Citation information:

Her, One-Soon. (2008) *Grammatical Functions and Verb Subcategorization in Mandarin Chinese*, Revised edition, Taipei: Crane Publishing

This book presents the first systematic classification of Mandarin verbs within a rigorously defined modern generative framework.

The classification is based on the subcategorized grammatical functions within Lexical-Functional Grammar (LFG). The following subcategorizable functions are identified in Chinese: SUBJ, OBJ, OBJ2, XCOMP, SCOMP, NCOMP, OBL_{LOCT}, OBL_{GOAL}, OBL_{THME}, OBL_{BNFC}. Twenty-six subcategories of verbs are identifed, and further subgroupings are made according to the different syntactic requirements of the subcategorized functions. In addition, four semantic subcategories are also distinguished along the distinction between action and state.

A number of the analyses of particular syntactic constructions differ from previous accounts. For example, topic is treated as a non-subcategorizable function. OBJ2, or the indirect object, is encoded as the second, rather than the first, postverbal NP, and [*ba* NP] is considered an oblique function. Also, *bei* is not a subject marker or a preposition; rather it is a verb subcategorizing for SUBJ, OBJ, and XCOMP.

The formalism developed here differs somewhat from the conventional LFG formalism and allows underspecified lexical entries. In this variant formalism, unification takes place at the same time as partial constituent structures are constructed. This allows earlier detection of functional structure violations, including incoherence of grammatical functions.

One-Soon Her has a Ph.D. in linguistics from the University of Hawaii and is now a Distinguished Professor of Linguistics at the National Chengchi University, where he has been teaching since 1992. He has published dozens of journal articles and is also the founder and co-editor of *Taiwan Journal of Linguistics*. He has another book published by Crane, entitled *Interaction and Variation in the Chinese VO Construction*.

文鶴出版有限公司
CRANE PUBLISHING CO.,LTD.
Book Number：002-0837

ISBN 978-986-147-273-7
00450
9 789861 472737

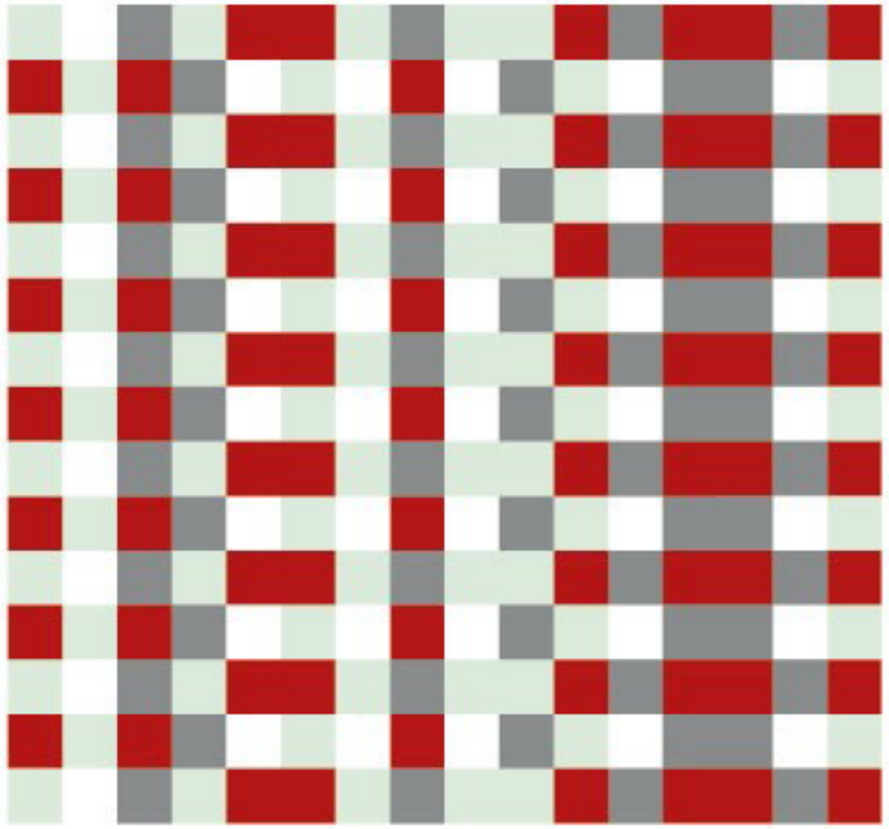
漢語研究01
Taiwan Journal of Linguistics
Book Series in Chinese Linguistics
Grammatical Functions and Verb Subcategorization in Mandarin Chinese
漢語中的語法功能及動詞分類
何萬順 One-Soon Her
Crane

漢語研究01
Taiwan Journal of Linguistics
Book Series in Chinese Linguistics

Grammatical Functions and Verb Subcategorization in Mandarin Chinese

漢語中的語法功能及動詞分類

何萬順 | One-Soon Her



Taiwan Journal of Linguistics
Book Series in Chinese Linguistics
Editors: One-Soon Her & Kawai Chui

GRAMMATICAL FUNCTIONS AND VERB SUBCATEGORIZATION IN MANDARIN CHINESE
漢語中的語法功能及動詞分類

One-Soon Her
何萬順

To the memory of my parents,
HER Yu-Shou 何於壽 & KUO Wen-Ying 郭文英

PREFACE TO THE REVISED EDITION

Since the book was first published in early 1991, I have received many comments, corrections, and suggestions. I myself have likewise noticed many typos. Also, back then the book was rather crudely typeset in the DOS version of WordPerfect. Therefore, when Angela Chen, one of Crane's able editors, and Crane's vice-president Paul Yang approached me about doing a revised edition, I was delighted. So they are the first two people I must thank, along with Fu-Gong Chang, Crane's president. I thank Mr. Chang for his continuous trust and support. He has successfully established Crane to be the country's most important bookseller and publisher in the field of linguistics and is thus to be credited for helping to promote the study of linguistics in Taiwan.

This revised edition has thus undergone quite an overhaul in formatting and style. For this I thank my assistants Li-Hsin Ning and Yu-Ying Ho. Li-Hsin's excellent skills with the word-processor Word, in particular, are largely responsible for the book's fresh new look. My wonderful friend Karen Chung has found many typos for me, and likewise my other assistants, Yi-Ting Hsieh, Wan-Ju Huang, Hsiao-Chien Feng, and Guang-Chung Wu.

The content of this revised edition, on the other hand, has not changed from the previous edition in any substantial way. The original verb subcategorization scheme proposed within Lexical-Functional Grammar, or LFG in short, in the book has since been reviewed extensively, from a Government and Binding (GB) perspective, by Feng-Fu Tsao in his 1996 Journal of Chinese Linguistics article, 'On verb classification in Chinese', where he has also proposed his own classification of Mandarin verbs under a GB/LFG hybrid framework. Meanwhile, a number of other important works also appeared, especially out of the Mainland China, dealing with the same general topic. The LFG framework, in the last two decades, has also made some significant advances and gained various new insights. To incorporate the advances in the LFG grammatical theory as well as in our knowledge of Mandarin verbs, I am now writing a new book, tentatively entitled *Grammatical Classification of Mandarin Verbs*, which should be published by Crane in 2010. However, this current book remains the first systematic classification of Mandarin verbs within a rigorously defined

modern generative framework and as such, I believe, is worth the attention of theoretical linguists, Chinese grammarians, as well as lexicographers.

I dedicate this revised edition to the loving memory of my parents, Yu-Shou Her and Wen-Ying Kuo.

ACKNOWLEDGEMENTS

This book is a revised version of my doctoral dissertation. I often think that the writing of a dissertation, and the making of a PhD for that matter, is more a collective effort than most people realize. I certainly have many people to thank, who have helped me in different ways in writing and publishing this work. The first of them is no doubt Professor Robert Cheng, my dissertation adviser, who has graciously undertaken this laborious task and has provided me with numerous valuable suggestions, criticisms, and encouragements. Throughout my study at the University of Hawaii, he has always expressed his interest in my work and has always been generous in building up my confidence. The one teacher that has singularly influenced me the most in shaping my conception of how linguistics as a scientific discipline should be pursued is Professor Stanley Starosta. I certainly hope that this influence is evident in the pages that follow. Professor Hsin-I Hsieh, on the other hand, has always inspired me to view linguistics in the larger picture of life. I thank him for his insights and the many stimulating conversations we had. Professor John Haig has carefully read the first draft of the dissertation and given me many detailed constructive comments and useful suggestions. Professor David Stampe opened the door into the maze of computational linguistics for me, and that definitely led to a drastic change in my life, for the better I hope. I also enjoy his style of interacting with people. I hereupon express my gratitude to all these teachers for their positive influence and generous help. Thanks also due to Dr. Chu-Ren Huang of Academia Sinica for his many insightful comments and Professor Chin-Tang Lo for being the outside member of the dissertation committee.

My special thanks go to Dr. Byron Bender, the chairman of the Department of Linguistics, who has provided me essential academic guidance over the years, to Claire Chang, who has very kindly helped me take care of so many administrative matters and has been a wonderful friend by being always supportive to me, to Professor Y. C. Li, an alumnus who has always been kind and encouraging to me and always showed interest in my work, and to my colleague Dr. Larry Browning, who has graciously read and provided detailed comments on most of my papers, including this dissertation, in terms of style and grammar. Also, I should not be thankless to Dr. Thomas Deterich of Portland State University, who turned me on to generative linguistics and

showed me what a good taste of syntactic argumentation should be. I want to thank Wendy, the secretary of the Linguistics Department of UH, for her kindness and help in the past six years.

For more than four years, I have been working for the company ECS, Inc. in Provo, Utah, developing machine translation systems. The president of the company, Mr. Larry Gibson, has been tremendously generous to me and most supportive to my efforts to complete my degree. I am sincerely grateful. I also want to extend my appreciation to my friends and colleagues at ECS, Dr. Larry Browning, Dr. Joseph Pentheroudakis, Dr. Dan Higinbotham, Hajime Wada (ABD), Jay Kim (ABD), Debby Masterson (ABD), Paul Warnick, M.A., Dave Rose, Reid Poole, Jim Hughes, Brice Poole, Kimbal Wirig, and Pat Reece. They are a wonderful group of people to work with. Many of the ideas in this dissertation were generated through my work experience here and conversations with my colleagues. I thank them very much for their emotional as well as intellectual support.

Finally, there is my family; without them, life itself would not mean much.

ABSTRACT

This book presents a classification of Mandarin verbs by the subcategorized grammatical functions within Lexical Functional Grammar (LFG). A brief description of LFG is provided. The following functions are identified as subcategorizable in Mandarin: SUBJ, OBJ, OBJ2, XCOMP, SCOMP, NCOMP, OBL_{LOCT}, OBL_{GOAL}, OBL_{THME}, OBL_{BNFC}, and the syntactic encoding of these functions is discussed. Twenty-six subcategories are identified, and further subgroupings are made when different syntactic requirements of the subcategorized elements are observed. Additionally, four semantic subcategories are distinguished along the traditional distinction of action versus state.

A number of our analyses of particular syntactic constructions in LFG differ from previous accounts. We consider topic a syntactic notion and argue against its subcategorizability in Mandarin. In relation to treating OBJ2, or indirect object, as encoded by the second, rather than the first, postverbal NP, we insist that [ba3 NP] should not be taken as object, rather it should be considered as an oblique function. To name one more, we reject bei4's status as either a subject marker or a preposition and analyze it as a verb requiring a pivotal construction subcategorizing for SUBJ, OBJ, and XCOMP.

To improve efficiency in computational processing and linguistic generalizations, the formalism in which our analyses are formulated differs than the conventional LFG formalism in the notations of phrase structure rules and the feature inheritance structure. The later makes under-specified lexical entries and thus a lexicon of reduced size possible. Also in our variant formalism unification takes place at the same time when partial constituent structures are constructed. This allows earlier detection of functional structure violations, including incoherence of grammatical functions.

While previous classifications of Mandarin verbs are often based on case roles and constituents, our classification is based on the notion of subcategorization for grammatical functions. Given the recognition of notions of subcategorization and grammatical relations in general linguistics, our study fills the gap in Chinese syntax due to the lack of a systematic, extensive account of Mandarin verbs concerning subcategorized grammatical functions.

TABLE OF CONTENTS

PREFACE TO THE REVISED EDITION V

ACKNOWLEDGEMENTS VII

ABSTRACT IX

LIST OF TABLES XV

LIST OF FIGURES XVI

LIST OF ABBREVIATIONS AND SYMBOLS XVII

PREFACE XXI

CHAPTER 1 LEXICAL FUNCTIONAL GRAMMAR AND A VARIANT FORMALISM 1

1.1 CHINESE LINGUISTICS AND MODERN LINGUISTIC THEORIES 1
1.2 LFG: A HISTORICAL AND CONTRASTIVE PERSPECTIVE 3
1.2.1 The Revolution: The Transformational Generative Grammar 4
1.2.2 The Second Stage: The Standard Theory 4
1.2.3 The Third Stage: Government and Binding Theory 5
1.2.4 Lexicase and the Formal Rigor 6
1.2.5 LFG and the Emphasis on Processing 6
1.2.6 Points of Convergence 7
1.3 LFG: A SKETCH 9
1.3.1 The Lexicon 9
1.3.2 The C-structure and Phrase Structure Rules 10
1.3.3 Unification: from C-structure to F-structure 12
1.3.4 From Form to Meaning 14
1.3.5 Further Readings 15
1.4 VLFG: A VARIANT LFG FORMALISM 16
1.4.1 The Informational Organization of the Lexicon 17
1.4.2 The Phrase Structure Rules 21
1.4.3 Unification: from C-structure to F-structure 23

CHAPTER 2 GRAMMATICAL FUNCTIONS IN AN LFG GRAMMAR OF MANDARIN CHINESE 29

2.1 GRAMMATICAL FUNCTIONS AND F-STRUCTURE 29
2.2 GRAMMATICAL FUNCTIONS AND THEMATIC ROLES 30
2.3 SYNTACTIC ENCODING OF GRAMMATICAL FUNCTIONS 31
2.4 SUBCATEGORIZATION IN LFG 33
2.5 CLASSIFICATION OF GRAMMATICAL FUNCTIONS 36
2.5.1 Subcategorizability 36
2.5.2 Semantic Restriction 38
2.6 SUBCATEGORIZABILITY OF GRAMMATICAL FUNCTIONS AND WELL-FORMEDNESS CONDITIONS ON F-STRUCTURE 39
2.7 GRAMMATICAL FUNCTIONS AS PRIMITIVES 45
2.8 THE STATUS OF TOPIC AND SUBJECT 45
2.9 SUBCATEGORIZABILITY OF TOPIC IN CHINESE 51
2.10 SYNTACTIC ENCODING OF SUBJ AND OBJ 61
2.11 OBJ2 AS A SEMANTICALLY RESTRICTED FUNCTION 62
2.12 COMP FUNCTION AND SUBTYPES 65
2.13 OBLIQUE FUNCTION AND SUBTYPES 66
2.14 SYNTACTIC ENCODING OF GRAMMATICAL FUNCTIONS IN CHINESE 67

CHAPTER 3 VERB SUBCATEGORIZATION IN MANDARIN CHINESE 69

3.1 EVIDENCE FOR SUBCATEGORIZED ELEMENTS 69
3.1.1 Obligatory Co-occurrence 70
3.1.2 Ontological Necessity 71
3.1.3 Constancy of Semantic Content 72
3.1.4 Distributional Restrictions 73
3.1.5 One per Sentence 74
3.1.6 Proximity to the Head 76
3.1.7 Semantic Selectional Restrictions 76
3.1.8 Possible Internal Gaps 76
3.1.9 Ability of Being an Indirect Question 77
3.1.10 Identification of a Subcategory 78
3.2 MANDARIN VERB SUBCATEGORIZATION IN VLFG 78
3.2.1 Activity and Stativity 79
3.2.2 Syntactic Subcategorization 85

3.2.3 Optionality in Subcategorization 88
3.2.4 Derivational Relations between Verb Classes 89
3.3 SUBCATEGORIZATION OF GRAMMATICAL FUNCTIONS 93
3.3.1 <SUBJ> 93
3.3.1.1 Intransitive Verbs 93
3.3.1.2 Weather Verbs 96
3.3.2 <SUBJ , OBL_{LOCT}> 97
3.3.3 <SUBJ , OBL_{GOAL}> 100
3.3.4 <SUBJ , OBJ> 103
3.3.4.1 Typical Transitive Verbs 103
3.3.4.2 Verbs Requiring a Locative OBJ 106
3.3.4.3 Locative Inversion Verbs 108
3.3.4.4 You3 'exist, have' 109
3.3.4.5 Verbs Requiring TOPIC 125
3.3.5 <SUBJ , OBL_{THME}> 126
3.3.5.1 Syntactic Category of ba3 127
3.3.5.2 The Double Direct Object Fallacy 128
3.3.5.3 Semantic Restrictions of [ba3 NP] 130
3.3.6 <SUBJ , OBL_{THME} , OBJ> 135
3.3.7 <SUBJ , OBL_{GOAL} , OBJ> 141
3.3.8 <SUBJ , OBJ , OBL_{LOCT}> 142
3.3.9 <SUBJ , OBL_{THME} , OBL_{LOCT}> 143
3.3.10 <SUBJ , OBL_{THME} , OBJ , OBL_{LOCT}> 144
3.3.11 <SUBJ , NCOMP> 145
3.3.11.1 Equational Verbs 145
3.3.11.2 Verbs Denoting Inherent Quality 149
3.3.12 <SUBJ , XCOMP> 151
3.3.12.1 Regular Type 151
3.3.12.2 Tough Construction 154
3.3.12.3 Modal Verbs 159
3.3.12.4 Clause Union Verbs 166
3.3.13 <SUBJ , SCOMP> 169
3.3.13.1 Regular Type 169
3.3.13.2 Verbs Requiring an Interrogative SCOMP 170
3.3.14 <SUBJ , OBJ , OBJ2> 173
3.3.14.1 Regular Type 173

