefers to be appeared before the light verb (e.g., 对可行性进行研 究 *dui kexingxing jinxing yanjiu* for\_practicability\_proceed\_research'to conduct research on practicability').

In addition, in Section 4.3.2 of Chapter 4, I have also shown that VO compounds in Taiwan tend to be followed by complicated object (e.g., 媲美这 个世界一流名声大噪的科技新产品 pimei zhege shijieyiliu mingshengdazao de keji xinchanpin rival with this worldclass famous DE technology new product 'rival with this world-class famous new technology product'). While complicated object tends to appear before the VO in Mainland Mandarin, as in adverbial-verb construction (e.g., 与这个世界 一流名声大噪的科技新产品媲美 yu zhege shijieyiliu mingshengdazao de keji xinchanpin pimei with\_this\_worldclass famous DE technology new product rival 'rival with this world-class famous new technology product').

Both tendencies have indicated that Taiwan Mandarin reveals the characteristic of OV word order (Cheng, 1985), while Mainland tends to have the characteristic of VO word order. This distinction actually corresponds with the rough rule of English complex NP shift, which was first studied by Ross (1967). Complex NP shift (or heavy NP shift), particular manifestation of shifting where a 'heavy' noun phrase (NP) appears in a position to the right of its canonical position under certain circumstances. Ross (1967) describes 'Heavy NP shift' as a rule which optionally moves a complex NP to the end of the first sentence up.

In that sense, Taiwan Mandarin is more like English to have heavy NP shift. When the object is collocated with complex modifiers, it prefers to be located before the main verb in Mainland Mandarin. While in Taiwan Mandarin, the complicated NP tends to appear on the right side of the main verb.

#### **5.2** Possible Applications in Computational Linguistics

In computational linguistics, automatic identification of language variations poses a special challenge in many Natural Language Processing (NLP) applications, e.g., information retrieval and machine translation. The module of generic linguistic tools shown in this thesis can actually have several potentially very useful applications in computational linguistics. For example, for light verb constructions, applications in machine translation, information extractions as well as error detection are all possible to be implemented based on the model proposed in my thesis. First, in translation, LVC is one of the most difficult constructions as there is less grammatical or contextual information to make the correct translation. As contextual selection information has been encoded for all light verbs in this thesis, the same approach can be applied to the other languages in the target-source pair to produce optimal pair. Second, in information extraction, selection of different light verbs often conveys subtle difference in meanings. The ability to differentiate similar light verbs in the same context could have great potential in extracting the subtle information change/increase in the same context. In addition, in error detection, the model utilized in my thesis can be readily applied to provide the correct context where a certain light very is preferred over another.

#### **5.3 Future Studies**

### 5.3.1 Correlation between Language Variation and Language Change for the Study of Light Verbs

In Section 4.6 of Chapter 4, the correlation between language variation and language change has been examined only in the study of VO compounds. The same methodology can also be applied to the study of light verbs.

According to the direction of lexicalization and grammaticalization (Brinton and Traugott, 2005), being semantically more bleached should be considered as a common direction of light verb change. That is, light verb should be developed from having more content to more light. Light verb which has higher degree in semantic impoverishment is considered as more developed light verb.

The impoverishment of light verb can be measured in different approaches. Dras (1995) proposes that a word's status regarding content-freeness is related to its frequency of occurrence in light constructions. That is, the frequency of being appeared in light verb constructions can be used to measure the degree of impoverishment of a light verb. In addition, Samardzic (2008) also claims that the taken complement of the light verb (lexical bias of the complement) is also related to the degree of impoverishment of light verb under certain context. Hence, the variation differences in degree of impoverishment between Mainland and Taiwan light verbs are able to be investigated in both ways: frequency of being used in light verb construction and the taken complement. Since the degree of impoverishment can actually indicates the different stages that Mainland and Taiwan light verbs are located in the process of change, the correlation between language change and language alternation (different alternative choices in taken complements) is possible to be examined for light verb constructions.

#### **5.3.2** Other Constructions in Other Languages

In my thesis, I have shown that with the assistance of large-scale comparable corpora and statistical tool, subtle grammatical variations can be detected very effectively. The ability of comparable corpus-based statistical approach to differentiate Mainland and Taiwan variants of Mandarin Chinese potentially contributes to overcoming the challenge of identification of subtle language/dialect variations among other grammatical constructions, as well as between other languages/dialects. The annotated semantic and syntactic features which are very effective in differentiation language variations in my thesis can also be used for future studies on other grammatical categories, as well as between other Mandarin varieties (e.g., Hong Kong Mandarin, Singapore

Mandarin, Malaysia Mandarin and so on).

## **APPENDIX I**

Taiwan examples	Taiwan examples	Taiwan examples
<做>答覆	<做>表示	<做>理性沟通
<做>澄清	<做>人身攻击	<做>环保
<做>呼吁	<做>考虑	<做>接触
<做>信心喊话	<做>冲刺	<做>竞赛
<做>回应	<做>调节性泄洪	