3.3.14.2 Idiomatic Expressions 174
3.3.15 <SUBJ , OBJ , OBL_{BNFC}> 176
3.3.16 <SUBJ , OBL_{THME} , OBJ , OBJ2> 177
3.3.17 <SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}> 179
3.3.18 <SUBJ , OBL_{THME} , OBL_{BNFC}> 180
3.3.19 <SUBJ , OBL_{GOAL} , OBL_{THME}> 181
3.3.20 <SUBJ , OBL_{GOAL} , OBL_{THME} , OBJ> 182
3.3.21 <SUBJ , OBJ , XCOMP> 183
3.3.21.1 Regular Type 183
3.3.21.2 Clause Union Verbs 186
3.3.21.3 "Promise" Verbs 189
3.3.21.4 Tough Construction 191
3.3.21.5 Bei4 193
3.3.21.5.1 Existing Analyses for Chinese bei4 194
3.3.21.5.2 Dismissal of bei4 as a Subject Marker 196
3.3.21.5.3 Dismissal of bei4 as a Preposition 197
3.3.21.5.4 Bei4 as a Verb 202
3.3.21.5.5 Conclusion 220
3.3.22 <SUBJ , OBJ , SCOMP> 221
3.3.22.1 Regular Type 221
3.3.22.2 Verbs Requiring an interrogative SCOMP 221
3.3.23 <SUBJ , OBL_{GOAL} , XCOMP> 222
3.3.23.1 Regular Type 222
3.3.23.2 Verbs Requiring an interrogative SCOMP 224
3.3.24 <SUBJ , OBL_{GOAL} , SCOMP> 225
3.3.24.1 Regular Type 225
3.3.24.2 Verbs Requiring an Interrogative SCOMP 226
3.3.25 <SUBJ , OBJ , OBJ2 , SCOMP> 226
3.3.26 <SUBJ , OBL_{GOAL} , OBJ , SCOMP> 227
3.4 SUMMARY OF SYNTACTIC SUBCATEGORIZATION OF VERBS 228
3.5 COMPARISON WITH PREVIOUS CLASSIFICATIONS 230

CHAPTER 4 GRAMMATICAL THEORIES AND NATURAL LANGUAGE PROCESSING 237

4.1 LEXICASE, GOVERNMENT AND BINDING, AND LFG 239
4.2 TASKS OF NATURAL LANGUAGE PROCESSING 241

4.3 PARSING NATURAL LANGUAGE 244
4.4 NATURAL LANGUAGE GENERATION 250
4.5 MACHINE TRANSLATION 253
4.6 SUMMARY 256

CHAPTER 5 CONCLUSION 257

5.1 THE vLFG FORMALISM 257
5.2 MANDARIN VERB SUBCATEGORIZATION IN LFG 259
5.3 SUGGESTED AREAS FOR FURTHER RESEARCH 262

BIBLIOGRAPHY 265

APPENDIX FEATURE INHERITANCE ENTRIES 277

LIST OF TABLES

Table 3.1 Semantic Subcategories of Mandarin Verbs..... 80
Table 3.2 Classification of Mandarin Modal Verbs 165
Table 3.3 Syntactic Classification of Mandarin Verbs 228
Table 3.4 Chao's Classification of Mandarin Verbs (Chao 1968)..... 231
Table 3.5 Li's Classification of Mandarin Verbs (Li 1971) 232
Table 3.6 CKIP's Classification of Mandarin Verbs (CKIP 1989)..... 234

LIST OF FIGURES

Figure 1.1 Co-description of C- and F-structure..... 14
Figure 1.2 Correspondence of Linguistic Representations in LFG 15
Figure 2.1 Correspondence between Conceptual Structure and Linguistic Structures..... 31
Figure 2.2 Lexical and Syntactic Encoding of Grammatical Functions..... 33
Figure 2.3 Classification of grammatical functions in Mandarin 38
Figure 3.1 Classification of Grammatical Functions..... 91
Figure 3.2 Possible Analyses for Chinese bei4..... 195
Figure 4.1 NLP Application of Linguistics 241
Figure 4.2 Correa's Variant GB Model (Correa 1987)..... 245

LIST OF ABBREVIATIONS AND SYMBOLS

A	Attributive Adjective in Chinese
ADJ	Adjunct Function in LFG
ag	agent
ATN	Augmented Transition Network
ben	beneficiary
BNFC	beneficiary
CAT	Category
CLS	Classifier
COMP	Inventory of Predicative Complements in LFG
exp	experiencer
FI	Feature Inheritance Entry
FS	Functional Structure
FUG	Functional Unification Grammar
GB	Government and Binding Theory
go	goal
GPSG	Generalized Phrase Structure Grammar
HPSG	Head-driven Phrase Structure Grammar
inst	instrument
LFG	Lexical Functional Grammar
loc	locative
LOCT	location
LXC	Lexicase Theory
MT	Machine Translation
N	Noun
NCOMP	Predicative Noun Complement in LFG
NLP	Natural Language Processing
NP	Noun Phrase
OBJ	Object Function in LFG
OBJ2	Second Object Function in LFG
OBL_{Θ}	Inventory of Oblique Functions in LFG
P	Preposition
PAS	Predicate Argument Structure
pat	patient
PRED	Predicates

PTCL	Particle
PP	Prepositional Phrase
PSR	Phrase Structure Rule
QTFR	Quantifier
RELN	(Semantic) Relation
RG	Relational Grammar
S	Sentence
SCOMP	Sentential Complement Function in LFG
SUBJ	Subject Function
TG	Transformational Grammar
th	theme
THME	theme
TOPIC	Topic Function in LFG
V	Verb
vLFG	the variant LFG Formalism
VP	Verb Phrase
XCOMP	Open Complement Function in LFG
XADJ	Open Adjunct Function in LFG
↑	the level of f-structure corresponding to my mother node of the c-structure
↓	the level of f-structure corresponding to my own node of the c-structure
=	unify with
≈	unify by default with
$=_c$	must be identical with
:	unify to be (usually a subcategorizable grammatical function)
ε	unify to be a member of (usually an adjunct or conjunct)
()	optional constituent (in phrase structure rules); OR path (in f-structure)
{}	disjunction (in phrase structure rules and schemata); OR conglomerated list (in f-structure)

X+	one or more instance of X (in c- or f-structure)
X*	zero or more instance of X (in c- or f-structure)
<XY>	X and Y in any order in vLFG PSR's; OR in LFG formalism, X and Y as subcategorized, thematically assigned functions, as part of the value of PRED
@	what preceded is a non-thematically assigned function in vLFG

PREFACE

This book first provides a brief description of the Lexical Functional Grammar (LFG) and a variant formalism of a unification grammar that most closely follows the theoretical constructs and formalism of LFG (Bresnan 1982, Sells 1985, Wescoat 1987, Kaplan and Zaenen 1987), including LFG's most distinctive feature: the division of constituent structure and functional structure. Within the LFG theory we will discuss the various grammatical functions in the grammar of Mandarin Chinese, and then, within this variant LFG formalism (vLFG), we will systematically present the various subcategories of verbs in Mandarin Chinese that require different subcategorization of grammatical functions. Our scheme of classification of Mandarin verbs within the vLFG formalism is therefore according to the grammatical functions that verbs subcategorize for. We shall argue for our analysis of grammatical functions and their subcategorizability in an LFG grammar of Mandarin Chinese. We shall also, in a relatively theory-independent manner, argue for our analysis of certain types of verbs of which there exist previous different accounts or whose syntactic constructions are known to be controversial in the literature of Chinese linguistics.

Concerning the Mandarin data in this book, we should clarify that, in order to illustrate precisely and concisely the relevant points, we have selected example sentences that contain the minimum necessary elements. The judgement on the grammaticality or acceptability is primarily the author's, a native speaker of Mandarin Chinese college-educated in Taiwan whose parents are from the central and northern regions of China. Most of the several other native speakers of Mandarin who have read previous versions of the book also confirmed the author's judgement. Furthermore, most of the sentences we use in the discussions do comport with examples cited in other works in Chinese linguistics (e.g., Chao 1968 and Li and Thompson 1981).

In CHAPTER 1, we will briefly review the historical development of Chinese linguistics in applying modern linguistic theories and the current state of Chinese linguistics in relation to contemporary grammatical theories and computational applications in processing Mandarin Chinese. We will briefly discuss the theory of LFG and its motivation in the context of the developmental stages of Chomsky's generative syntax of the past three decades and also in contrast with two other contemporary grammatical

theories, Government and Binding Theory (GB) and Lexicase Theory (LXC). We will then sketch the theoretical constructs and formal notations of LFG as it is conventionally formulated. We then present the vLFG formalism employed in this book. We shall see that although the vLFG and the conventional LFG are very close in almost all aspects of linguistic theoretical constructs, the major differences between them include the timing of the building of the functional structure, the informational organization of the lexicon and lexical entries, and the notation of lexical entries and phrase structure rules. The vLFG formalism allows a feature inheritance structure in the lexicon, and unlike conventional LFG that builds a functional structure only after the entire constituent structure is built, the vLFG builds a partial functional structure whenever its corresponding constituent is properly formed. We will describe in detail the formalism and notations of grammar writing within the vLFG.

In CHAPTER 2, we will first discuss what we consider grammatical functions in LFG and the different types among them. We will show how grammatical functions are lexically and syntactically encoded in a grammar. More specifically, we will identify a set of grammatical functions for the grammar of Mandarin Chinese and also justify our classification of subcategorizable and non-subcategorizable (adjunctive) functions in Mandarin. Among the subcategorizable functions, we will identify the semantically restricted ones versus the semantically unrestricted ones. We will devote a more lengthy discussion on the status of subject and topic, especially regarding the subcategorizability of TOPIC in Mandarin Chinese. The following subcategorizable functions are identified: SUBJ, OBJ, OBJ2, XCOMP, SCOMP, NCOMP, OBL_{LOCT}, OBL_{GOAL}, OBL_{THME}, OBL_{BNFC}. We will also discuss the syntactic encoding of these functions.

CHAPTER 3, the core of this book, first discusses the notion of subcategorization and the generally-accepted criteria for distinguishing subcategorized constituents from adjuncts. Then, we will present a semantic classification of Mandarin verbs based on two binary features of activity and process. After giving some reasons of not allowing optional arguments in subcategorization and introducing an example of morpholexical rule in LFG, we proceed to present our classification of Mandarin verbs according to the grammatical functions they subcategorize for. We will discuss each subcategory of verbs and argue for our analysis for certain verb classes when

it differs from previous accounts. Noticeably, more detailed discussion is devoted to the following elements that have been known to be controversial: ba3, bei4, you3, and shi4 etc. Twenty-six subcategories are identified, and within several of them further subgroupings are made when differences in syntactic requirements of the subcategorized elements are observed. Finally we will offer a brief comparison of our classification schemes with three other existing ones, Chao (1968), Li (1971), and CKIP (1989).

In CHAPTER 4 we will discuss the practical application and theoretical implication of the study in this book. We shall argue that the vLFG grammar has been shown to be suitable for the description of Mandarin verbs and that while its theory is generally constrained in similar ways as the LFG theory, its formalism is expressive and thus appropriate for explicit formulation of linguistic generalizations for computational applications. We shall examine in what ways LFG is relevant to the various tasks under the general domain of Natural Language Processing (NLP), such as parsing, generation, and machine translation. In contrast, we will also discuss the two other grammatical theories, Lexicase and Government and Binding Theory, regarding their NLP application. We will argue for the advantage of LFG, especially in the vLFG formalism we are adopting in this study, for computational applications of NLP.

Finally, in CHAPTER 5, we will briefly reiterate the contributions we hope this book will make to the study of Mandarin Chinese and the study of LFG in terms of our subcategorization of Mandarin verbs, accounts of certain controversial syntactic constructions, and the several alternatives in the consideration of grammar formalism and processing procedure. We will point out areas for further research in order to gain a more complete account of Mandarin verbs in LFG and conclude the book.

CHAPTER 1
LEXICAL FUNCTIONAL GRAMMAR AND A VARIANT FORMALISM

In this chapter we will review briefly the development of linguistic approaches for analyzing Mandarin Chinese, and we will discuss briefly the current state of Chinese linguistics in relation to current linguistic theories and computational application. We will argue for the importance for linguists to work within and contribute to a linguistic framework that is truly "generative" in nature. We will then look at the development and the primary motivation of LFG in the context of the developmental stages of Chomsky's generative theory of language. Also, to provide different perspectives, we will compare LFG with two other current linguistic theories, GB and LXC in the context of these developmental stages. We will sketch the theoretical constructs and formalism of LFG as it is conventionally formulated. Then we will describe the variant formalism of LFG in which we will formulate our analyses. For ease of discussion, we will use the term "vLFG" to refer to our variant formalism, one that is different in the organization of the lexicon, notation of phrase structure rules and the timing of the building of functional structure.

1.1 Chinese Linguistics and Modern Linguistic Theories

For more than half a century Chinese linguists have applied modern linguistic theories to the analysis of Chinese languages, and most intensively to Mandarin Chinese, the language that has the largest number of native speakers on earth as reported by the 1989 World Almanac. Since almost all modern linguistic theories claim to follow western scientific principles and are developed out of the west, the application of new, innovative theories to Mandarin Chinese always comes noticeably more slowly, compared to their application to western languages. The traditional, indigent school of philology and the study of phonology, especially in the area of rhyme, existed in China for centuries before the western Structuralist school established linguistics as an independent discipline in modern sciences early in this century. This rich resource of traditional works has provided a solid basis for Chinese historical

linguistics and dialectal studies and still occupies an important position in Chinese linguistics.

The Structuralist framework dominated the modern linguistic scene of the study of Chinese from the late 1940's well into the 1970's, almost two decades after Chomsky revolutionized the study of language in the late 1950's with his generative theory of linguistics. In fact, even today a large number of Chinese linguists, especially in Mainland China, still publish papers in a general linguistic framework best described as Structuralist (e.g., Zheng and Chen 1989 and Yang 1989). Structuralist linguists have contributed greatly to the study of Mandarin Chinese; among them there is the most influential figure in Chinese linguistics to date, the late Dr. Y.R. Chao, whose monumental work laid the foundation for the study of modern Chinese.

Many Chinese linguists worked diligently within the framework of Transformational Grammar, including its later version the Standard Theory, advocated by Chomsky as the first generative theory, in its heyday of the 60's and 70's (e.g., Huang 1966 and Li 1972). However, their number dwindled in the 80's as the Chomskyan school of linguistics came to be highly abstract and deviated from Chomsky's original vision of generative linguistics. In the meantime, alternative theories to the mainstream Chomskyan practice flourished, reacting to the inadequacies that many theorists perceived in the conventional Chomskyan theory. Many of these alternative theories insist strongly upon their "generative" nature, in its original sense of "formal" and "explicit" intended by Chomsky in his revolutionary work. Although the Chomskyan school of Government and Binding Theory is still often perceived as the mainstream and the application of the alternative theories in Chinese has come slowly, several of them have been applied in the analysis of Mandarin Chinese; they include Case Grammar (e.g., Li 1971 and Teng 1975), Lexicase (Starosta 1985 and Her 1985-6), Categorial Grammar (e.g., Liu 1987 and Zhang 1989), Generalized Phrase Structure Grammar (e.g., Huang 1987 and Sheu 1987), and Lexical Functional Grammar (e.g., Huang 1989 and Her 1989b).

Having abandoned Transformational Grammar, which has been recognized as excessively powerful and thus non-revealing, many Chinese linguists, who followed Chomsky's earlier theory, stayed outside of the GB camp, and, like many other Chinese linguists, often resorted to analyzing Chinese linguistic data in a most general, undefined framework of linguistics, or one that

Starosta (1988) describes as "eclecticism." Truly, if linguistics is to be recognized as a discipline of modern science, it is crucial, or even mandatory, one might argue, for its workers to contribute collectively in a theoretically well-defined framework within this discipline. Furthermore, as a scientific pursuit, a sound linguistic theory has to adhere to the scientific methodology known as the hypothetico-deductive method (Starosta 1987). Logically, only a genuine generative linguistic theory where generalizations must be made formally and explicitly is falsifiable and thus compatible with the hypothetico-deductive spirit.

The rapid advancement of computer technology and soaring demand of natural language processing in recent years have provided a serious testing ground for linguistic theories and analyses. An informal or inexplicit linguistic theory or analysis cannot survive the scrutiny of computational applications. Although, unfortunately, many projects of natural language processing do not utilize any well-established linguistic theory for their linguistic analysis, in recent years LFG has been gaining increasing popularity in computational applications, especially in the Republic of China (Huang and Chen 1989). We will discuss in greater details the relevance of linguistic theories in computational applications in CHAPTER 4.

In terms of linguistic analysis, several areas of Chinese grammar have already been explored in the LFG framework, including Mandarin [NP-de] constructions (Huang 1987), sentence particles and questions (Shiu 1989, Shiu and Huang 1989), relative clauses (Hu 1989), topicalization (Chen 1989), bei4 sentences (Her 1989), lexical discontinuity (Huang 1988), and the linking relations between thematic roles and grammatical functions in Chinese (Huang 1989, 1989a). This present study of grammatical functions and verb subcategorization, which has been preliminarily reported in Her (1989a), will constitutes yet another significant step in the application of LFG in Mandarin Chinese.

1.2 LFG: A Historical and Contrastive Perspective

Lexical Functional Grammar, a generative lexicalist unification grammar theory, was first systematically introduced by Kaplan and Bresnan (1982). Like almost all other contemporary generative grammatical theories, LFG has its roots deeply embedded in Chomsky's early generative syntax and yet was

developed as an improvement of and a reaction to some of the inadequacies that the LFG theorists observed in the directions that the mainstream Chomskyan grammarians chose to follow. To understand the motivation and the development of LFG, it should reveal meaningful insights to first look at LFG within the context of the developmental stages of Chomsky's generative syntactic theory and in contrast with other contemporary grammatical theories. Among other current syntactic theories we choose to compare LFG with Lexicase (LXC) and Government and Binding Theory (GB), which provide different perspectives than LFG and thus serve well as contrast.

1.2.1 The Revolution: The Transformational Generative Grammar

Nearly all contemporary generative syntactic theories share a common ancestry: Chomsky's revolutionary work of generative transformational syntax of the late 1950's. We believe each of the three theories, in the areas it chooses to emphasize, represents a different reaction to, or extension of, Chomsky's earlier interpretation of syntax. From late 1950's to the present, Chomsky's syntactic theory has roughly undergone three perceivable developmental phases. Syntactic Structures (Chomsky 1957) revolutionized syntactic theorizing and trumpeted the advent of the era of generative grammar, and also firmly established the study of linguistics as a scientific pursuit. The proposal was that the objective of a grammar is to "generate" all and only the infinite grammatically acceptable strings of a natural language. Thus the key point is that grammars should be 'generative', in a mathematical sense. Therefore, the criterion of explicit, formal and falsifiable formulation of linguistic statements and generalizations was greatly emphasized. The claim was that this goal can be obtained with a transformational grammar but not with a phrase structure grammar nor the earlier Structuralist approach to language.

1.2.2 The Second Stage: The Standard Theory

The 'Standard Theory' of Aspects of the Theory of Syntax (Chomsky 1965) posited two levels of syntactic representation: a deep structure as the basis for meaning interpretation and a surface structure as the basis for phonological interpretation. The crucial linking between the deep and surface structure is

accomplished by transformations. The focus of syntactic theory was that how a grammar should model the mental mapping between meaning and sound, and thus linguistic analyses should be psychologically real. The requirement of explicitness and formalness began to lose its earlier visibility.

The Standard Theory soon divided into several revised and extended versions by the late 1970's. The most salient and probably also most important trend of syntactic research was the restraining of powerful theoretical constructs such as transformation rules. Within the Chomskyan camp of transformational grammar the greatest effort has been in the elimination of various structure-specific transformations; however, layers of highly abstract constraints have to be devised to allow for one single general transformation: Move-α. Other theories of different approaches, including LXC and LFG, on the other hand have totally ruled out the validity of transformations theoretically and treat syntax as a purely surface phenomenon. As a matter of fact, LXC, dating back to the early 1970's, was probably the first generative theory of syntax that was entirely transformationless.

1.2.3 The Third Stage: Government and Binding Theory

Lectures on Government and Binding (Chomsky 1981), introducing the Government and Binding Theory, embodied research within the Standard Theory and extended models concerned with the constraining of transformations and the attaining of the explanatory power of how only grammars learnable based on the primary data should be allowed in the theory of language. In other words, GB attempts to provide a theory where a grammar of a natural language can be inferred through a set of universal principles and the setting of certain universal parameters. In the pursuit of linguistic parameters and universal grammar, standards of explicit, formal, and detailed formulation of analysis of specific syntactic constructions waned and were even reproved. While there is a considerable amount of GB cross-language research on parameters of language variation and universal principles, one rarely finds an explicit formulation of an analysis of a specific syntactic construction of a particular language.

1.2.4 Lexicase and the Formal Rigor

While GB has compromised considerably the standards of formal explicitness and psychological realism, LXC and LFG to a significant degree represent serious efforts to reemphasize some of the worthwhile characteristics of earlier phases of Chomsky's generative theory. Both LFG and LXC claim their respective linguistic theory to be a universal model for all languages and therefore do stress upon the standard of explicit rigor, psychological reality, and the universality of their theory. LFG however has strong emphasis on the computational and psycholinguistic processing of language, and the LXC literature has demonstrated a most serious commitment to the formal and explicit formulation of linguistic generalizations. One of LXC's strongest objection to GB is that it is not clear whether GB can still be considered 'generative' in the original sense intended by Chomsky in the first stage of the late 1950's. Starosta, the primary theorist of LXC who had substantial training in physics, devoted an entire volume readdressing the issue of 'generative grammar' as a hypothetico-deductive science (Starosta 1987), and on numerous occasions repeated the generative aspect of LXC and that the goal of a LXC grammar is to generate all and only the acceptable phrases of which sentences are a subset. The LXC literature is therefore largely composed of detailed, explicit accounts of grammatical phenomena of various natural languages, most of which are non-IndoEuropean.

1.2.5 LFG and the Emphasis on Processing

LFG's concern for the processing aspect of language can no doubt be partially attributed to the two primary architects of the theory, Kaplan and Bresnan. The formal conception of LFG evolved in the mid-1970's from earlier work in both Transformational Grammar and computational linguistics. Kaplan was a psychologist and did experimental work on human sentence processing and computational natural language processing. He was one of the designers of Augmented Transition Network (ATN) grammar, a computationally-oriented grammar which also served as one of the precursors of LFG. When making a transition from Transformational Grammar to LFG, Bresnan (1978) argued that the model she was proposing was psychologically

more realistic. This point was again crucially emphasized in the most important compilation of LFG work, The Mental Representation of Grammatical Relations (Bresnan 1982). LFG thus differs in being a linguistic theory with the goal to also serve as the grammatical basis of a computationally precise and psychologically realistic model of natural languages. Consequently, many of the theoretical decisions have been influenced by this perspective (Bresnan 1982, Sells 1985). The fact that a great number of research projects of natural language processing employ LFG-style formalisms also reveals LFG's commitment as a processing-oriented syntactic theory.

Another striking similarity that LFG shares with the Standard Theory is that as the Standard Theory identifies deep structure as the basis for semantic interpretation and surface structure as the basis for phonological interpretation, with transformational rules as the linkage, LFG also posits two levels of syntactic representation: the c-structure (constituent structure) which is the basis for phonological operation and f-structure (functional structure) from which the semantic representation is derived, and functional descriptions provide the linkage between c- and f-structures. However, it should be quickly pointed out that the similarity between deep/surface structure and c-/f-structure stops here. While deep and surface structures are two separate strata in the derivation process, c-structure and f-structure are associated with each other at any given point of the derivation and thus are co-descriptions of the same word string.

1.2.6 Points of Convergence

Despite the fact that these three theories make different choices of what to emphasize and differ in the assumptions they make for the basis of a syntactic theory, and also employ drastically different formalism, there are two significant points of convergence among them and perhaps other contemporary theories as well: the reduced role of transformations and the increased role of the lexicon. How to limit or eliminate the power of transformations and the shifting of emphasis to the lexicon thus are the two major trends in the study of syntax in the past two decades. GB has reduced the earlier various ad hoc and powerful transformational rules to just one: Move-α (move anything to anywhere). LFG and LXC eliminated entirely the

theoretical validity of transformations and employed different analyses and/or lexical/morphological processes to account for syntactic phenomena that are previously accounted for by transformations.

The lexicon plays a central role in Lexical Functional Grammar and Lexicase, as their names suggest. Lexicase takes the most extreme position in that the lexicon of a language is the entire grammar of that language and all linguistic generalizations are expressed by lexical rules, which must be feature-preserving, i.e., they can only add but cannot delete or change features. In LFG, every lexical entry has a set of functional expressions associated with it, and the f-structure of a phrase or a sentence is the result of unification of lexical functional structures according to the functional specifications associated with phrase structure rules that build the phrase or sentence. What most transformations used to perform now is handled by lexical rules, such as Passivization, Equi, and Raising. The significant similarity among all three theories is therefore that the clause structure of a verb to a large extend can be predicated by its semantic predicate structure. And the argument structure of a predicate is specified in the lexicon. The Projection Principle of GB and the Principle of Function-Argument Biuniqueness of LFG ensure that the predicate argument structure is realized structurally and may not be altered in essential ways. Such conditions are not necessary in LXC where contextual features associated with lexical items dictate entirely the possible phrase structures of a clause, and implied case relations, i.e., thematic relations roughly, are also specified in the lexical entry.

To view LFG through this historical and comparative frame is interesting and revealing; it seems that each of the theories discussed here, in their particular emphasis, presents a different reaction to Chomsky's later directions and represents a developmental stage of Chomsky's theory of syntax. However, we by no means imply any accusation of LXC and LFG of reversing progress, for both theories do address the important issue of explanatory adequacy. Rather, we respect the commitment to formal rigor and processing efficiency in grammar writing on the part of LFG and LXC. We will come back to this point again when we discuss these three theories in CHAPTER 4 regarding their application in computational tasks of natural language processing.

1.3 LFG: A Sketch

In this section, we will provide a brief description of the LFG theory and its formalism, based on Kaplan and Bresnan (1982), Sells (1985), Kaplan (1989), and Huang (1989c). This description thus serves as the basis for the comparison between the conventional LFG and our vLFG discussed in the next section. The best-known feature of LFG theory is probably that its grammar assigns two types of syntactic representations to a sentence or phrase, c-structure to reflect its constituent structure and f-structure for its grammatical/functional structure. The idea of c- and f-structure division is similar to that of the ID/LP format (ID = immediate dominance; LP = linear precedence) of stating phrase structure rules in GPSG. The difference is that while the ID rules are an abstraction of pure constituency away from LP rules, which describe the linear ordering of constituents, f-structure is an abstraction of the grammatical, relational, or functional information away from both phrasal constituency and ordering. Therefore, c-/f-structure division and ID/LP format are totally compatible. Many LFG practitioners do adapt the ID/LP format in stating phrase structure rules (e.g., Huang and Mangione 1985).

The c-structure serves as the basis for phonological interpretation while the f-structure for semantic and discoursal interpretation. Lexical items have their functional information defined in the lexicon. The LFG context-free phrase structure rules are augmented with functional annotations. The c-structure formed by the PSR's is thus also annotated with functional expressions. The f-structure is then formed by instantiating the functional annotations on the c-structure.

1.3.1 The Lexicon

Capitalizing on the linguistic trend of having lexical information account for various grammatical phenomena, in LFG the lexicon plays an essential role. All the functional or grammatical information of a sentence comes from the words in it and the functional specifications on phrase structure rules. In the lexicon, all lexical entries are fully inflected and thus have complete grammatical information. A lexical entry contains its categorial specification and a set of functional expressions known as "schemata," which are to be

instantiated, or to take place, when the c-structure of the sentence is reached. The English word 'forces', as in a construction like "Mary forces John to go" would have the following entry:

1. forces V,
   (↑ PRED) = 'force <(SUBJ) (OBJ) (XCOMP)>'
   (↑ OBJ) = (↑ XCOMP SUBJ)
   (↑ SUBJ PERSON) $=_c$ 3rd
   (↑ SUBJ NUMBER) $=_c$ SG

For practical purposes, we can interpret ↑ as "the level of f-structure associated with my mother node in the c-structure" and ↓ as "the level of f-structure associated with my own node in the c-structure." The syntactic category of this lexical entry is V. The first functional equation defines (indicated by =) this entry's semantic form and the predicate structure. It requires three arguments, or subcategorizes for three grammatical functions, SUBJ, OBJ, and XCOMP. The second equations defines the "control" relation between the matrix object (John) and the missing subject of the embedded non-finite clause XCOMP, with the former being the controller and the later the controlled. The subject of "to go" is therefore to be identified (unified) with "John." The third and fourth equations are functional constraints (indicated by $=_c$). Unlike an equation of definition (=), where unification actually takes place, a constraint ($=_c$), where no unification is performed, only checks whether such attribute-value pair exists or not. Thus the two constraints in this entry make sure that its subject has the attribute PERSON with value 3rd and the attribute NUMBER with value SG. If there is any constraint not satisfied, the functional structure is considered ill-formed. Many agreement features are checked in this manner. As we shall see later, the functional equations, or schemata, on a lexical entry are instantiated only after the c-structure is constructed.

1.3.2 The C-structure and Phrase Structure Rules

Basically an LFG grammar consists of a lexicon, a set of context-free phrase structure rules (PSR's) functionally annotated, and certain well-formedness conditions on f-structures. The PSR's describe the

constituent structure of a sentence, and in fact they may also be viewed as constraints on the well-formedness of c-structures. Only the syntactic categories, not the functional equations, of the words contained in a word string are relevant to its c-structure. Let's see a simple example, "Mary loves John" assuming the following lexicon and PSR's.

2. a. Mary N,
(↑ PRED) = 'Mary'
(↑ PERSON) = 3rd
(↑ NUMBER) = SG

b. John N,
(↑ PRED) = 'John'
(↑ PERSON) = 3rd
(↑ NUMBER) = SG

c. loves V,
(↑ PRED) = 'love <(SUBJ) (OBJ)>'
(↑ SUBJ PERSON) $=_c$ 3rd
(↑ SUBJ NUMBER) $=_c$ SG

3. a. S → NP VP
(↑ SUBJ)=↓ ↑=↓

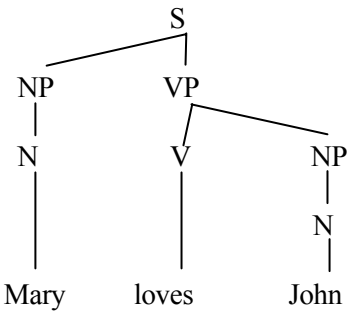
b. VP → V (NP) (NP)
↑=↓ (↑ OBJ)=↓ (↑ OBJ2)=↓

c. NP → (DET) (A) N
↑=↓ ↓ε(↑ ADJ) ↑=↓

If all we are concerned with is the c-structure of the sentence, both the functional equations on the lexical entries and the functional annotations, also known as "schemata," on the PSR's can be ignored. The only relevant feature is the lexical item's categorial specification indicating the preterminal category under which the lexical item may be inserted. Obviously then, the notion of c-structure is entirely compatible with that of the more commonly known tree

structure. These PSR's thus assign the following tree structure correspondence to the word string. In the tree representation of c-structures, we will borrow the Lexicase convention that a straight vertical line indicates that the lower node is the head of its mother node, but this does not mean that we assume the Lexicase position that all c-structure constructions have a lexical head.

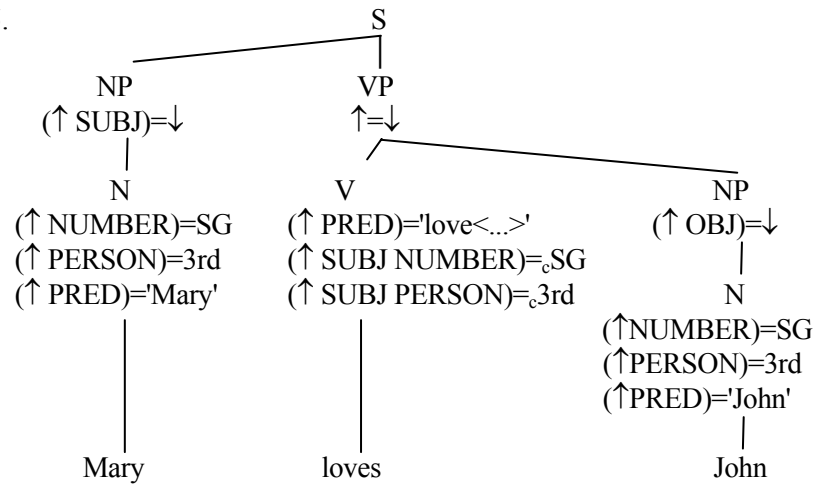
4.



1.3.3 Unification: from C-structure to F-structure

Of course c-structure rules in LFG are different from simple context-free PSR's in that they are augmented with functional annotations; therefore, they assign to a sentence its c-structure with particular specifications to regulate the manner of unification in constructing the f-structure of the sentence. Again, let's look at our previous example. The functionally annotated PSR's would assign the following c-structure with functional schemata to the sentence.

5.



Note that the functional schemata originating in the lexicon are not formally distinct from the ones coming from the augmented PSR's. Through the instantiation of the functional schemata and the operation of unification, the f-structure of the sentences is then composed. Therefore, f-structure is constructed only after the c-structure is constructed. Readers interested in the formal, detailed procedures of deriving f-structure correspondence from c-structure through functional descriptions should refer to Kaplan and Bresnan (1982) or Wescoat (1987). We will now illustrate in the figure below the correspondence between the c- and f-structure, or the co-description of c- and f-structures, of the sentence "Mary loves John."

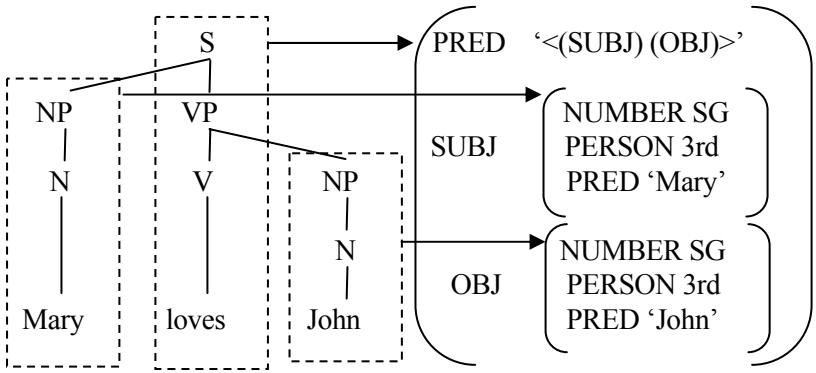


Figure 1.1 Co-description of C- and F-structure

As mentioned before, LFG also posits certain well-formedness conditions on f-structures. A final f-structure associated with a final c-structure still needs to satisfy all the well-formedness conditions. We will discuss these conditions together with grammatical functions in CHAPTER 2, since most of the conditions make reference to the notion of "subcategorizable functions."