Appendix I. Specfic usages of 做 zuo in Taiwan Mandarin

# **APPENDIX II**

Mainland examples	Mainland examples	Mainland examples
<搞>倾销	<搞>计划种植	<搞>接待
<搞>一刀切	<搞>对外开放	<搞>阶级斗争
<搞>教育	<搞>社会主义	<搞>专题讲座
<搞>包产到户	<搞>劳动制度配套改革	<搞>运输
<搞>单一化	<搞>军民共建	<搞>突击运动
<搞>冬种	<搞>大流通	<搞>技改
<搞>服务	<搞>假出口押汇手段	<搞>勘察
<搞>副业	<搞>隐性奖金分配	<搞>加价
<搞>集中供热	<搞>股份制试点	<搞>精神文明建设
<搞>短期行为	<搞>调查	<搞>公费旅游
<搞>种族压迫	<搞>多头重复检查	<搞>清障
<搞>花炮塑料加工	<搞>实业	<搞>蔬菜批发
<搞>加工	<搞>技术承包服务	<搞>加工配套
<搞>零售	<搞>收购、兼并	<搞>钻井
<搞>批发	<搞>代办铁路运输	<搞>销售
<搞>科技承包	<搞>了租赁、拍卖等	<搞>市场化
<搞>生产	<搞>生产自救	<搞>养殖加工
<搞>生活后勤	<搞>服装培训	<搞>蕃茄杂交授粉
<搞>小家电发明	<搞>育种	<搞>灌溉
<搞>竹编工艺	<搞>服装设计	<搞>有偿服务
<搞>市场调查	<搞>股份制试点	<搞>物质文明建设
<搞>测量	<搞>制作	<搞>直销
<搞>经济发展规划	<搞>集团承包和科技示	<搞>航运
	范	
<搞>游行	<搞>统一战线事业	<搞>防洪建设
<搞>集体商业	<搞>加工服务	<搞>海水养殖
<搞>边贸	<搞>应酬	<搞>种养
<搞>贩运	<搞>福利	<搞>设计
<搞>营养袋育苗密植移	<搞>社会主义市场经济	<搞>地方保护
栽		
<搞>训练	<搞>基础性服务	<搞>原料粗加工
<搞>专业修剪	<搞>镭射	<搞>专项检查
<搞>拥军优属工作	<搞>绿化	<搞>评比
<搞>水土保持	<搞>工业性试验	<搞>专业种植养殖
<搞>包干制	<搞>农业综合开发试验	<搞>反季节生产
<搞>现货交易	<搞>现代化建设	<搞>玉米加工
<搞>市场经济	<搞>形式主义	<搞>试验
<搞>资产阶级自由化	<搞>羊绒加工	

Appendix II. Specfic usages of 搞 gao in Mainland Mandarin

### References

- Agresti, A., and Kateri, M. (2011). *Categorical data analysis* Springer Berlin Heidelberg, pp. 206-208.
- An, X. R. (2014). 基于正态分布的词频分析法高频词阈值研究 (The Research on the Threshold of High-Frequency words based on the Normal Distribution in Word Frequency Analysis). Journal of Intelligence, 33(10), 129-136.
- Bates, D., and Maechler, M. (2009). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-32. HYPERLINK http://CRAN.R-project.org/package=lme4
- Bergem, D. R. V. (1995). Acoustic and lexical vowel reduction (Studies in Language and Language Use 16). Amsterdam: IFOTT.
- Bod, R., and Hay, J. (2003). *Probabilistic linguistics*. Massachusetts: Mit Press.
- Bresnan, J., Cueni, A., Nikitina, T., and Baayen, R. H. (2007a). Predicting the dative alternation. *Cognitive Foundations of Interpretation*, 69-94.
- Bresnan, J. (2007b). Is syntactic knowledge probabilistic? Experiments with the English dative alternation. *Roots: Linguistics in search of its evidential base*, 75-96.
- Bresnan, J., and Hay, J. (2008). Gradient grammar: an effect of animacy on the syntax of give in New Zealand and American English. *Lingua*, 118 (2), 245–259.
- Bresnan, J., and Ford, M. (2010). Predicting syntax: Processing dative constructions in American and Australian varieties of English. *Language*, 86(1), 168-213.
- Brinton, L. J., and Traugott, E. C. (2005). *Lexicalization and language change*. Cambridge University Press.

- Britain, D. (2016). 'Sociolinguistic Variation'Subject Centre for Languages, Linguistics and Area Studies Good Practice Guide. Retrieved from http://linguistics.oxfordre.com/view/10.1093/acrefore/9780199384655.001 .0001/acrefore-9780199384655-e-364
- Butt, M. (1995). *The structure of complex predicates in Urdu*. Center for the Study of Language (CSLI).
- Butt, M. (2010). The light verb jungle: Still hacking away. *Complex predicates in cross-linguistic perspective*, 48-78.
- Butt, M., and Geuder, W. (2001). On the (semi) lexical status of light verbs. Semi-lexical Categories, 323-370.
- Bybee, J. (1976). Word frequency in lexical diffusion and the source of morphophonological change. In W. Christie (eds.), *Current progress in historical linguistics*. Amsterdam: North Holland. 96, 105.
- Bybee, J. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (eds.), Usage-based models of language. Stanford: CSLI. 65–85.
- Bybee, J. (2002). Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change. *Language variation and change*, 14(3), 261-290.
- Cai, W. L. (1982). "进行"带宾问题 (Issues on the Complement of 'jinxing'). Chinese Language Learning, 3, 7-11.
- Cedergren, H. J., and Sankoff, D. (1974). Variable rules: Performance as a statistical reflection of competence. *Language*, 333-355.
- Chao, Y. R. (1968). *A grammar of spoken Chinese*. University of California Press, Berkeley: California.
- Chen, C. M. (1995). 谈述宾短语带宾语的几个问题 (Several issues on VO

phrase taking Objects). Journal of Jinan University (Philosophy and Social Edition), 1, 116-122.