Thus, in LFG, both phrase structure rules and information associated with lexical items contribute a local co-description of partial information to the final c- and f-structure of the sentence. The advantage of this division is that it allows separate encoding of external constituent structure, which varies across different languages, and the internal relational structure of grammatical functions which is largely invariable universally. The two structures are two independent but parallel planes associated to each other. Therefore, the mapping from a c-structure to an f-structure is purely procedural and not derivational, which is the crucial difference between c-/f-structure division and the bistratal deep/surface structure distinction.

1.3.4 From Form to Meaning

The notion of correspondence in LFG is an important one: it provides a general way of correlating different kinds of linguistic information in separate domains through modular specifications. We have seen that the word string,

through the mediation of PSR's, first maps to a c-structure correspondence, which is then mapped to an f-structure correspondence by the instantiation of functional schemata and the operation of unification. Kaplan (1989) suggests that the correspondence idea may be extended for the entire mapping between the external form of an utterance and its internal representation of meaning, as shown in the following configuration.

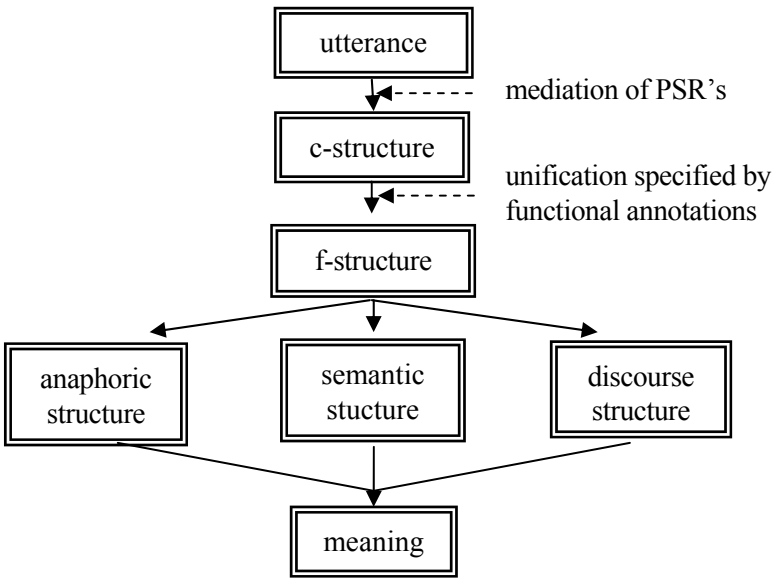


Figure 1.2 Correspondence of Linguistic Representations in LFG

By extending the same conceptual framework of description and correspondence, the same existing mathematical and computational techniques in mapping c- to f-structure can be applied to other systems of linguistic information. For the semantic component in LFG, refer to Halvorsen (1983).

1.3.5 Further Readings

The most important compilation of works in LFG to date is still Bresnan (1982), of which Chapter 4 (i.e., Kaplan and Bresnan 1982) describes

systematically LFG theory and formalism. Two other chapters that are also important in terms of theoretical significance are Chapter 1 (Bresnan 1982a), which justifies the advantages of lexical processes in a lexical theory over transformational analyses, in the case of passivization, and Chapter 5 (Bresnan 1982b) where a theory of control relations and complementation in LFG is presented. The section on LFG of Sells (1985) constitutes a well-balanced and comprehensive introduction to the theory and formalism. Likewise, Chapter 4 of Horrocks (1987) is a clearly illustrated and explicated introduction on LFG. As for detailed, step-by-step instructions on working with LFG's formalism and unification process, Wescoat (1987) is a most practical manual.

Kaplan (1989) provides a description of LFG's formal architecture and a summary of the recent developments of LFG's formalism since Kaplan and Bresnan (1982): functional uncertainty, functional precedence, and the natural-class organization of grammatical functions (also known as the Lexical Mapping Theory). The Lexical Mapping Theory started with the pioneering work of Levin (1986) and was formally introduced in Bresnan and Kanerva (1989). The extended technique of functional uncertainty is described in Kaplan and Zaenen (1989a), and Kaplan and Zaenen (1989) present a discussion on functional precedence in LFG. For the operation of unification, Shieber (1986) is still the best reference.

1.4 vLFG: A Variant LFG Formalism

In this section, we will describe the formalism and specific notations of grammar writing within the vLFG employed in this study. We will illustrate that vLFG maintains most of LFG's theoretical constructs, including its most distinctive feature: the division of constituent structure and functional structure. The major differences of vLFG include the timing of building the functional structure, the notation of lexical entries and phrase structure rules, and the informational organization of the lexicon.

The syntactic component of a vLFG grammar, like LFG, contains a lexicon, a set of phrase structure rules augmented with functional annotations, and a set of well-formedness constraints on f-structures. To reach a c-structure and a corresponding f-structure for a word string, individual lexical items contained in the string are first looked up in the lexicon. When a matching lexical entry

is found, all feature inheritance (FI) entries specified in it will be expanded. When the lexical item is fully specified, it can then be inserted under a category equivalent to its categorial specification. When phrase structure rules apply to assign a partial c-structure of the sentence, unification takes place at the same time to assign a partial f-structure to the c-structure. Thus, whenever a final c-structure is reached, a final f-structure is also reached, at the same time. We will now step through each one of these stages and describe the various notations of lexical entry, feature inheritance entry, and augmented phrase structure rules, and also the operative mechanisms associated with them.

1.4.1 The Informational Organization of the Lexicon

The vLFG utilizes a feature inheritance structure in the lexicon that does not exist in LFG. A vLFG lexicon is therefore composed of lexical entries and feature inheritance entries. Another grammatical theory that fully takes advantage of such a scheme of feature inheritance in the informational structure of its lexicon is the Lexicase framework (Starosta 1988). We will illustrate the different structures of lexical entries in LFG and vLFG.

6. LFG:

a. WORD X, `categorial feature
(↑FEATURE1) = VALUE1 `functional schemata
(↑FEATURE2) = VALUE2
(↑FEATURE3) = VALUE3

b. forces V,
(↑ PRED) = 'force <(SUBJ) (OBJ) (XCOMP)>'
(↑ OBJ) = (↑ XCOMP SUBJ)
(↑ SUBJ PERSON) $=_c$ 3rd
(↑ SUBJ NUMBER) $=_c$ SG

7. vLFG:

a. WORD:

```
(  CAT   X                                  )   `keyword string
(  FS                                       )   `dag
(        (  FEATURE1   VALUE1  )            )   `category (CAT)
(        (  FEATURE2   VALUE2  )            )   `f-structure (FS)
(                                           )
(  FI-1                                     )   `feature
(                                           )   `inheritance
                                                `entry (FI)
```

b. forces V,

```
(  FS                                       )
(       (  FORM   'force'  )                )
(                                           )
(  FI-V-PIVOTAL                             )
(  FI-3RD.SG                                )
```

c. FI-V-PIVOTAL:

```
(  CAT   V                                  )
(  FS                                       )
(        (  PRED <SUBJ , OBJ , XCOMP>  )    )
(                                           )
(  (↑ OBJ) = (↑ XCOMP SUBJ)                 )
```

d. FI-V-3RD.SG:

```
(  FS                                       )
(       (  PERSON   3RD  )                  )
(       (  NUMBER   SG   )                  )
```

A lexical entry in vLFG is composed of two fields: one, the keyword string, and two, the definition of the word, i.e., the grammatical information associated with that word, organized in a bracket format. Since the information specified in this format is completely equivalent to a directed acyclic graph, or "dag" in short, we will refer to them as dags, following the common usage in the literature of unification grammars. Thus, a lexical entry is composed of a keyword string and a dag depicting its grammatical information, which includes the categorial specification of the lexical item and its f-structure. The information represented in the above bracket format is identical with the one below; however, for the sake of typographical ease we will adopt the following format.

8. a. WORD: 'keyword string

```
[ CAT X                                   'dag
  FS   [ FEATURE1   VALUE1
         FEATURE2   VALUE2
       ]
  FI-1
]
```

b. forces:

```
[ FS   [ FORM   'force'   ]
  FI-V-PIVOTAL
  FI-3RD.SG
]
```

Information that is idiosyncratic to the lexical item is specified in the word dag. Information that is shared with a word class, be it functional or otherwise, can be specified through feature inheritance entries. A feature inheritance (FI) entry is very similar in structure to a lexical entry, except that it contains information common to a class of lexical items, as illustrated below. The information contained in the entry FI-1 is shared by a class of words that have this FI-1 specified in their word entries.

9. FI-1: `entry name
[FS [FEATURE3 VALUE3 `dag
FEATURE4 VALUE4
]
]

It is easily noticeable that an entry of feature inheritance is very similar in structure as a lexical entry. When a lexical entry has its FI entries expanded to become fully specified, the expansion is achieved through the unification of the lexical dag and the FI dags. It is crucial to point out that the unification between a lexical dag and a feature inheritance dag is by way of default, or extension as it is sometimes called. It is also equivalent to the LFG mechanism of priority union proposed in Kaplan (1987). If there is a conflict in terms of the value of certain feature, unification by extension will not fail; rather, the value of the lexical dag is preserved and the conflicting value of the feature inheritance entry will be overwritten. Assuming that value1 does not equal value2 below, we will contrast unification, signaled by =, with the operation of extension, signaled by ≈:

10.a. Unification (=)
[feature1 value1] = [feature1 value2]
→
fails, no unification

b. Extension (≈)
[feature1 value1] ≈ [feature1 value2]
→
[feature1 value1]

By allowing unification by extension, idiosyncratic behaviors of a lexical item can be fully accounted for and the generalizations can be stated in a maximally general way. A lexical entry may inherit information from more than one feature inheritance entry, and a feature inheritance entry may in turn call upon other feature inheritance entries and acquire more information from them. A feature inheritance structure thus not only maximizes the economy of the size of the lexicon but also fully captures the generalizations of word

classes while still allows the idiosyncrasies of individual lexical items. A homograph is indicated by a disjunction mark {}:

```
11.   WORD:                                          'keyword string
      {     [ CAT   X                                'homonym 1
              FS   [ FEATURE1   VALUE1
                     FEATURE2   VALUE2
                   ]
              FI-1
            ]
            [ CAT   Y                                'homonym 2
              FS   [ FEATURE1   VALUE1
                     FEATURE3   VALUE3
                   ]
              FI-2
            ]
      }
```

1.4.2 The Phrase Structure Rules

A syntactic rule in the vLFG consists of three parts: rule name, pattern matching, and actions. Like LFG, phrase structure rules are augmented with functional expressions.

```
12. PSR-1:                                           'rule name
        X     <↓ FEATURE1> =c VALUE1                 'pattern
        Y:    GRAMMATICAL FUNCTION
        →                                            'action
        Z
```

The pattern matching section specifies what kinds of syntactic elements are needed in order to build a higher category. A constituent may assign a certain grammatical function, indicated by either: or ε followed by a function name; the former usually assigns a subcategorizable function, while the latter assigns an adjunctive function. We will illustrate these two symbols with their LFG equivalents below. The only required part of the action section is the

specification of the higher category that is to be built. When the higher category is built, in the meantime its corresponding f-structure is also built. In the event that the building of the f-structure fails, the higher category will not be built. Other actions, if any, then perform the specified operations on the f-structure thus far built by the rule. Again, if any of the operation of the actions fails, the higher category still will not be built. We will illustrate the differences between the notation of LFG and vLFG below.

13. LFG:

a. S → NP VP
(↑ SUBJ)=↓ ↑=↓

b. VP → V (NP)
↑=↓ (↑ OBJ)=↓

c. NP → (DET) (A) N
↑=↓ ↓ε(↑ ADJ) ↑=↓

14. vLFG:

a. PSR-1:

NP: SUBJ
VP
→
S

b. PSR-2:

V
(NP): OBJ
→
VP

c. PSR-3:

(DET)
(A) ε ADJ
N
→
NP

Given a lexicalist linguistic theory like LFG, we feel that it is more suitable intuitively to express in a phrase structure rule the constituents first: X Y -> Z, rather than Z -> X Y. The former indicates that when two elements, X and Y, are found, Z is built and that between X and Y the one not assigned with any grammatical function is the head. In terms of both parsing and generation, this scheme comports better with the lexicalist view of grammar where it is the actual lexical items that dictate the structure of a string and activate the parser or generator for processing. The pattern section allows optionality and disjunction, and the specification of partial ordering is also facilitated.

15.	a. [X Y]	`X immediately precedes y
	b. < X Y Z >	`X, Y and Z in any order
	c. { X Y }	`either X or Y, but not both
	d. < 1-2; X Y Z >	`X, Y and Z in any order except that `X must precede y; this specifies `partial ordering
	e. (X)	`X is optional
	f. X+	`one or more X
	g. X*	`zero or more X; equivalent to (X+)

1.4.3 Unification: from C-structure to F-structure

In conventional LFG, phrase structure rules first build the entire c-structure of the sentence. Unification then proceeds according to the specified functional equations on the c-structure to build a corresponding f-structure (Wescoat 1987). In the vLFG however, as we have just seen from the vLFG rule format, when any legal portion of the c-structure of a sentence is built, it must have simultaneously an f-structure built to correspond to it. In other words, when a phrase structure rule applies to build a higher category, the

functional equations specified by that rule are applied at the same time. And only when the f-structure associated with the c-structure is also sound, can the higher category be built. As explicated by Figure V in Wescoat (1987:10), LFG does allow this alternative. We will give a simple example to illustrate some of the points we have covered thus far, assuming the rules in 12 and 13 above, the following lexicon in vLFG, and the lexicon of 2 in LFG, given the sentence: "John loves Mary."

16. a. Mary:

```
[ CAT   N
  FS   [ FORM 'Mary'
         PERSON 3rd
         NUMBER SG
       ]
]
```

b. John:

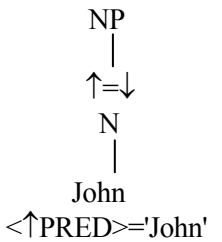
```
[ CAT   N
  FS   [ FORM 'John'
         PERSON 3rd
         NUMBER SG
       ]
]
```

c. loves:

```
[ CAT   V
  FS   [ PRED   <SUBJ , OBJ>
         FORM 'love'
         TENSE PRESENT
         SUBJ   [ PERSON 3rd
                  NUMBER SG
                ]
       ]
]
```

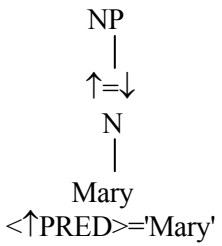
17. LFG:

a.



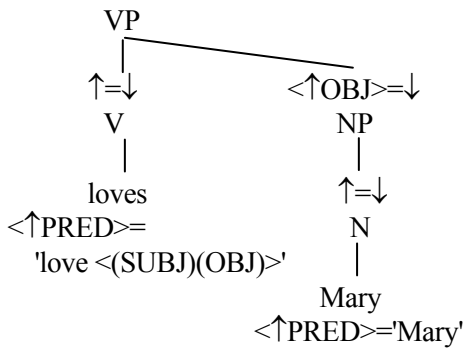
`no unification
`no f-structure

b.

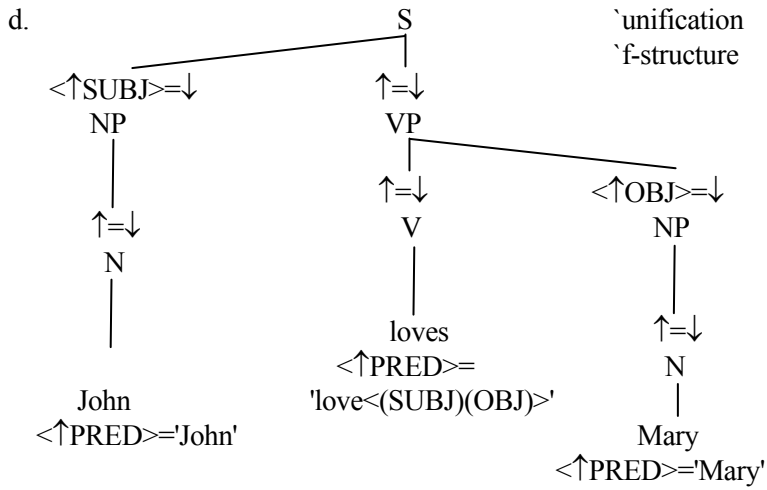


`no unification
`no f-structure

c.



`no unification
`no f-structure



Only when a c-structure that expands the whole word string is reached does unification proceed to build the f-structure according to the functional annotations on the final c-structure:

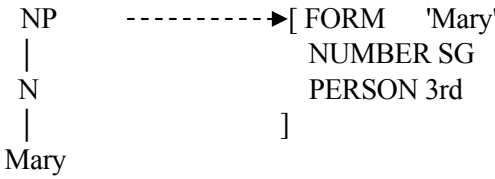
18.

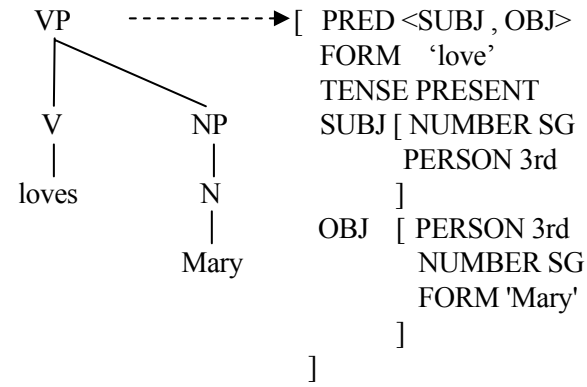
$$\begin{pmatrix} \text{PRED} & \text{'love <(SUBJ)(OBJ)>'} \\ \text{SUBJ} & [\ \text{PRED 'John'}] \\ \text{OBJ} & [\ \text{PRED 'Mary'}] \end{pmatrix}$$

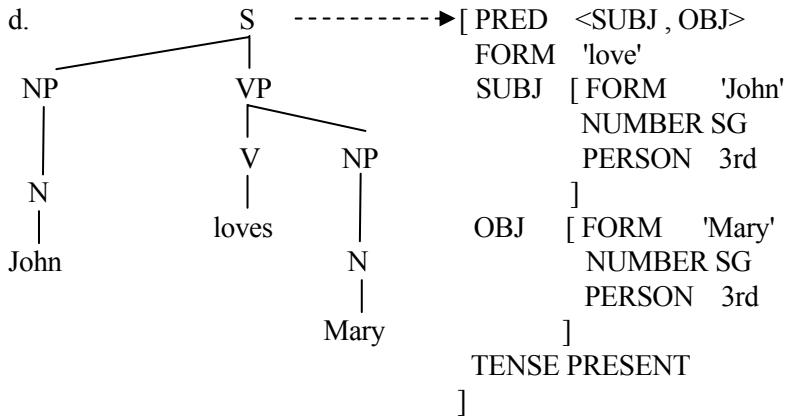
The vLFG on the other hand builds a c-structure and a corresponding f-structure simultaneously in every step:

19. vLFG: c-structure ----------→ f-structure

a.


b.


c.




The advantage of vLFG in this respect is that violations of grammatical constraints in the f-structure can be detected earlier in the unification process and thus paths with f-structure violations will not be pursued any further. Any constituent formed would have to have a well-formed f-structure as well. LFG, on the other hand, would pursue all paths of c-structures that the PSR's allow and violation of f-structures can only be detected when the final c-structure is reached because only then does unification proceed to build the f-structure. This difference in the timing of the building of f-structure thus has rather significant implications on computational efficiency as well as the psycholinguistic processing of natural languages. The vLFG approach in this respect resembles a dependency grammar such as Lexicase in that the constituent structure of a word string cannot exist independently from the participating words in that string. The dependency requirements and restrictions are observed in all steps of constructing the constituent structure of a sentence. The conventional LFG approach, however, in constructing the c-structure is only concerned with the requirements and restrictions, in terms of categorial features, specified by the PSR's.

CHAPTER 2
GRAMMATICAL FUNCTIONS IN AN LFG GRAMMAR OF MANDARIN CHINESE

In this chapter, we will first discuss the place of grammatical functions in LFG and the relation between thematic roles and grammatical functions. We will present a classification of grammatical functions in a grammar of Chinese, in terms of their subcategorizability and semantic restrictions. More specifically, we will discuss the notion of subcategorization and each grammatical function in the context of LFG as well as the grammar of Mandarin Chinese. Due to the especially confusing status of topic and subject in the literature of Chinese linguistics, we will devote a more lengthy discussion to clarify their use. Furthermore, we will argue for our treatment of TOPIC as non-subcategorizable grammatical function in the grammar of Chinese. We will also examine the semantic restrictions on OBJ2 in the context of ditransitive verbs and its syntactic encodings in Chinese.

2.1 Grammatical Functions and F-structure

Grammatical relations, such as subjects, objects, and modifiers, are known as "grammatical functions" in LFG. For a sentence to be well-formed it must have a well-formed c-structure as well as a well-formed functional structure. A functional structure of a sentence contains the functional information of a sentence, and thus is the representation of the hierarchies or internal relationships of grammatical functions among all the components of the sentence. To illustrate, wo3 'I' functions grammatically as the subject of the verb hao3 'fine' in the sentence wo3 hao3 'I am fine', which has the following f-structure in LFG.

1. [PRED 'hao3 <(SUBJ)>'
SUBJ [PRED 'wo3']
]

Functional structures can also be viewed as functions from attributes to values in a mathematical sense. For example, the f-structure of the above example sentence wo3 hao3 can be understood as a function and when we apply this function to the attribute subject we get the value wo3. Thus, the "functional" part of the name of the LFG theory has both grammatical as well as mathematical significance.

2.2 Grammatical Functions and Thematic Roles

In the theory of Lexical Functional Grammar, grammatical relations are lexically encoded by mapping thematic roles in the predicate argument structure (PAS) to grammatical functions. The mapping process between PAS and grammatical functions is one of the foci of recent LFG developments known as the Lexical Mapping Theory, highlighted by Bresnan and Kanerva (1989). Positions in the PAS are identified with individual thematic roles, such as "agent," "theme," "location," etc. The Lexical Mapping Theory associates each thematic role with a subcategorized grammatical function, such as SUBJ, OBJ, SCOMP, etc. The Function-Argument Biuniqueness Principle ensures that a unique function is mapped with each thematic role and a unique thematic role to each function associated with the PAS. However, the relation between thematic roles and grammatical functions may not always be one-to-one because of the possibility of non-thematic grammatical functions, e.g., the raised subject in "He seems to be sick" or the object of the idiom string "kick the bucket." In addition, lexical rules, such as Passivization and Locative Inversion, in Lexical Mapping Theory may change the thematic roles of grammatical functions. Most recently, Lexical Mapping Theory has been preliminarily and rather sketchily applied to Mandarin Chinese (Huang 1989). It is certainly an area worth much further exploration.

Once again, the notion of correspondence is important here: Lexical Mapping Theory assigns the thematic structure to the predicate structure of grammatical functions, or the structure of subcategorization. The theory of LFG, with this recent addition of Lexical Mapping Theory, therefore can be

further extended to model the mapping relations from conceptual structures to linguistic representations (Huang 1989), as shown below.

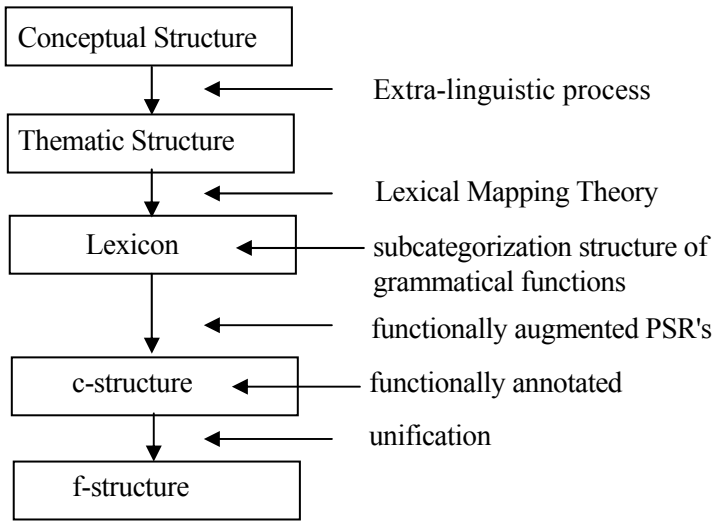


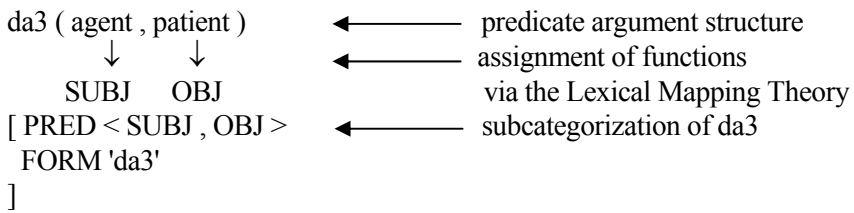
Figure 2.1 Correspondence between Conceptual Structure and Linguistic Structures

In this study, we are primarily interested in the possible patterns of grammatical functions allowable in the predicate structure of Mandarin verbs. We will have little to say about their semantic/thematic counterparts and the mapping process between them. Since grammatical functions have little intrinsic semantic significance, our focus is on the surface syntactic structure required by the verbs. Even so, due to the intimate relation between thematic roles and grammatical functions, often semantic criteria are used to test whether a certain element is a subcategorized argument of the verb, as we shall see in later sections.

2.3 Syntactic Encoding of Grammatical Functions

In LFG, grammatical functions are considered universal and they constitute a small, finite set. However, these universal functions are encoded in the syntactic categories in various manners in different languages. For instance, in

a so-called configurational language like English or Chinese, grammatical functions are assigned to c-structure positions, while in a non-configurational language like Japanese or Malayalam, grammatical functions are assigned to case features. In other words, the correspondence between grammatical functions and syntactic categories is very language-dependent and needs to be specified in the grammar of any particular language. The correspondence between syntactic categories and grammatical functions again need not to be one-to-one. More specifically, the same grammatical function may be assigned by different c-structure positions, and in turn the same c-structure position may assign different grammatical functions. However, the Direct Syntactic Encoding Principle in the universal grammar requires that all non-lexical rules preserve function assignment; that is, syntactic rules can never delete any function or replace one function name with another. This principle ensures that the syntactic encoding of grammatical functions applies directly to surface structures without the mediation of any syntactic or functional derivation. The following figure illustrates the mapping from PAS to the surface syntactic structure, given the sentence Li3si4 da3 Zhang1san1 'Lisi hit Zhangsan'.



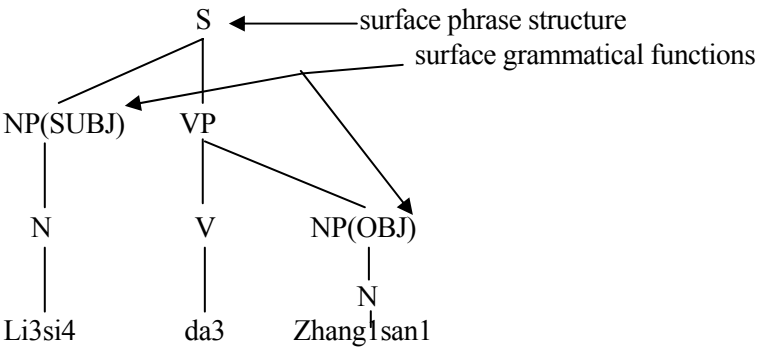


Figure 2.2 Lexical and Syntactic Encoding of Grammatical Functions

2.4 Subcategorization in LFG

In the literature of linguistics, it is well-accepted that the term "subcategorization" refers to the syntactic nature of the elements which a lexical item is required to combine with (Pollard and Sag 1987:113). Thus, subcategorization is taken to be a syntactic notion. Various members of the syntactic category of verb are most commonly classified into different subcategories according to the number and nature of their required dependent, or "subcategorized," elements, although subcategorization restrictions are not limited to verbs. In LFG, the dependency relationships between verbs and their subcategorized elements are characterized in terms of grammatical functions, such as SUBJ, OBJ, XCOMP, etc., not in terms of categories, such as NP, VP, S, etc. The subcategorization of a verb is specified as part of PRED's value in its lexical entry. For instance, in the following entries, da3 'hit', a verb, subcategorizes for SUBJ and OBJ, while wo3 'I', a noun, does not subcategorize for any grammatical function.

2. LFG:

a. da3 V, `hit
(↑ PRED) = 'da3 <(SUBJ) (OBJ)>'

b. wo3 N, `I
(↑ PRED) = 'wo3'

Thus, in LFG, PRED contains a special value which is composed of the semantic form of the lexical item and also contains its subcategorization, if there is any. It is important to point out that the PRED feature in LFG is divided into two features in the vLFG: FORM, for the semantic form of the word, and PRED, for the predicate structure, the subcategorization. PRED therefore takes a list of grammatical functions as its value, since in LFG subcategorization is specified in terms of grammatical functions rather than categories. The PRED feature in our vLFG is therefore rather similar to the SUBCAT of syntax in the Head-driven Phrase Structure Grammar (HPSG) and the vLFG's FORM value is similar to HPSG's RELN (semantic relation) (Pollard and Sag 1987). We illustrate with an example:

3. vLFG:

a. da3: 'hit

```
[ CAT V
  FS   [ PRED < SUBJ , OBJ >
         FORM 'da3'
       ]
]
```

b. wo3: 'I

```
[ CAT N
  FS   [ FORM 'wo3' ]
]
```

Therefore, while in LFG a lexical entry with a PRED value is known as a "semantic form" and a semantic form that subcategorizes for grammatical functions is known as a "lexical form," in vLFG a semantic form is a lexical

entry with the FORM value, and a lexical form is one with a FORM as well as PRED value.

The concept of subcategorization for grammatical functions, rather than syntactic categories, allows certain category-independence in grammatical processes and generalizations. Grimshaw (1982) and Sells (1985:157) provide several arguments for the functional basis of subcategorization. Lexicase takes a similar position, where subcategorization is stipulated in terms of required and implied case roles, not surface syntactic categories such as NP, VP, and S. It should also be pointed out that our use of the terms "subcategorize" and "subcategorization" is actually theory-dependent, in that in LFG their use has a somewhat different meaning from its use in the GB theory. In LFG verb subcategorization is stipulated in terms of grammatical functions and all required functions, including non-thematic ones, are taken to be "subcategorized" (Bresnan 1982:288-292).

In GB theory, "subcategorize" is used in a different sense. GB recognizes only "strict subcategorization," whose features must be "strictly local" in the deep structure, a constituent structure. Thus, subject, which is not a constituent dominated by VP, according to the PSR, S → NP VP, is regarded as "outer argument" and thus non-subcategorizable; only "inner arguments" can be subcategorized. As Starosta (1986) has pointed out in greater detail, this GB division of two types of grammatical arguments, which is a direct result of the recognition of a VP category and the stipulation of an INFL and ARG category, is an unnecessary complication in an otherwise straightforward situation and has missed several otherwise statable generalizations of grammatical arguments.

Similar to the position taken in the theory of HPSG where subjects are treated as subcategorized-for, in LFG, although VP is still a recognized category, the strict locality of subcategorized functions is required in the f-structure, not the c-structure, of the verb that subcategorizes for them. Since the verb is the lexical head of the f-structure of its clause, which, unlike the deep structure VP of GB, does contain the function SUBJ locally, subject is subcategorizable (Ford, Bresnan and Kaplan 1982:773). Actually, in LFG, SUBJ is the function that is required in every verb subcategorization, according to the universal subject constraint in Lexical Mapping Theory (Bresnan and Kanerva 1989). A subcategorization without the SUBJ function is considered ill-formed. In this study of Mandarin verb subcategorization, we

also adopt this position, assuming that the absence of surface subjects in some Chinese sentences is due to discourse factors and thus they are discoursally recoverable.

2.5 Classification of Grammatical Functions

Grammatical functions fall into two different classes: subcategorizable functions and non-subcategorizable functions, and subcategorizable functions can be divided into two types: semantically restricted and semantically unrestricted. Note that in the LFG literature, "subcategorizable" and "non-subcategorizable" are also referred to as "governable" and "non-governable" respectively.

2.5.1 Subcategorizability

A grammatical function is subcategorizable if there is at least one lexical item that subcategorizes for it. A grammatical function is considered subcategorized by a certain lexical item when it appears in the PRED's value of that lexical item. In the following example of the verb "seem" in sentences like "He seems to be sick," two functions are subcategorized by it, SUBJ and XCOMP.

4. seem V, `LFG
(↑ PRED) = 'seem <(↑XCOMP)> (↑SUBJ)'

Note that here SUBJ is also an example of a subcategorized, and yet non-thematic, argument indicated by the notation that it appears outside of the angle-brackets. Non-thematic arguments are subcategorized grammatical functions that are not linked with thematic roles in the PAS. The verb "seem" in this particular construction requires only one thematic role; however, in terms of grammatical functions, it requires two. The following notation in our vLFG formalism illustrates the same fact. A function in the PRED's value followed by an @ sign is a non-thematic argument.

5. seem:: `vLFG

```
[ CAT V
  FS    [ FORM   'seem'
          PRED   < XCOMP , SUBJ @ >
        ]
]
```

Thus, to put it simply, a function is subcategorizable if it appears within the PRED's value of at least one lexical item. In other words, only subcategorizable functions may appear within the PRED's value of a lexical form, a form with a PRED attribute, and may receive assignment form thematic roles in the PAS. A non-subcategorizable function thus may never appear within the PRED's value and may never receive a thematic assignment. For Chinese verbs, we identify SUBJ, OBJ, OBJ2, OBL_{Θ} (oblique function which includes subtypes OBL_{THME} (theme), OBL_{GOAL} (goal), OBL_{BNFC} (beneficiary), OBL_{LOCT} (location)), and COMP (complement function that includes subtypes XCOMP, SCOMP, and NCOMP) as subcategorizable grammatical functions. TOPIC, ADJUNCTS (adjunctive function that has two subtypes ADJ and XADJ), and POSS are non-subcategorizable.