- Chen, N. P. (1987). 现代汉语名词类的扩大 (The Expansion of Modern Chinese Noun Class). *Zhongguo Yuwen*, 5, 379-389.
- Cheng, J. (2016). 基于语料库的汉语历时高频词对比研究 (A research of Diachronic Analysis about Chinese High-frequency Words Based on) (Master's thesis). Anhui University.
- Cheng, R. L. (1985). A comparison of Taiwanese, Taiwan Mandarin, and Peking Mandarin. *Language*, 352-377.
- Cheng, R. L. (1990). 词汇扩散理论在句法变化里的应用——兼谈台湾官话 "有"字句的句法变化 (Lexical Diffusion Theory and its application in syntax: A discussion of the changes in you constructions in Taiwan Mandarin). *Language Teaching and Linguistic Studies*, 1, 66-73.
- Chin, Andy. (2007). A sociocultural comparison of Chinese news headline verbs between Hong Kong and Taiwan. Proceedings of The First International Workshop on Intercultural Collaboration -IWIC 2007. 444-450.
- Chinese Knowledge and Information Process Group (CKIP Group). (2006). Sinica Corpus user manual, retrieved from http://ckip.iis.sinica.edu.tw/CKIP/paper/Sinica%20Corpus%20user%20 manual.pdf.
- Chu, Z. X., and Zhang, Q. (2013). 海峡两岸"透过"用法的多样性与倾向性考察 (The Diversity and Tendentiousness Investigation of touguo in Both Sides of the Taiwan Straits). *Applied Linguistics*, 4, 70-79.
- Cui, S. J. (2008). 现代汉语潜显现象研究 (On the Implicit and ExplicitPhenomenon in Modern Chinese). Chengdu: Bashu Press.
- Curme, G. O. (1935). Parts of speech and accidence. Boston: D.C. Health and

Company.

- Dai, Z. M., and Zhao, Y. F. (2009). "港台式"语法及其对内地汉语的影响 (Grammar of Hong Kong/Taiwan Mandarin and the influence on Mainland Mandarin). In Li, X. X. (Eds.), 海峡两岸现代汉语研究 (Research on Modern Chinese Cross Strait) (pp. 191-205). Hong Kong: Educational and Cultural Press Limited,
- Diao, Y. B. (1998). 也谈"动宾式动词+宾语"形式 (Also on VO+O construction). Language Planning, 6, 39-41.
- Diao, Y. B. (1998a). 台湾话的特点及其与内地的差异 (The Characteristic of Taiwan Mandarin and its Differences between Mainland Mandarin). *Zhongguo Yuwen*, 5, 387-390.
- Diao, Y. B. (2000). 差异与融合: 海峡两岸语言应用对比 (Divergence and Convergence: A Cross-Strait Language Comparison). Nangchang: Jiangxi Education Press.
- Diao, Y. B. (2004). 现代汉语虚义动词研究 (Research on Delexical Verb in Modern Chinese). Dalian: Liaoning Normal University Press.
- Diao, Y. B. (2012a). 两岸四地"获"字句对比考察 (The Comparative Study of the Sentence with huo in the Four Places across the Taiwan Strait). *TCSOL*, 2, 67-76.
- Diao, Y. B. (2012b). 两岸四地现代汉语常用词"进行"使用情况对比考察与 分析 (Comparison and Analysis of the Usage of jinxing in the Four Places across the Taiwan Strait). *Journal of Wuling*, 3, 116-124.
- Diao, Y. B. (2012c). 两岸四地的"遭"字句及其与"被"字句的差异 (The Contrast Study about zao and bei Sentences in the Four Places across the Taiwan Strait). *Language Teaching and Linguistic Studies*, 5, 70-77.

- Diao, Y. B. (2012e). 台港澳地区"搞"的使用情况及其与内地的差异 (On the Usage of Gao in the Regions of Taiwan, Hong Kong and Macao and the Differences with the Mainland). *Journal of Weinan Normal University*, 9, 80-83.
- Diao, Y. B. (2012f). 两岸四地"将"字句及其使用情况考察 (Research on Jiang Construction in Modern Chinese in the Four Places across the Taiwan Strait and Its Usages). *Journal of Nanyang Normal University*, 5, 37-43.
- Diao, Y. B. (2012g). 两岸四地"有+VP"形式考察 (The Study of you+VP in the four places across the Taiwan Strait). *Academic Journal of Liyun*, 1, 27-51.
- Diao, Y. B. (2013). 两岸四地"被"字句对比考察 (The Comparative Study of the Sentence with BEI in the Four Places across the Taiwan Strait). *Linguistic Researches*, 2, 17-23.
- Ding X. X., and Yuan X. M. (1998). 对"动宾式动词+宾语"句式增多的思考. (Thoughts On the Increasing Popularity of VO+O Construction). *Language Planning*, 3, 30-32.
- Doğruöz, A. S., Nakov, P., and Tower, T. (2014). Predicting Dialect Variation in Immigrant Contexts Using Light Verb Constructions. In *EMNLP* (pp. 1391-1395).
- Donohue, J. C. (1973). Understanding scientific literatures: a bibliometric approach. Massachusetts: MIT Press.
- Dras, M. (1995). Automatic identification of support verbs: A step towards a definition of semantic weight. *arXiv preprint cmp-lg/9510007*.