Sells (1985:155-156) suggests that there is some evidence in English for subcategorizable POSS (for instance, "the professor's knowledge" but not "*chemistry's knowledge"); however, such an argument cannot be established in Chinese, where "jiao4shou4 de zhi1shi4" and "hua4xue2 de zhi1shi4" are both acceptable. We do not find any kind of "agency or sentient feeling" consistently associated with the function POSS in Chinese. As we will argue in our discussion of the verb you3, which denotes both possession and existence, in Chinese the concept of possession is subsumed by the more general concept of existence. Further it is possible for POSS to exist in Chinese without a lexical head, such as lao3shi1 de 'the teacher's' which refers to some understood object in the discourse that belongs to the teacher. We therefore treat POSS as a non-subcategorizable function. The following chart illustrates the division of grammatical functions in Mandarin Chinese.

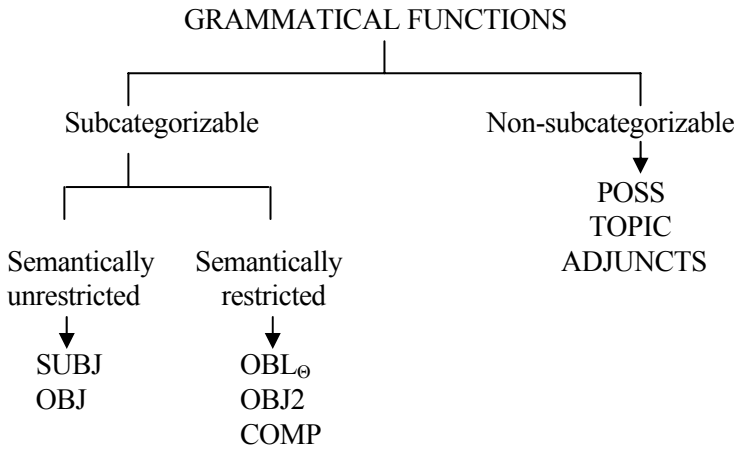


Figure 2.3 Classification of grammatical functions in Mandarin

2.5.2 Semantic Restriction

The distinction between "semantically restricted" and "semantically unrestricted" functions relates to Lexical Mapping Theory that pairs a grammatical function with its thematic role. The unrestricted functions may be linked to any of the thematic roles; for instance, though SUBJ is often linked to agent, it may also be theme, patient, goal, or any other thematic role. Also, only semantically unrestricted functions may be non-thematically assigned, such as the SUBJ of the verb "seem" in our previous example. Semantically restricted functions are more intrinsically related to their semantic content and thus may only be linked to certain thematic roles. For example, OBL_{GOAL} is linked only with the thematic role goal in a PAS. In other words, while semantically unrestricted functions may be semantically variant, semantically restricted ones are usually semantical invariant.

2.6 Subcategorizability of Grammatical Functions and Well-Formedness Conditions on F-structure

For any given f-structure to be well-formed, it must satisfy three conditions: Consistency, Completeness, and Coherence. The concept of subcategorizable, or governable, functions is involved in the last two conditions.

6. a. Consistency (or Functional Uniqueness):
In a well-formed f-structure, any attribute may have at most one value.

b. Completeness
An f-structure is locally complete if and only if it contains all the subcategorizable grammatical functions that its predicate subcategorizes.
An f-structure is complete if and only if all its subsidiary f-structures are locally complete.

c. Coherence
An f-structure is locally coherent if and only if all the subcategorizable grammatical functions that it contains are subcategorized by a local predicate.
An f-structure is coherent if and only if all its subsidiary f-structures are locally coherent.

The one exception to the Consistency condition is the attribute ADJUNCTS, which may have more than one value, in a conglomerated list, which is indicated by curly brackets, {}, in an f-structure. For example, xiao3 hei1 gou3 'little, black dog' would have the following f-structure.

7.
```
[ FORM 'gou3'
  ADJ   { [ FORM 'xiao3' ]
          [ FORM 'hei1' ]
        }
]
```

The conditions of Completeness and Coherence ensure that all subcategorizable functions an f-structure contains are indeed subcategorized by a local predicate, and that all subcategorized functions are all indeed present locally. Sub-functions of ADJUNCTS, i.e., ADJ and XADJ, being non-subcategorizable, may appear (or be absent) freely and thus are exempted from all the above three conditions. In our vLFG formalism we impose one more well-formedness condition: Comprehensibility.

8. Comprehensibility
In a well-formed f-structure, no attribute may have the value ANY.

ANY, along with OPT and NONE, are three special values that need to be explained. Both ANY and OPT are placeholders, meaning that they always succeed in unification (Shieber 1986:43-44). However, an f-structure with OPT value does not constitute any violation. NONE is quite the opposite in that it always fails if unified with any other value. We will give an example of violation of each of the four conditions.

9. a.
```
Inconsistent:                 `unification fails because the
[ NUMBER SG ] =               `result would assign more than one
      [ NUMBER PL ]           `value, SG and PL, to NUMBER
```

b.
```
Incomplete:                   `incomplete because PRED
[ PRED < SUBJ , OBJ >         `subcategorizes two functions but
  FORM 'construct'            `one of them, OBJ, is not contained
  SUBJ [ FORM 'I' ]           `in the f-structure locally
]
```

c.
```
Incoherent:                    `incoherent because PRED only
[ PRED < SUBJ >                `subcategorizes one function and
  FORM 'sleep'                 `yet the f-structure contains
  SUBJ   [ FORM 'I' ]          `another subcategorizable function,
  OBJ    [ FORM 'mary' ]       `OBJ, not subcategorized by the
]                              `local PRED
```

d. Incomprehensible

```
[ PRED < SUBJ >
  FORM 'swam'
  SUBJ [ FORM 'deer'          `incomprehensible because in SUBJ
         NUMBER ANY           `the value of number is ANY
         DEFINITE +
       ]
]
```

We can see that the condition of Functional Uniqueness or Consistency can be viewed as a general constraint on unification: whenever an attribute has conflicting values, unification fails. However, Completeness, Coherence, and Comprehensibility are constraints on the linguistic well-formedness of an f-structure. In conventional LFG, since the c-structure expanding the whole string has to be built first and then the corresponding f-structure is built, Completeness and Coherence are checked only when an f-structure corresponding to a final c-structure is built. In our vLFG formalism, similarly the Functional Uniqueness is always checked whenever unification takes place and the Completeness Condition and Comprehensibility Conditions are checked only after a final f-structure associated with the entire word string is reached. However, the important difference here is that in vLFG Coherence is checked whenever a partial f-structure containing a subcategorizable function is reached, while in LFG the Coherence Condition, like Completeness Condition, is checked only when a final f-structure is reached. This difference has significant psycholinguistic and computational implications. Although we will not discuss, nor justify, these implications in any detail and will simply note that statistical and psycholinguistic studies will need to be done to substantiate our claims, we will show some examples that intuitively indicate some advantages of the vLFG over LFG formalism.

10. a.*John slept Mary the bed.
    b.*John slept Mary the bed, I slept in the chair, and Cindy slept in the sofa.

Given the intransitive "sleep" which subcategorizes <SUBJ> only, the conventional LFG formalism would assign a well-formed c-structure to 10a

which is then ruled out only because when the f-structure is built it is found incoherent due to the existing unsubcategorized OBJ and OBJ2.

```
10a-f-LFG:                                  `[John slept Mary the bed.]
        [ SUBJ  [ PRED 'John' ]
          PRED 'sleep <(SUBJ)>'
          OBJ   [ PRED 'Mary' ]             `incoherent
          OBJ2  [ PRED 'bed'                `incoherent
                  DEFINITE +
                  NUMBER SG
                ]
          TENSE PAST
        ]
```

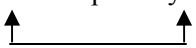
However, in the vLFG formalism, there is no well-formed c-structure assigned to the entire string of 10a, because when "slept" is combined with "Mary" and "the bed" to form a VP, an f-structure is being built simultaneously to correspond to it and this f-structure will then be found incoherent due to the unsubcategorized OBJ and OBJ2 in relation to "sleep" and therefore the VP category expanding "slept Mary the bed" or "slept Mary" will never be built. Thus, in the process of the vLFG analysis, there is no c-structure nor f-structure assigned to the entire string of 10a.

```
10a-f-vLFG:                                 `[John slept][Mary][the bed]
      [ SUBJ [ FORM 'John' ]
        PRED <SUBJ>
        TENSE PAST
        FORM 'sleep'
      ]
      [ FORM 'Mary' ]
      [ FORM 'bed'
        DEFINITE +
        NUMBER SG
      ]
```

Our approach, again, is more similar to a lexicalist word-dependency theory, e.g., Lexicase, in that the c-structure of a word string is intrinsically tied with the idiosyncratic dependency requirements of words in that string. Which approach reflects native psychological processing more faithful is a matter for empirical study, but sentences like 10b might give some indication that our approach is more favorable. Conventional LFG would predict that native speakers detect the ungrammaticality of 10b only when they finish reading or listening to the whole string, which is composed of three clauses and the part that is ill-formed is the very first one. Our approach, along with that of dependency grammars, predictS that the ungrammaticality is detected before the end of the string, which is intuitively more correct.

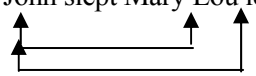
For the following sentences, the conventional LFG would assign c-structures that would not be allowed in vLFG. From a computational point of view, our formalism is thus intuitively more efficient.

11. a. When John slept Mary left.



`c-structure allowed in LFG

b. When John slept Mary Lou left.



`c-structures allowed in LFG

These c-structures with incoherent f-structures would not be allowed in the vLFG formalism and thus such paths would not be pursued further. For 11b, for example, two final c-structures are possible in LFG, one with incoherent f-structure, the other with its f-structure well-formed. In vLFG formalism, again similar to dependency grammars, only one c-structure with a well-formed f-structure is allowed for 11b. We will list below the two f-structures of 11b allowed in LFG.

11b-f-LFG-1: `[[When John slept Mary] Lou left.]

```
[ SUBJ [ PRED 'Lou' ]
  PRED 'leave <(SUBJ)>'
  TENSE PAST
  ADJ { [  CFORM 'when'
           SUBJ [ PRED 'John' ]
           PRED 'sleep <(SUBJ)>'
           OBJ [ PRED 'Mary' ]          `incoherent
           TENSE PAST
        ]
      }
]
```

11b-f-LFG-2: `[[When John slept] Mary Lou left.]

```
[  SUBJ   [ PRED 'Mary Lou' ]
   PRED   'leave <(SUBJ)>'
   TENSE PAST
   ADJ {  [  CFORM 'when'
             SUBJ [ PRED 'John' ]
             PRED 'sleep <(SUBJ)>'
          ]
      }
]
```

Recall that LFG theory imposes the Direct Syntactic Encoding Principle which bans syntactic rules from deleting or replacing any grammatical function. Given this constraint, an incoherent function, once having come to existence in an f-structure, will always be there, and therefore the f-structure is bound to be incoherent. To allow the analysis process to pursue paths that are doomed to failure is counter-intuitive and reduces efficiency. Based on this observation, vLFG formalism checks coherence whenever a subcategorizable function enters an f-structure.

2.7 Grammatical Functions as Primitives

Grammatical functions are taken to be primitives in LFG. This is an important commonality that LFG shares with Relational Grammar (RG), where grammatical functions are called grammatical relations. This is of course one of the major features that separate LFG and RG from the mainstream Chomskyan generative grammar. If grammatical functions are taken to be primitives, that means they cannot be further defined in terms of other syntactic primitives. Thus, to talk about the "definitions" of grammatical functions such as SUBJ and OBJ in the framework of LFG is rather illogical. Actually, an important rationale to take grammatical functions as primitives in LFG is that, although grammatical functions have certain properties and characteristics, these properties are not universal nor sufficient for a set of universal definitions of grammatical functions. Thus, in giving a set of definitions of SUBJ, OBJ, etc., based on surface syntactic configuration in Mandarin Chinese within the LFG framework, Huang (1989) has misconstrued the nature of grammatical functions in LFG. Rather, he should have discussed the correspondence between syntactic categories and grammatical functions, or more specifically what c-structure configurations assign what grammatical functions in Mandarin Chinese, since in this language there is little morphological encoding.

2.8 The Status of Topic and Subject

There are two grammatical functions, SUBJ and TOPIC, that we will discuss in this section in detail due to their rather confusing status and various uses in discussions of Chinese grammar. This confusing state of the use of "subject" and "topic" is no doubt in part attributed to the fact that in Mandarin Chinese there is no morphological markers for either of them.

The confusion also stems from the use of "topic" both as a syntactic notion as well as a semantic/discoursal notion in different discussions. Again, in a language like Japanese, topic, being a parallel notion to subject, will have to be taken as a syntactic notion. In numerous discussions of Chinese linguistics, authors use "topic" without specifying whether they are referring to it as a syntactic notion or a semantic one and from their discussions one finds that it is often intended as both.

Cheng (1983), for instance, which otherwise constitutes an insightful discussion on syntactic devices encoding focus in Chinese, makes no explicit statement regarding whether 'topic' is a syntactic notion or a semantic/discoursal one and it seems that the term is used for both. His statement that shi-predication serves to distinguish a topic from a subject suggests that topic is a notion parallel to that of subject, a syntactic notion. Assertions like "the topic is the slot for unfocused elements" (ibid:97) implies the same. However, from the many sentences that he gives as examples of topicalization one has to conclude that topic is used as a semantic notion as well. For example,

1. a. Wo3 du4lejia4 de di4fang1 shi4 Mao3yi1.
I vacationed DE place be Maui
'The place where I vacationed was Maui.'

b. Zuo2tian1 shang1hai4 ta1 de ren2 shi4 Lao3Wang2.
Yesterday harmed he DE person be Laowang
'The person who harmed him yesterday is Laowang.'

First of all, shi4 is recognized as the main verb in both sentences. If topic is a syntactic function here, any preverbal element considered as the topic cannot be the subject at the same time (for the same obvious reason the subject can not be the object at the same time). If I am right in assuming that Cheng also recognizes that shi4 in these two sentences has a subject, then he must be using 'topic' as a semantic notion here for he considers the entire constituents preceding shi4 the topic.

Similar confusion of syntax and semantics is also frequently found in the use of "subject" and, though to a less extend, "object." Within the field of modern generative linguistics, subject and object are notions within the domain of syntax, not semantics. Chomsky's use of 'logical subject' and 'logical object' is thus unfortunate in terms of their imprecision and subjectivity. Although it should be recognized that certain semantic functions are closely related to the notions of subject and object, a syntactic category and its semantic function should not be confused. Chu (1984:137) initially voiced the same concern about this unsound confusion.

> When talking about subject and object, one has to make sure what subject and object are. While it may be easy to define 'semantic subject' and 'semantic object', it is rather difficult to define 'syntactic subject' and 'syntactic object'! (Translation mine)

However, we must also object to Chu's use of the terms "semantic subject" and "semantic object." That an agent-like or actor-like NP is the subject and a theme-like or patient-like NP is the object is a rather common misconception. The obscure status of subject and topic in Chinese linguistics can be indirectly attributed to the exaggerated prominence of topic in Chinese. Chao (1968:69) first stated that the semantic relation of subject and predicate in Mandarin is that of topic and comment. Thus, though not stated explicitly by Chao, it will have to be assumed that to him "topic" is a semantic concept and "subject" a syntactic one, with a correspondence as depicted below.

SYNTACTIC RELATIONS:	SUBJECT	PREDICATE
	↓	↓
SEMANTIC FUNCTIONS:	TOPIC	COMMENT

However, when Li and Thompson (1981:15) declare that Mandarin is a "topic-prominent" language and that it is typologically different from "subject-prominent" languages such as English, it is quite unclear whether they consider "topic" a syntactic or semantic notion. Their statement "in addition to the grammatical relations of 'subject' and 'direct object', the description of Mandarin must also include the element 'topic'" (ibid.) seems to indicate that they treat topic as a syntactic notion parallel to that of subject and object; yet, they go on to characterize topic in semantic terms as if it were a semantic notion. In fact the same kind of imprecision exists also in their use of the term subject. Starosta in the following criticism clearly points out this inadequacy:

> Li and Thompson failed to recognize the importance of the category of subject in a grammar of Chinese because of their informal and subjective 'functional' approach. Thus they define the category of subject notionally in terms of a 'doing'

> or 'being' relationship (Li and Thompson 1981:87), an approach taken by traditional school grammars but discredited within modern linguistics since the early thirties and especially since the advent of case grammar.
>
> The importance of a grammatically defined category of subject within a relatively rigorous linguistic framework has been amply demonstrated by John Hou, who describes a number of Chinese grammatical processes that crucially involve the category of subject (Hou 1979:47-59, 102, 110-131, 165-193, 205-208), and it comes out even in an informal 'functional' treatment such as Li and Thompson's grammar, where the syntactic distribution of certain classes is stated using what is in effect the position of the subject rather than the topic as the point of reference (Li and Thompson 1981:174, 175, 181, 318, 320, 340, 350, 356) (Starosta 1985a:260-261).

The dilemma for Li and Thompson is if they consider topic a semantic notion, then the dichotomy of "topic-prominent" languages versus "subject-prominent" languages is meaningless since topic and subject are notions of two different domains, one of semantics, the other syntax. In order for this dichotomy to be significant, topic and subject would have to be parallel notions within the same division of linguistics. Since it is well established that subject is a syntactic notion, it follows that topic should be too.

This issue has gained more attention recently, after Her (1985-6) initially voiced the concern and stated explicit definitions of topic and subject as syntactic notions in the Lexicase framework. Li (1988) explicitly recognizes topic as a semantic notion and subject as a syntactic one, while Huang (1989a) explicitly states that both topic and subject are to be taken as syntactic notions, following the convention of Lexical Functional Grammar. Tsao (1987) and earlier papers on Mandarin topics clearly treat topic as a discourse notion beyond the scope of sentences. Note that we do not object to any well-defined use of the term "topic"; however, we have to stress the futility and confusion caused by an un-defined use of the term.

What we have to point out is that those who do not treat topic as a syntactic notion, such as Tsao (1987), Li (1988), Li and Thompson (1981) and Chao (1968), have nonetheless neglected the issue as to what topic, as a semantic or discourse notion, corresponds to in terms of its grammatical relation in syntax. Let's look at the following famous example of topic from Li and Thompson (1976). The topic is italicized.

2. *Nei4 chang2 huo3*, xing4kui1 xiao1fang2dui4 lai2 de kuai4.
that CLS fire fortunate fire-brigade come DE quick
'That fire, fortunately the fire-brigade came quickly.'

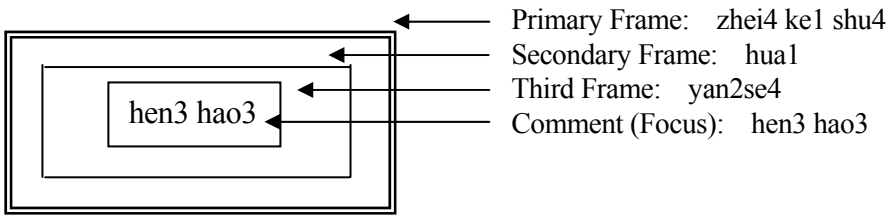
The question unanswered by those who consider topic a semantic or discourse notion is: what is the syntactic, grammatical function of the italicized element, the initial NP? It is certainly not subject nor object. What is it? In this study, we use "topic" strictly as a syntactic notion, following the convention in the LFG theory. Furthermore, we recommend the use of "frame" to denote the semantic/discourse function of topic, following Her (1989c) where the following generalizations on subject, topic, frame, and focus in Chinese are presented.

1) Frame is a semantic/discourse notion which denotes the very center of the old or background information.
2) Focus is a semantic/discoursal notion which denotes the very center of the new or foreground information.
3) The unmarked order in Mandarin is old/background information, general (/whole/universe) preceding specific (/part/scope).
4) Topic is a syntactic notion. The topic of a sentence, being always preverbal and before the subject, usually encodes the semantic/discoursal frame; however, when a topic encodes contrast, a semantic notion subsumed by focus, it does not encode frame and thus necessarily encodes the focus.
5) Subject is always preverbal; therefore, according to 3) above, if there is no topic its unmarked function is likely the frame encoding background information unless there are discoursal principles or phonological principles dictating otherwise.

6) The most prominent position in a Chinese sentence is its predicate whose unmarked semantic relation function is thus the focus.

Therefore, Chao's (1968:69) famous statement could now be extended to be the following: the semantic relation between topic/subject and predicate in Chinese is that of frame and comment. Chafe (1976:50) characterizes the semantic function of the topic as that of setting "a spatial, temporal or individual framework within which the main predication holds." What we have done here is to name that semantic function "frame" and to reserve the term "topic" for its grammatical function, a syntactic notion. Within our defined use, a subject cannot be the topic, nor vice versa, for they are two parallel syntactic notions; yet, while a topic functions semantically as the frame, so may the subject. When there is no topic present, the subject may well be interpreted as the primary frame semantically or discoursally, confirming Chao's observation. The term "frame" also accommodates the concept of topic-chain nicely: in a topic-chain construction, the discoursal interpretative frame is being narrowed down by secondary frames, as depicted in the following illustration:

3. Zhei4 ke1 shu4, hua1, yan2se4 hen3 hao3.
this CLS tree flower color very nice
'As for this tree, its flowers have very nice colors.'



Finally, to be fair, we would like to repeat the concern voiced in Her (1985-6): the confusion of grammatical relations and their semantic functions in the use of "topic," "subject," and "object" is certainly not limited to discussions in Chinese linguistics; rather it is unfortunately common in typological studies, such as word order, in the Greenbergian tradition. A more

precise and well-defined use of these terms can no doubt provide more lucid and revealing results.

2.9 Subcategorizability of TOPIC in Chinese

We have argued in the previous discussion that it is best to treat TOPIC as a syntactic notion in Mandarin Chinese. The LFG use of TOPIC is completely compatible with our use in that TOPIC is a grammatical function within the domain of syntax. However, we still need to justify our position in considering TOPIC a non-subcategorizable function in the grammar of Mandarin Chinese. As we have mentioned in the previous section, in LFG, TOPIC could be either subcategorizable or non-subcategorizable, which has to be specified in the grammar of a particular language. It is important to point out that in any given grammar a grammatical function must be either subcategorizable or non-subcategorizable, never both; the theory of LFG does not allow that possibility. While a subcategorizable grammatical function is always governed by the conditions of Coherence and Completeness, a non-subcategorizable function never is.

The only discussion in the literature on the subcategorizability of TOPIC in Mandarin Chinese is that of Huang (1989a), where he presents a small set of twenty or so Mandarin verbs that seem to require a topic for their sentences to be complete. Based upon such data, Huang argues that these verbs thus subcategorize for a TOPIC in Mandarin Chinese. The following are two examples of such verbs discussed by Huang (1989a).

1. a. Zhei4 jian4 shi4, ni3 zuo4zhu3.
this CLS matter you make-master
'You'll take charge of this matter.'

b.*Ni3 zuo4zhu3.
you make-master
'You'll take charge.'

c.*Ni3 zuo4zhu3 zhei4 jian4 shi4.
you make-master this CLS matter
'You'll take charge of this matter.'

2. a. Yu3yan2xue2, ta1 na2shou3.
linguistics he take-hand
'He is good at linguistics.'

b.*Ta1 na2shou3.
he take-hand
'He is good at it.'

c.*Ta1 na2shou3 yu3yan2xue2.
he take-hand linguistics
'He is good at linguistics.'

Huang analyzes these verbs as having a subcategorization pattern of <(TOPIC) (SUBJ)>. Remember that even if there is only one lexical item that subcategorizes for a grammatical function, such function would necessarily be rendered subcategorizable in that language. We have no significant disagreement with Huang's data which largely comport with native speakers' intuition faithfully; yet, we are rather hesitant to accept his conclusion that accordingly verbs like zuo4zhu3 'take charge of' subcategorize for a TOPIC and thus TOPIC should be considered subcategorizable in Mandarin Chinese. Such an analysis has serious implications that may be in direct conflict with other well-established facts in Mandarin Chinese.

First of all, concerning Huang's data, we should point out that 1b and 2b, although in isolation they do seem odd and incomplete, like many Chinese sentences with missing arguments, are acceptable if given an appropriate discourse context. In the following dialogue, due to the discoursally rich context, both 1b and 2b are acceptable responses.

Q: Zhei4 jian4 shi4, ni3 cai1 shei2 zuo4zhu3?
this CLS matter you guess who make-master
'Guess who takes charge of this matter?'

A: Ni3 zuo4zhu3. (1b)
you make-master
'You do.'

Q: Zhei4 shi4 yin1wei4 yu3yan2xue2 shei2 zui4 na2shou3?
this be because linguistics who most take-hand
'And this is because who is best at linguistics?'

A: Ni3 zui4 na2shou3. (2b)
you most take-hand
'You are.'

This is a very important point to make because, as we will discuss in greater length in the first section of the next chapter, the strongest test for a constituent to be considered as subcategorized-for by a head verb is the obligatory co-occurrence of this constituent and the verb. Therefore, if it were a fact that the topic of verbs like zuo4zhu3 'take charge of' is indeed obligatory, then our position that TOPIC should not be classified as a subcategorizable function would have been considerably weakened. However, since in Mandarin arguments are often not overt, this observation by no means indicates that TOPIC is therefore not subcategorized.

Huang posits that both TOPIC and SUBJ are subcategorizable in Chinese within the LFG framework; however, it is not entirely clear whether this is an acceptable claim in LFG or not. Bresnan (1982:287-288) has proposed that the subcategorizability of TOPIC functions as a parameter that distinguishes "topic-oriented" languages from "subj-oriented" languages. If her proposal is to be taken seriously, then, though not stated explicitly in Bresnan (1982), it would imply that subcategorizable TOPIC is allowed only when SUBJ is considered non-subcategorizable. That is, TOPIC and SUBJ can not be both subcategorizable in a language. For the dichotomy of "subject-oriented" versus "topic-oriented" languages to be significant, this position of mutually-exclusive subcategorizability of TOPIC and SUBJ seems rather unavoidable. Thus, Huang's proposal would imply that SUBJ be non-subcategorizable in Mandarin Chinese, although this is not the position he takes. The essential question is: can an LFG grammar of Mandarin Chinese survive with subcategorized TOPICs but without subcategorized SUBJs? The answer is clearly negative. Not to look far, sentences like 1a and 2a provide counter-examples already. While, as Huang has discovered, there are some twenty verbs in Chinese which seem to require a topic, nearly all Chinese

verbs require the subcategorization of a subject. (As a matter of fact, in this study we take the position that all verbs require a SUBJ, which is also the position assumed in the Lexical Mapping Theory.) The evidence for subcategorized SUBJs in Chinese is simply too overwhelming to ignore (e.g., Starosta 1985:260-261).

Furthermore, leaving the question of SUBJ's subcategorizability behind, one still finds other serious problems with the notion of subcategorized TOPICs in Chinese. In LFG the Coherence Principle requires that a subcategorizable function be subcategorized at all times in a well-formed f-structure. For instance, in English whenever there is a SUBJ in an f-structure, that SUBJ has to be subcategorized by some element on the same level of the f-structure for the f-structure to be coherent. The almost non-restrictive use of topics in Chinese sentences cannot possibly fulfill such a rigid but necessary requirement. Observe the occurrences of TOPICs (italicized) in the following examples.

3 a. *Yu2*, wo3 zhi3 xi3huan1 zhun1yu2.
fish I only like trout
'As for fish, I only like trout.'

b. *Zhang1san1*, wo3 hui4 ma4 ta1.
Zhangsan I will scold him.
'As for Zhangsan, I will scold him.'

c. *Mei3guo2*, wo3 you3 qin1qi4.
U.S. I have relatives
'As for the U.S., I have relatives.'

d. *Na4 chang2 che1huo4*, wo3men tai4 xing4yun4 le.
that CLS car-accident we too lucky LE
'As for that car accident, we were too lucky.'

To put it simply, virtually all well-formed Chinese sentences without a topic can have a topic attached at the sentence-initial position and still be well-formed. To make TOPIC subcategorizable would have to mean that not only all verbs in Mandarin Chinese subcategorize a TOPIC but also that,

except the twenty-odd verbs noted by Huang, all the verbs subcategorize for TOPIC, but only optionally. Such an analysis certainly bears no linguistic merit and would make the claim of subcategorized TOPICs vacuous; we certainly do not want to be accused of missing the forest for the trees.

Another potential problem arises from the topic-chain construction in Mandarin Chinese. The boldfaced portion is the first topic, and the italicized, the second topic.

4.a. **Xiao3shi2hou4**, *yu2*, wo3 zhi2 chi1 zun1yu2.
Childhood fish I only eat trout
'In my childhood, as for fish, I would only eat trout.'

b. **Zhei4 jian4 shi4**, *Zhang1san1*, ta1 mei2 you3 cuo4.
this CLS matter Zhangsan he not have fault
'About this matter, as for Zhangsan, he has no fault.'

c. **Mei3guo2**, *nan2jia1zhou1*, kong1qi4 zui4 huai4.
U.S. south-California air most bad
'As for the U.S, in southern California the air is the worst.'

d. **Nei4 ke1 shu4**, *hua1*, yan2se4 hen3duo1.
that CLS tree flower color many
'As for that tree, its flowers have many colors.'

The first problem is, as we have stated earlier, none of the topics here should be considered subcategorized. Second, in a topic-chain construction, all TOPICs would have to be subcategorized, since TOPIC is now considered subcategorizable in Chinese. The relationship between the first and the second topic is certainly not that of coordination. Thus, within this analysis there are two possible f-structures for topic-chain construction, one with a topic containing another, the other with TOPIC being a conglomerated list containing more than one value. Take 4c for example:

```
4c-f1.  [ SUBJ   [ FORM 'kong1qi4' ]
          FORM   'huai4'
          PRED   < SUBJ , TOPIC >
          ADJ    { [ FORM 'zui4' ] }
          TOPIC [ FORM   'nan2jia1zhou1'
                  PRED   < TOPIC >
                  TOPIC [ FORM 'mei3guo2' ]
                ]
        ]

4c-f2.  [ SUBJ   [ FORM 'kong1qi4' ]
          FORM   'huai4'
          PRED   < SUBJ , TOPIC >
          ADJ    { [ FORM 'zui4' ] }
          TOPIC { [ FORM   'nan2jia1zhou1' ]
                  [ FORM 'mei3guo2' ]
                }
        ]
```

If the first one is taken to be the right f-structure, then it implies that all nouns in Chinese must, optionally, subcategorize a TOPIC function; another vacuous proposal indeed. However, if the second f-structure is taken to be the correct one, then TOPIC must be recognized as another exception to the Consistency condition, a characteristic rather unique to non-subcategorizable adjunctive functions. None of the other subcategorizable functions, such as SUBJ, OBJ, OBJ2 , COMP and OBL_{Θ} may violate the Consistency condition. Therefore, neither of the two possible f-structures seems appropriate for the topic chain construction. On the other hand, if TOPIC is to be considered non-subcategorizable, then either of the two f-structures is workable, although given our previous description of frames we would probably prefer the first choice where the primary frame is contained within the secondary frame.

Based upon all the evidence above, we have to reject the notion of subcategorizable TOPICs in Mandarin Chinese. Thus, we maintain that in Mandarin Chinese the function SUBJ is subcategorizable but TOPIC is not. Yet, how do we account for the twenty or so verbs such as zuo4zhu3 'take charge of' and na2shou3 'be good at' that Huang cited, if TOPIC can not be

subcategorized? The solution we propose is to still recognize verbs like zuo4zhu3 'take charge of' as subcategorizing SUBJ and OBJ. Because TOPIC is usually a placeholder of old, or background, information, we designate a attribute-value pair of [BACKGROUND +] in TOPIC. On the other hand, since OBJ is always encoded in a postverbal position, embedded in the predicate of a clause, we designate [BACKGROUND -] in OBJ through the appropriate phrase structure rule. (See section 2.14 of this chapter for phrase structure rules.) In order to make sure that the OBJ is always "missing" and that the missing OBJ is always to be unified anaphorically with the matrix TOPIC and thus satisfy completeness and coherence conditions, we impose the attribute-value pair of [BACKGROUND +] in the OBJ in the lexical entry of zuo4zhu3 'take charge of'.

```
zuo4zhu3 :
        [ CAT V
          FS      [ FORM    'zuo4zhu3'
                    PRED    < SUBJ , OBJ >
                    OBJ     [ BACKGROUND +
                              FORM ANY
                            ]
                  ]
        ]
```

Thus, the verb does not subcategorize a TOPIC; rather, it subcategorizes SUBJ and OBJ with the specification of an attribute-value pair [BACKGROUND +] in its OBJ. 1c and 2c are ill-formed for their OBJs appear postverbally and thus have [BACKGROUND -] which is in conflict with [BACKGROUND +] of OBJ specified in their lexical entries; unification thus fails. To account for the ungrammaticality of 1b and 2b where neither an overt OBJ nor a TOPIC is present, we impose [FORM ANY] in the OBJ of the lexical entry, for in such cases the value of OBJ's FORM remains to be ANY in the final f-structure, and thus such f-structures are ill-formed (incomprehensible).

```
1b-f [ FS   [ FORM   'zuo4zhu3'                    `*Ta1 zuo4zhu3.
              PRED   <SUBJ , OBJ>
              SUBJ   [ FORM 'ta1'
              OBJ    [ BACKGROUND +
                       FORM ANY                    `incomprehensible
                     ]
            ]
     ]
```

Note our solution still recognizes that zuo4zhu3 'take charge of' requires two thematic roles which map into two grammatical functions, SUBJ and OBJ. Thus, compared to other verbs of similar meanings such as fu4ze2 'be responsible for', semantic generalizations remain. Furthermore, our solution entirely avoids the problems that Huang (1989) has recognized in linking thematic roles to the TOPIC function in the Lexical Mapping Theory of LFG. The difference between zuo4zhu3 and fu4ze2 is thus purely syntactic, in that zuo4zhu3 may never take an overt OBJ and its required OBJ has to be satisfied by linking the TOPIC anaphorically with the missing OBJ, and this is precisely how the missing OBJ of fu4ze2 of 5a below is satisfied. Thus, in our analysis, 1a is entirely equivalent to 5a in terms of both c- and f-structures.

1. a. Zhei4 jian4 shi4, ni3 zuo4zhu3.
this CLS matter you make-master
'You'll take charge of this matter.'

b.*Ni3 zuo4zhu3.
you make-master
'You'll take charge.'

c.*Ni3 zuo4zhu3 zhei4 jian4 shi4.
you make-master this CLS matter
'You'll take charge of this matter.'

5. a. Zhei4 jian4 shi4, ni3 fu4ze2.
this CLS matter you be-responsible
'You'll be responsible for this matter.'

b.*Ni3 fu4ze2.
you be-responsible
'You'll be responsible.'

c. Ni3 fu4ze2 zhei4 jian4 shi4.
you be-responsible this CLS matter
'You'll be responsible for this matter.'

Furthermore, it may be generalized that for 1b and 5b alike their incompleteness arises from the unfulfilled OBJ, not the unfulfilled TOPIC. The identification of a missing OBJ with the matrix TOPIC is a rather general rule in Chinese, as exemplified in the following sentences, and thus no special specification is needed to identify the missing OBJ of verbs like zuo4zhu3 'take charge of' with its matrix TOPIC.