Du, Q. E. (2010). 现代汉语形式动词研究 (A Tentative Study on Dummy Verbs

*of Contemporary Chinese Language)* (Master's thesis). Shanghai: Shanghai Normal University.

- Dunn, O. J. (1964). Multiple comparisons using rank sums. *Technometrics*, 6, 241-252.
- Fang, Q. M. (2013). 基于口语库统计的两岸华语语气标记比较研究 (A Contrast Study of the Mood Markers between Taiwanese Mandarin and Mandarin Chinese: Based on the Spoken Chinese Corpus). *TCSOL*, 3, 58-65.
- Fang, Q. M. (2014). 基于口语库统计的两岸华语指示标记比较研究 (A Contrast Study of the Demonstrative Markers between Taiwanese Mandarin and Mandarin Chinese: Based on the Spoken Chinese Corpus). *Linguistic Sciences*, 13(2), 131-139.
- Fang, Q. M. (2015). 基于口语库统计的两岸华语程度副词比较研究 (A Contrast Study of the Degree Adverbs between Taiwanese Mandarin and Mandarin Chinese: Based on the Spoken Chinese Corpus). *Shijie Huawen Jiaoxue*, 01, 88-98.
- Fidelholtz, J. L. (1975). Word frequency and vowel reduction in English. *Chicago Linguistic Society*, 11, 200–213.
- Gao, G. S. (1998). "动宾式动词+ 宾语"的搭配规律 (The Collocation Rules for VO+O construction). *Language Planning*, 6, 36-38.
- Gregory, M. L., Raymond, W. D., Bell, A., Fosler-Lussier, E., and Jurafsky, D. (1999). The effects of collocational strength and contextual predictability in lexical production. *Chicago Linguistic Society*, 35, 151–166.
- Grimshaw, J., and Mester, A. (1988). Light verbs and θ-marking. *Linguistic inquiry*, 205-232.

- Hazen, K. (2011). Labov: language variation and change. *The SAGE Handbook* of Sociolinguistics, 24.
- Her, O. S. (1991). Variation of transitivity in Mandarin Chinese VO compound verbs. *NACCL-3*, Cornell University.
- Her, O. S. (1996). Variation of the VO construction in Chinese: a synchronic account. *Linguistics*, 34(4), 733-752.
- Hong, J. F., and Huang, C. R. (2008). 語料庫為本的兩岸對應詞彙發掘 (A Corpus-Based Approach to the Discovery of Cross-Strait Lexical Contrasts). *Language and Linguistics*, 9(2), 221-238.
- Hook, P. E. (1974). The Hindi compound verb: What it is and what it does. *Readings in Hindi-Urdu Linguistics*. Delhi: National Publishing House.
- Hook, P. E. (1991). The emergence of perfective aspect in Indo-Aryan languages. *Approaches to grammaticalization*, 2, 59-89.
- Hopper, P. J., and Thompson, S. A. (1980). Transitivity in grammar and discourse. *Language*, 251-299.
- Hu, Q. G. (2003). 述宾式动词及其宾语的语义类型 (VO verb and the Semantic Type of its Taken Complement). *Journal of Yumeng*, 24(1), 110-112.
- Hu, Y. S., and Xiao F. (1995). Research on Verbs (动词研究). Kaifeng: Henan University Press.
- Hua, Y. M. (1997). 试探动宾动词加宾语流行的根由 (An Investigation on the Motivation of VO+O construction). *Language Planning*, 10, 40-41.
- Huang B. R., and Liao X. D. (2007). *Modern Chinese 现代汉语*. Beijing Higher Education Press.
- Huang, C. R. (2009). Tagged Chinese Gigaword Version 2.0. Philadelphia:

Lexical Data Consortium, University of Pennsylvania. ISBN 1-58563-516-2 <u>http://www.ldc.upenn.edu/Catalog/catalogEntry.jsp?catalogId=LDC200</u> 9T14

- Huang, C. R. (2016). Endurant vs Perdurant: Ontological Motivation for Language Variations. PACLIC 30, 15.
- Huang, C. R., and Ahrens, K. (2003). Individuals, kinds and events: classifier coercion of nouns. *Language Sciences*, 25, 4, 353–373.
- Huang, C. R., Chen, K. J., and Lai, C. X. (1995). Mandarin Chinese Classifier and Noun-Classifier Collocation Dictionary. Taipei: Mandarin Daily Press.
- Huang, C. R., and Lin, J. X. (2013). The ordering of Mandarin Chinese light verbs. *Proceedings of the 13th Chinese Lexical Semantics Workshop*. In Ji D. and Xiao G. (Eds.), CLSW 2012, LNAI 7717, pp. 728-735. Heidelberg: Springer.
- Huang, C. R., Lin, J. X., and Zhang, H. R. (2013). World Chineses based on comparable corpus: The case of grammatical variations of jinxing. 澳门语 言文化研究, pages 397-414.
- Huang, C. R., Lin, J. X., Jiang, M. H., and Xu, H. Z. (2014). Corpus-based Study and Identification of Mandarin Chinese Light Verb Variations. COLING Workshop on Applying NLP Tools to Similar Languages, Varieties and Dialects. Dublin, August 23.
- Huang, C. R., Simon P., and Hsieh S. K. (2007). Automatic Discovery of Named Entity Variants. Proceedings of the Association of Computational Linguistics Annual Meeting. 153-156.
- Huang, C. T. J. (1984). Phrase structure, lexical integrity, and Chinese compounds. *Journal of the Chinese Language Teachers Association*,

- Huang, G. Y. (1988). 台湾当代小说的词汇语法特点 (The lexical and grammatical property of Taiwan Contemporary Fiction). *Zhongguo Yuwen*, 3, 194-201.
- Hwang, J. D., Bhatia, A., Bonial, C., Mansouri, A., Vaidya, A., Xue, N. W., and Palmer, M. (2010). PropBank annotation of multilingual light verb constructions. *Proceedings of the Fourth Linguistic Annotation Workshop, ACL 2010*, 82–90.
- Institute of Applied Linguistic Ministry of Education. (2009). Introduction to CN corpus (国家语委语料库科研成果简介). [EB/OL] [2007-10-15]. www.china2languange.gov.cn.
- Iyeiri, Y. (2010). Verbs of Implicit Negation and their Complements in the History of English. John Benjamins Publishing.
- Jespersen, O. (1965). A Modern English Grammar on Historical Principles. Part VI, Morphology. London: George Allen and Unwin Ltd.
- Jurafsky, D., Bell, A., Gregory, M., and Raymond, W. D. (2001). Probabilistic relations between words: Evidence from reduction in lexical production. In Bybee, J. and Hopper, P. (eds.), *Frequency and the emergence of linguistic structure*. Amsterdam: Benjamins. 229–254.
- Kearns, K. (1998). *Light verbs in English* (Master's thesis), New Zealand: University of Canterbury.
- Kubler, C. C. (1985). *The development of Mandarin in Taiwan: A case study of language contact* (Vol. 9). Taiwan Student Book Press.
- Kwong, O. Y., and Tsou B. K. (2004). The usage and perception of judgement terms in the Pan-Chinese context. *Proceeding of 5th Chinese Lexical Semantics Workshop (CLSW5).* 220-227.

- Kwong, O. Y., and Tsou B. K. (2006). Regional variation of domain-specific lexical items: Toward a Pan-Chinese lexical resource. *Proceedings of the Fifth SIGHAN Workshop on Chinese Language Processing*. 9-16.
- Kwong, O. Y., and Tsou B. K. (2008). Extending a thesaurus with words from Pan-Chinese sources. Proceedings of the 22nd International Conference on Computational Linguistics. 457-464.
- Labov, W. (1994). Principles of linguistic change: Internal factors. Oxford: Blackwell.
- Labov, W. (2006). *The social stratification of English in New York city*. Cambridge University Press.
- Labov, W., and Wang, W. S. -Y. (2014). 语音变化前沿问题演讲录 (A Speech on the Frontier of Sound Change). *Language Teaching and Linguistic Studies*, 1, 1-12
- Language Institute of Chinese Academy of Sciences. (2005). 现代汉语词典 (第5版). Modern Chinese Dictionary (5th Edition). Beijing: Commercial Press.
- Lee, Y. M. S. (2010). A linguistic approach to emotion detection and classification (Doctoral dissertation), The Hong Kong Polytechnic University).
- Li, Y. M. (2010). 全球华语词典 (A Dictionary of Global Huayu (Chinese Language). Beijing: The Commercial Press.
- Li, T. (2015). 台湾国语与大陆普通话口语中程度副词使用情况对比分析 (Comparative Analysis of the Degree Adverbs Usages between Taiwan Mandarin and Mainland Mandarin Spoken Language). (Master's thesis).
   Guangzhou, Jinan University.
- Li, L. D. (1990). 现代汉语动词 (Verbs in Modern Chinese). Beijing: China

Social Science Press.

- Li, C. N., and Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* University of California Press.
- Li, Y. H. A. (1985). Case, Tense and Clauses. Language, 51, 1990.
- Li, Y. N. (2014). 现代汉语"动宾宾"结构在海峡两岸的变异情况探究 (The Study On Variation of Modern Chinese VO+O structure between China Mainland and Taiwan District) (Master's thesis), Guangzhou: Jinan University.
- Li, Y. Z., and Wu. Y. C. (2017). On the evolutionary mechanism of disyllabic transitive verbs in Chinese. *Journal of Chinese Linguistics*.
- Liao, L. P. (1999). 台湾小说单音形容词的重叠形式-海峡两岸词的重叠形式 的对比研究 (The Reduplication Form of Monosyllabic Adjectives in Taiwan Novels - A Contrastive Study of the Reduplication Forms of Cross-Strait Words). *Journal of Xuzhou Normal University*, 3: 66-68.
- Lin, J. X., Shi, D. X., Jiang M. H., and Huang, C. R. (To appear). Variation in World Chineses. (In Chu-ren Huang, Zhuo Jing-Schmidit and Barbara Meisterernst, eds.), *The Routledge Handbook of Chinese Applied Linguistics*
- Lin, J. X., Xu, H. Z., Jiang M. H., and Huang, C. R. (2014). Annotation and classification of light verbs and light verb variations in Mandarin Chinese. COLING Workshop on Lexical and Grammatical Resources for Language Processing. Dublin, August 24.
- Ling, D. X. (1999). 试论双音节动宾式动词带宾语的基本规律 (An Analysis on Verb-object Type + Object). *Chinese Language Learning*, 5, 9-13.
- Liu, D. W. (1998a). 关于动宾带宾现象的一些思考 (上) (Some Thoughts on VO compounds taking Objects (1)). *Language Planning*, 1, 22-26.

- Liu, D. W. (1998b). 关于动宾带宾现象的一些思考 (下) (Some Thoughts on VO compounds taking Objects (2)). *Language Planning*, 3, 28-29.
- Liu, Y. F. (2005). 形式动词及其宾语 (Form Verbs and their objects). Journal of Southwest University for Nationalities (Humanities and Social Sciences), 5, 353-356.
- Liu, Y. J. (1993). 动宾式动词与所带宾语之间的语义关系 (The Semantic Correlation between VO compounds and its Taken Complement). *Chinese Language Learning*, 4, 48-53.
- Liu, Y., and Li, J. X. (1998). "动宾式动词+ 宾语"的变换形式及宾语的语义 类型 (The Transformation of VO<sub>1</sub>+O<sub>2</sub> structure and the Semantic Type of O<sub>2</sub>). *Journal of Jianghan University*, 5, 011.
- Losiewicz, B. L. (1992). *The effect of frequency on linguistic morphology*(Doctoral dissertation), University of Texas at Austin.
- Lu, B., and Tsou, B. K. (To appear). Mining parallel knowledge from comparable patents. *In Ontology Learning and Knowledge Discovery Using the Web: Challenges and Recent Advances*. IGI Global.
- Luo, X. R. (1998). "动宾式动词+ 宾语"规律探究 (Analysis on VO+N). Language Planning, 5, 27-30.
- Lü, S. X. (1980). 现代汉语八百词 (Eight Hundred Modern Chinese Words). Beijing: Commercial Press.
- Mao, F. S. (2002). 两岸三地外语影片片名的翻译比较 (A Comparison Study of Translation of Foreign Films in the Three Places across the Taiwan Strait). *Journal of Xi'an International Studies University*. 4, 19-23.
- Mao, Z. M. (2007). 台湾书面语中的闽南方言特点初探-以台湾报刊语言为
  例 (The Code-Mixing of Taiwanese in Taiwan Mandarin Newspapers) (Master's thesis). Shanghai: East China Normal University.

- Masolo, C., Borgo, S., Gangemi, A., Guarino, N., Oltramari, A., Oltramari, R., Schneider and Horrocks, I. (2002). *WonderWeb deliverable D17*. the WonderWeb library of foundational ontologies and the DOLCE ontology.
- Macutek, J., Popescu, I. I., and Altmann, G. (2007). Confidence intervals and tests for the h-point and related text characteristics. *Glottometrics*, 15, 45-52.
- Moonwomon, B. (1992). The mechanism of lexical diffusion. *Paper presented at the Annual Meeting of the Linguistic Society of America*, Philadelphia.
- Mohanan, T. (1994). Argument structure in Hindi. Center for the Study of Language (CSLI).
- Nagy, T., Vincze, V., and Farkas, R. (2013). Full-coverage Identification of English Light Verb Constructions. In Proceedings of the International Joint Conference on Natural Language Processing, pages 329-337.
- Nevalainen, T. (2006). Introduction to Early Modern English. Edinburgh University Press.
- Ogura, M. (1993). The development of periphrastic do in English: A case of lexical diffusion in syntax. *Diachronica*, 10(1), 51-85.
- Phillips, B. S. (1980). Old English an on: A new appraisal. *Journal of English Linguistics*, 14, 20-23.
- Phillips, B. S. (1983). ME diphthongization, phonetic analogy, and lexical diffusion. *Word*, 34(1), 11-24.
- Phillips, B. S. (1984). Word frequency and the actuation of sound change. *Language*, 60, 320–342.
- Pinheiro, J. C., and Bates, D. M. (2000). Mixed-effects Models in S and S-Plus. New York: Springer.

Qian, C. Y. (2011). 现代汉语"动宾式复合词带宾语"结构分析 (Structural

Analysis on VO+O construction in Modern Chinese) (Master's thesis), Shanghai: Fudan University.

- Qiu, R. T. (1994). 名动词质疑------评朱德熙先生关于名动词的说法 (Doubts on Nominal Verbs). *Chinese Learning*, 6, 15-20.
- Qiu, Z. P. (1990). 大陆和台湾差别词典 (Dictionary of Mainland and Taiwan Chinese). Jiangsu: Nanjing University Press.
- Qiu, Z. Q., and Fan, D. B. (1994). 台湾语言现状的初步研究 (A Preliminary Study on the Current Situation of Taiwan Language). *Zhongguo Yuwen*, 4, 254-261.
- Rao, C. R. (1984). 动宾组合带宾语 (On VO+O). Zhongguo Yuwen, 6, 413-418.
- Ren, H. B., and Wang, G. (2005). 基于语料库的现代汉语离合词形式分析 (The Analysis of Split Word in Modern Chinese Based on the Large scale Corpus). *Linguistics Sciences*, 4(6), 75-87.
- Richards, J, J. Platt, and H. Weber. (1985). Longman Dictionary of Applied Linguistics. Harlow: Longman.
- Robins, R. H. (1964). *General Linguistics: An Introductory Survey*. London: Longman.
- Rosen, S. T. (1990). Argument structure and complex predicates. New York: Garland.
- Ross, J. R. (1967). *Constraints on variables in syntax* (Doctoral Dissertation), Massachusetts: MIT.
- Samardzic, T. (2008). Light Verbs and the Lexical Category Bias of Their Complements (Doctoral dissertation, PhD thesis, Department of Linguistics, University of Geneva).

Schuchardt, H. (1885/1972). On sound laws: against the neogrammarians. In

Theo Vennemann, T., and Wilbur, T. (eds.), *Schuchardt, the Neogrammarians, and the Transformational Theory of Phonological Change*, 39-72.Frankfurt: Athen äum Verlag,

- Shan, B. S. (2007). 制约二价 VO 式动词价位的因素 (Factors that Affect the Valency of Bivalent Verbs). *The Northern Forum*, 6, 65-68.
- Shen, J. X., and Zhang, J. Z. (2013). Function of Form Verbs (也谈形式动词的 功能). *TCSOL*, 2, 8-17.
- Sheng, Y. Q. (1996). 海峡两岸同源异形计算机术语辨析 (Differentiation and Analysis of Computer Terms with the Same Origins but in Different Forms Used in Different Sides of the Taiwan Straits). *Terminology Standardization and Information Technology*. 19-20.
- Shi D. X. (2011). 名詞和名詞性成分 (Nouns and Nominals). Peking University Press.
- Shi, G. H., Li, X. J., and Li, X. (2003). 两岸现代汉语常用词典 (Modern Chinese Cross-Strait Dictionary). Beijing: Beijing Language and Culture University Press.
- Shi, M. Z. (1999). 述宾复合词的语法特点 (The Grammatical Property of VO compound). *Language Teaching and Linguistic Studies*, 1, 123-134.
- Simon, P., Huang, C. R., Hsieh, S. K., and Hong, J. F. (2007). Transliterated Named Entity Recognition Based on Chinese Word Sketch. *International Journal of Computer Processing of Oriental Languages* 21(1). 19-30.
- Sinclair, J. (1996). Preliminary recommendations on corpus typology. *EAGLES* Document TCWG-CTYP/P (available from http://www. ilc. pi. cnr. it/EAGLES/corpustyp/corpustyp. html).
- Tang, S. S. (2005). "进行,加以"带宾的句法语义考察 (Syntactic and Semantic Studies on Complement of 'jinxing, jiayi'). Journal of

Qinzhou Normal College, 2, 23-26.

- Tang, Z. X. (1995). 中国大陆,台湾,香港,新加坡汉语词汇方面若干差异举例 (Examples of Lexical Differences between Mainland China, Taiwan, Hong Kong and Singapore). *Journal of Xuzhou Normal University*, 1, 103-109.
- Tottie, G. (1991). Lexical diffusion in syntactic change: frequency as a determinant of linguistic conservatism in the. *Historical English syntax*, 2, 439.
- Tseng, H. I. (2003). 当代台湾国语的句法结构 (The Syntax Structures of Contemporary Taiwanese Mandarin) (Master's thesis), National Taiwan Normal University.
- Tsou, B. K., Yuen, R. W., Kwong, O. Y., La, T. B. Y., and Wong, W. L. (2005). Polarity classification of celebrity coverage in the Chinese press. In *Proceedings of 2005 International Conference on Intelligence Analysis*. http://analysis.mitre.org/proceedings/index.html (accessed 30 September 2011).
- Tsou, B. K., and Kwong, O. Y. (2006). Toward a Pan-Chinese thesaurus. the Fifth International Conference on Language Resources and Evaluation (LREC 2006), 2391–2394.
- Tu, Y. C., and Roth, D. (2011). Learning English light verb constructions: Contextual or statistical. In Proceedings of the Workshop on Multiword Expressions: from Parsing and Generation to the Real World. Association for Computational Linguistics.
- Vogt, H. (1954). Contact of languages. Word, 10, 365-74.
- Wang, H. D. (1997). "动宾式动词+ 宾语" 规律何在? (What is the pattern for VO+O construction). *Language Planning*, 8, 30-31.

- Wang, S., and Huang, C. R. (2011). Compound Event Nouns of the 'Modifierhead' Type in Mandarin Chinese. *In PACLIC* (pp. 511-518).
- Wang, W. S. -Y. (1969). Competing changes as a cause of residue. *Language*, 45, 9-25.
- Wang, W. S. -Y. (Eds.). (1977). *The lexicon in phonological change* (Vol. 5).Walter de Gruyter.
- Wang, W. S. -Y. (1979). Language change: a lexical perspective. *Annual Review* of Anthropology, 8(1), 353-371.
- Wang, T. K., and Li, X. J. (1996). 两岸词汇比较研究管见 (A Comparative Study of Lexical Differences between cross strait). *The World of Chinese Language*, 81, 24-31.
- Wang, W. S. -Y., Ke, J., and Minett, J. W. (2004). Computational studies of language evolution. *Computational linguistics and beyond*, 65-106.
- Wang, X. C., and Wang, Y. F. (2006). "动宾动词+宾语"结构及现象探略 (An Analysis of VO+O construction). Journal of Qingdao University of Science and Technology (Social Science Edition), 22(3), 116-120.
- Wei, X. M. (1984). *国语演变之研究 (Changes in the Mandarin Language in Taiwan)*. Taipei: National Taiwan University Press, 71-81.
- Wei, L., and Sheng, Y. Q. (2000). 大陆及港澳台常用词对比词典 (Comparative Dictionary of Commonly Used Words in Mainland China, Hong Kong, Macau and Taiwan). Beijing: Beijing University of Technology Press.
- Xiao, F. (1955). "进行" (On light verb jinxing). Chinese Knowledge.
- Xie, S. Y. (1990). "进行"类动词研究综述 (Review on Verbs like 'jinxing'). Journal of Ning Xia University, 3, 106-111.

- Xing, F. Y. (To appear). *全球华语语法研究 (Global Chinese Grammar)*. Beijing: Commercial Press.
- Xing, G. W. (1997). 一种似乎要流行开来的可疑句式——动宾式动词+ 宾语 (A dubious pattern in the process of popularity). *Language Planning*, 4, 21-23.
- Xu, D. H. (1995a). 两岸词语差异之比较 (Lexical Differences between Mainland and Taiwan Mandarin). *Proceeding of 1<sup>st</sup> symposium on Cross-strait Lexical and Character Differences*. 31-35.
- Xu, D. N. (1988). 双宾同指与双宾异指 (Co-referential and different references of Ditransitive Verb). *Language Teaching and Linguistic Studies*, 2, 35-45.
- Xu, X. R. (1995b). 海峡两岸新词语中若干词义衍生和规范的考察 (A Study on Derivation and Standardization of New Words in Cross-Strait). *Proceeding of 1<sup>st</sup> symposium on Cross-strait Lexical and Character Differences*. 228-238.
- Yang, H. M. (2001). "VO+ N" 与语义, 结构的兼容与冲突——汉语动宾组合带宾语结构中的语义问题 (The compatibility and conflict of the meaning and structure on the 'VO+N' structure). *Chinese Language Learning*, 1, 28-34.
- Yang, H. M., and Zhou J. (2006). 汉语语法的动态研究 (Dynamic Study on Chinese Grammar). Beijing: Beijing University Press.
- Yao, Y. (2014). Predicting the use of BA construction in Mandarin Chinese discourse: A modeling study with two verbs. *Proceedings of the 28th Paclic*. Phuket, Thailand.
- Yao, Y., and Liu F. H. (2010). A working report on statistically modeling dative variation in Mandarin Chinese. *Proceedings of the 23rd International*

- Yue-Hashimoto, A. (1993). 语法演变中的词窠: 汉语语法的词窠扩散 (The Lexicon In Syntactic Change: Lexical Diffusion In Chinese Syntax). *Journal of Chinese Linguistics*, 213-254.
- Zeng, R. F. (1995). 两岸语言词汇整理之我见 (Opinion on Coress-strait Language Differences). *Proceeding of 1<sup>st</sup> symposium on Cross-strait Lexical and Character Differences*. 1-11.
- Zhao, Y. F. (2008). 浅谈两岸三地同实异形词及其规范问题 (Study on the heterography from the Three Places Cross-Strait and the issue of Standardization). *Modern Chinese*, 3, 16-18.
- Zhao, C. L., and Shi, D. X. (2014). 两岸四地汉语"有信心"句式的异同 (Syntactic Difference of youxinxin between Mainland Chinese and Hong Kong, Macau and Taiwan Chinese). *Chinese Language Learning*, 2, 27-36.
- Zhang, B. (1999). "动宾结构十宾语"的条件及发展趋势 (The requirement and tendency of VO+O construction). *Research in Ancient Chinese Language*, 3, 2-6.
- Zhang, M., (2000). Syntactic change in Southeastern Mandarin: How does geographical distribution reveal a history of diffusion? In Memory of Professor Li Fang-Kuei: *Essays of Linguistic Change and the Chinese Dialects*. Taipei and Seattle: Academia Sinica and University of Washington, 197-242.
- Zhang, N. (1994). 港台与大陆书面语语法差异 (The Grammatical Differences between the written Chinese in Mainland China, Hong Kong and Taiwan). *Journal of Shangdong University*, 4, 31-38.
- Zhang, Y. S. (2010). 从错配到脱落: 附缀 "于" 的零形化后果与形容词, 动词 的及物化 (From reanalysis to dropping the consequences of the zero-form

of clitic yu and the transitivization of adjectives and intransitive verbs). *Zhongguo Yuwen*, 2, 135-145.

- Zhou, F. Q. (2009). 两岸三地外国人名地名翻译异同对比 (Translating English Proper Names into Chinese: Three Approaches in Chinese Mainland, Hong Kong and Taiwan). *Journal of Wanxi College*, 4, 131-134.
- Zhou, X. B. (1987a). "进行""加以"句型比较 (Sentence Pattern Comparison of 'jinxing' and 'jiayi'). *Chinese Language Learning*, 6, 1-5.
- Zhou, G. (1987b). 形式动词的次分类 (Subdivision of Dummy Verbs). Chinese Language Learning, 1, 11-14.
- Zhu, D. X. (1985). 现代书面汉语里的虚化动词和名动词 (Dummy Verbs and NV in Modern Chinese). Journal of Peking University (Humanities and Social Sciences), 5, 1-6.
- Zhu, H., Guo, J. X., and Chen, S. P. (2002). "VN+ O" 结构的语义认知考察. (The semantic and cognitive investigation on VN+O structure). *Linguistic Research(Special Issue)*, 1, 132-135.
- Zhu, J., and Sheng, X. H. (2008). "动宾结构带宾语"格式成因探究 (On Formation Causes of Structure VN+O). *Chinese Language Learning*, 3, 37-42.
- Zhu, J. S., and Zhou, W. W. (1990). 台湾"国语"词汇与普通话的主要差异 (The Lexical Differences between Taiwan Mandarin and Mainland Mandarin). *Journal of Anhui University*, 1, 91-100.
- Zou, J. Y., and Mo, Y. H. (2013). 港台书面语的历史与现状 (The Present and the Past of Written Hong Kong and Taiwan Chinese). In Feng, S. L (Eds.), 汉语书面语的历史与现状 (Written Chinese: the Present and the Past). Beijing: Peking University Press, 69-73.