6. a. Xiao3hai2zi, ta1 hui4 xi3huan1.
kids he will like
'Kids, he will like.'

b. Huai4 dian4ying3, wo3 bu4 yao4 ta1 kan4.
bad movie I not want he see
'Bad movies, I don't want him to see.'

c. Zhei4 jian4 shi4, wo3 qiang2po4 ni3 fu4ze2.
this CLS matter I force you be-responsible
'I force you to be responsible for this matter.'

d. Zhei4 jian4 shi4, wo3 qiang2po4 ni3 zuo4zhu3.
this CLS matter I force you make-master
'I force you to take charge of this matter.'

Complex sentences like 6d that involve verbs like zuo4zhu3 'take charge of' in the subordinate clause still provide another argument for our solution. In our analysis, 6a-d are all treated the same, with a missing OBJ in the XCOMP that has to be linked with the matrix TOPIC. However, in Huang's analysis

their f-structures will be very different, with 6a-c's XCOMP missing OBJ that has to be linked with the matrix TOPIC and yet with 6d missing XCOMP's TOPIC that also has to be linked with the matrix TOPIC. Aside from missing this generalization, another dilemma arises: in the XCOMP it is impossible for zuo4zhu3 'take charge of' to have an overt TOPIC that it requires.

7. a.*Ta1 hui4 zhei4 jian4 shi4 zuo4zhu3.
he will this CLS matter make-master
'He will take charge of this matter.'

b.*Wo3 qiang2po4 ni3 zhei4 jian4 shi4 zuo4zhu3.
I force you this CLS matter make-master
'I force you to take charge of this matter.'

Similar facts from the observation of relative clauses and pseudo-cleft sentences also indicate that our solution captures the generalizations while Huang's account misses them.

8. a. Ta1 fu4ze2 de shi4 dou1 shi1bai4 le.
he be-responsible DE matter all fail LE
'The matters that he was responsible for all failed.'

b. Ta1 zuo4zhu3 de shi4 dou1 shi1bai4 le.
he make-master DE matter all fail LE
'The matters that he took charge of all failed.'

c. Ta1 fu4ze2 de bu2 shi4 na4 jian4 shi4.
he be-responsible DE not be that CLS matter
'What he is responsible for is not that matter.'

d. Ta1 zuo4zhu3 de bu2 shi4 na4 jian4 shi4.
he make-master DE not be that CLS matter
'What he takes charge of is not that matter.'

Again, in our analysis, both 8a and 8b are equivalent in c- and f-structures; the missing OBJ of the relative clause is to be identified with the relativized

noun by long distance dependency rules. Yet, in Huang's account, 8a and 8b share the same c-structure but have different f-structures; while fu4ze2's missing OBJ is considered relativized, it is the TOPIC of zuo4zhu3 'take charge of' that is relativized. Therefore, for 8a, long distance dependency rules have to identify two different relativized elements. Within our analysis, 8c and 8d are also equivalent in both c- and f-structures, with the missing OBJ of the headless relative clause identifiable with the NP within shi4's predication. In Huang's account, still, 8c and 8d would have the same c-structure but not the same f-structure; thus, in 8c it is the missing OBJ that is identifiable with the shi's predication, while in 8d it would have to be the missing TOPIC.

To summarize, although Huang's data is correct, his conclusion of subcategorized TOPICs in Mandarin Chinese cannot be accepted without serious compromises. Given the conditions of completeness and coherence in LFG and the fact that almost all Chinese verbs may co-occur with one topic or more, we conclude that TOPIC should be treated as a non-subcategorizable function in Chinese. To account for Huang's data, we propose to treat verbs like zuo4zhu3 'take charge of' as subcategorizing SUBJ and OBJ with a special condition on its OBJ to account for the fact that they may never take an overt OBJ and that their missing OBJ has to be linked with the matrix TOPIC. Furthermore, we list several sentence types to show that our analysis captures the generalizations among similar syntactic constructions while in Huang's account these generalizations would be missed.

2.10 Syntactic Encoding of SUBJ and OBJ

We follow the very straightforward definitions of subject and object offered by Her (1985-6), Li (1988), and Huang (1989). In LFG terms, we thus recognize that in a syntactic configuration the NP position immediately preceding the predicate, usually a verb phrase, encodes the SUBJ function, and the postverbal NP position adjacent to a transitive verb assigns the OBJ function. Our analysis here therefore necessarily entails the implication that in a ditransitive construction the first NP following the verb is to be identified as the OBJ and the second NP OBJ2. This position of course is contradictory to the conventional wisdom. We will discuss and clarify the use and status of OBJ2 in LFG and in a Chinese grammar below.

2.11 OBJ2 as a Semantically Restricted Function

While OBJ refers to the direct object, OBJ2 is what is traditionally and conventionally known as the "indirect object" of a ditransitive verb. According to Bresnan (1982b:287,294) as well as Sells (1985:155), OBJ2, along with SUBJ and OBJ, is classified as a semantically unrestricted grammatical function. However, this position has been corrected in Bresnan and Kanerva (1989) where only SUBJ and OBJ are classified as [-r] (semantically unrestricted) and OBJ2 and OBL_{Θ} [+r] (semantically restricted). It is now recognized that OBJ2 always relates to the Theme role and thus is referred to as OBJ_{th}. (We will however maintain the use of OBJ2 for its simplicity and more widely known use.)

In this study we adopt the later position of Bresnan and Kanerva (1989) that OBJ2 is a semantically restricted function. Under this premise, it seems more appropriate to have the second NP in the predicate of a ditransitive construction assigned to an OBJ2 function which is invariably linked to a theme role, while the first NP after the verb, though it may not realize the whole range of thematic roles, does show more freedom in terms of the thematic roles it may be associated with, as shown in the following examples, where the first postverbal NP and its corresponding thematic role is italicized.

1. a. Wo3 gei3 *ta1* qian2. `V + *Beneficiary* + Theme
   I give he money
   'I give him money.'

   b. Wo3 qiang3 *ta1* qian2. `V + *Maleficiary* + Theme
   I rob he money
   'I rob him of money.'

   c. Wo3 qian4 *ta1* qian2. `V + *Source* + Theme
   I owe he money
   'I owe him money.'

   d. Wo3 jing4 *ta1* yi1 bei1 jiu3. `V + *Goal* + Theme
   I toast he a glass wine
   'I raised a glass of wine to toast him.'

Another piece of evidence comes from the bei4 constructions that are semantically the affective voice counterparts of active voice constructions of transitive and ditransitive verbs. We will have a detailed account of the bei4 sentences in section 3.3.21.5.

2. a. Jing3cha2 da3-le *ni3*.
policeman hit LE you
'The policeman hit you.'

a' *Ni3* bei4 jing3cha2 da3-le.
you BEI policeman hit LE
'You were hit by the policeman.'

b. Jing3cha2 gei3-le *ta1* fa2dan1.
policeman give LE he ticket
'The policeman gave him a ticket.'

b' *Ta1* bei4 jing3cha2 gei3-le fa2dan1.
he BEI policeman give LE ticket
'He was given a ticket by the policeman.'

b"*Fa2dan1 bei4 jing3cha2 gei3-le ta1.
ticket BEI policeman give LE he
'A ticket was given to him by the policeman.'

c. Na4 ge huai4 lao3shi1 jiao1 wo3 fa4wen2.
that CLS bad teacher teach I French
'That bad teacher taught me French.'

c' Wo3 bei4 na4 ge huai4 lao3shi1 jiao1 fa4wen2.
I BEI that CLS bad teacher teach French
'I was taught French by that bad teacher.'

c"*Fa4wen2 bei4 na4 ge huai4 lao3shi1 jiao1 wo3.
French BEI that CLS bad teacher teach I
'French was taught to me by that bad teacher.'

d. Ta1 qiang3 le wo3 qian2.
he rob LE I money
'He robbed me of money.'

d' Wo3 bei4 ta1 qiang3 le qian2.
I BEI he rob LE money
'I was robbed of my money by him.'

d"*Qian2 bei4 ta1 qiang3 le wo3.
money BEI he rob LE I
'I was robbed of my money by him.'

There are two kinds of evidence we can derive from the above sentences. First, note that, as exemplified with 2a-2a', it is always the OBJ (italicized), a semantically unrestricted function, of the active voice construction that is semantically equivalent to the SUBJ, also semantically unrestricted, of the affective voice counterpart. Thus, in order to make this generalization, it is the first NP following the ditransitive verb that should be considered OBJ, as shown by 2b-b' and 2c-c'. Furthermore, the second NP of a ditransitive verb does not usually have the SUBJ counterpart in a bei4 construction, as shown by 2b" and 2c"; this again indicates that the second NP is semantically restricted and thus should not be considered OBJ.

The assumption that there is a hierarchy for the assignment of SUBJ, OBJ and OBJ2 (Bresnan 1982b:294) and our treatment of [ba3 NP] as an oblique function (discussed in the next chapter) also force the interpretation that the first NP following a verb is OBJ and the second NP OBJ2. The hierarchy predicts that OBJ2 may be assigned only if OBJ has been assigned already. In other words, OBJ2 may not exist in the absence of OBJ.

3. a. Wo3 gei3 *ta1* qian2. `V + *OBJ* + OBJ2
I give he money
'I give him money.'

b. Wo3 ba3 qian2 gei3 *ta1*. `OBL_{THME} + V + *OBJ*
I BA money give he
'I give him money.'

If we were to take the second NP to be the OBJ, then 3b would be without an OBJ but with an OBJ2 and thus violate the assumed hierarchy. Yet, under our analysis where the first postverbal NP (italicized) is taken to be the OBJ, this hierarchy will still be maintained in sentence 3b. Based upon these arguments we thus conclude that in Mandarin Chinese the OBJ2 function is encoded by the second NP position in the predicate of a ditransitive verb.

2.12 COMP Function and Subtypes

According to Bresnan (1982:9), the grammatical function of complement, or COMP, has multiple subtypes, which are characteristically distinguished in a grammar by a limited set of categorial features. At this point we should clarify our usage of several terms. "Argument" refers to a subcategorized element, while "complement" is specifically reserved for a predicative argument. Non-subcategorized constituents are often referred to as "adjuncts." In our grammar of Chinese, we distinguish three complement types: NCOMP, SCOMP, and XCOMP. In terms of syntactic encoding, they all occupy postverbal positions. As their names suggest, NCOMP is marked by the constituent category of NP and SCOMP by an S. XCOMP is an open complement typically assigned to an infinitival VP that has at least its SUBJ missing, which needs an antecedent from a higher clause as its "controller." (Due to its close relation with VP, in earlier LFG literatures, XCOMP is also often referred to as VCOMP.) Because of the postulation of c- and f-structures, binding and control relationships are accounted for in LFG in terms of f-structure, unlike GB where such relationships have to refer to external constituent structure directly. LXC also lacks an independent structure designating the grammatical functions of syntactic constituents and thus also has to refer to the external constituent structure and postulate devices such as the argument identification table and argument identification rules.

The newly developed Lexical Mapping Theory (Bresnan and Kanerva 1989), although it provides an elaborate system of mapping thematic roles to

grammatical functions, does not explain how or what thematic roles are mapped to the function COMP and its subtypes. Actually, the lack of discussion of the thematic roles of complements is not unique in LFG. Since the early Case Grammar, discussions of case relations, or thematic relations in more recent terminology, have been primarily concerned with noun phrases and prepositional phrases; little attention has been paid to verb or sentence complements.

2.13 Oblique Function and Subtypes

The oblique function (OBL_{Θ}) also belongs to the inventory of universal functions, and again according to Bresnan (1982:9), it may have multiple subtypes, which are distinguishable within the grammar of a particular language by a closed set of categorial or morphological features. Similar to English, the multiple subtypes of the oblique function in Mandarin Chinese are marked by the finite class of prepositions, including ba3, gei3, dui4, zai4, dao4, etc. (For a criticism on the inappropriate use of the term "coverb" and the justification to recognize the word class of prepositions in a generative grammar of Chinese, refer to Starosta (1985).) Thus, categorially all subtypes of the oblique function are encoded by a PP in Chinese. As stated before, we recognize four subtypes of the oblique function in our grammar of Chinese: OBL_{THME}, OBL_{GOAL}, OBL_{BNFC}, and OBL_{LOCT}.

Each individual subtype of OBL_{Θ} is marked by a number of different prepositions. Since the oblique function is classified as semantically restricted, it is rather logical that each subtype is marked only by prepositions that are semantically compatible with the thematic role linked with the particular subtype. Thus, OBL_{BNFC} is marked by preposition gei3; OBL_{LOCT} by zai3 and dao4; OBL_{GOAL} by gen1, dui4, xiang4, he2, and tong2; and OBL_{THME} by ba3. In terms of the c-structure position they correspond to, OBL_{THME} and OBL_{GOAL} appear preverbally and OBL_{BNFC} and OBL_{LOCT} postverbally. Since each subtype of the oblique function is characteristically linked to individuating thematic roles, the recent development of the Lexical Mapping Theory in its mechanisms of linking certain thematic roles to the oblique function provides means to link thematic roles to specific subtypes of the oblique function, for instance the linking of Theme to OBL_{THME}.

2.14 Syntactic Encoding of Grammatical Functions in Chinese

Based on our discussion above, there are three syntactic devices in Chinese which encode grammatical functions: word order, syntactic categories, and prepositions. We thus assume the following PSR's for our analysis of verb subcategorization. We also take the position that TOPIC is a place holder for old, background information and thus offers definite reference. SUBJ, which always precedes the verb in our analysis, also tends to be definite unless it is syntactically marked to be indefinite. Thus, SUBJ has a default (≈) value of [DEFINITE +]. Also, [ba3 NP], which assigns the function OBL_{THME} in our analysis, being always preverbal, has a strong tendency for being definite; we thus give it a default value [DEFINITE +] as well. Postverbal nouns tend to be indefinite, unless they are syntactically or inherently definite. We will not discuss this in further detail but will note that our position in general agrees with that taken by Li and Thompson (1975), Chu (1979), Light (1979), and Sun and Givon (1985).

1. PSR-S':

NP ε TOPIC	(↓ DEFINITE) = +
	(↓ BACKGROUND) = +
S	
→ S'	

2. PSR-S:

(NP) : SUBJ	(↓ DEFINITE) ≈ +
VP	
→ S	

3. PSR-PP:

P
NP
→ PP

4. PSR-VP:

<(PP) : OBL_{GOAL}	(↓ PCASE) =$_c$ GOAL
(PP) : OBL_{THME}	(↓ PCASE) =$_c$ THME
	(↓ DEFINITE) ≈ +
>	
V.	
(PTCL)	`aspect particle
(NP) : OBJ	(↓ DEFINITE) ≈ -
	(↓ BACKGROUND) = -
(NP) : OBJ2	(↓ DEFINITE) ≈ -
(NP) : NCOMP	(↓ DEFINITE) ≈ -
(PP) : OBL_{BNFC}	(↓ PCASE) =$_c$ BNFC
(PP) : OBL_{LOCT}	(↓ PCASE) =$_c$ LOCT
(S) : SCOMP	
(VP) : XCOMP	
→ VP	

CHAPTER 3
VERB SUBCATEGORIZATION IN MANDARIN CHINESE

This chapter constitutes the core of this study of Mandarin verb subcategorization of grammatical functions within a variant formalism of Lexical Functional Grammar. As we have mentioned in the previous chapter, subcategorization requirements or restrictions within the LFG framework are stated in terms of grammatical functions, such as SUBJ, OBJ, OBJ2, XCOMP, etc., in the functional structure, a level of syntactic representation corresponding to and yet distinct from the constituent structure. First, however, we will present the generally accepted principles for determining whether a certain element should be considered as subcategorized-for by a verb or not. Then, after a discussion of the semantic subcategorization of Chinese verbs according to the distinction of stativity and activity, we will list all the subcategories of Mandarin verbs in terms of their subcategorized grammatical functions. We will discuss each of them in detail and illustrate with examples. In cases where our analysis differs from previous accounts or justification is needed, we shall provide evidence and argue for our analysis.

3.1 Evidence for Subcategorized Elements

The notion of subcategorization is a crucial one in most modern linguistic theories. The term standardly refers to differences among verbs as to the syntactic nature of the various constituents they co-occur with to form a complete sentence. However, in Chinese it is not always straightforward to determine whether a constituent is subcategorized for by a verb. In other words, not all arguments are clearly distinguishable from adjuncts. Although this may be a relatively easy matter in many European languages like English and French where such subcategorized constituents often appear obligatorily, in languages like Chinese and Japanese where such items may appear optionally, decisions about the subcategorization status are consequently harder (Ernst 1989). While in the linguistic literature there is no general agreement as to how the difference between subcategorized constituents and

non-subcategorized ones should be characterized in theoretical terms, there are a number of well-accepted, rule-of-thumb syntactic and semantic tests available which serve to make the distinction (e.g. Pollard and Sag 1987, Ernst 1989, Bresnan 1987, Starosta 1988, and Huang 1989a). Many of the following criteria serving to make this distinction are based on the notion that a verb's lexical entry is the proper repository for idiosyncratic information about the verb and elements it co-occurs with. It should also be noted that most of these criteria should not be taken to be sufficient conditions or definitive tests individually; rather, a stronger case can be made when they are used in conjunction with one another. We will now go through these conditions or tests and illustrate with examples in Chinese.

3.1.1 Obligatory Co-occurrence

A constituent is generally considered as subcategorized for by a verb if its co-occurrence with the verb is obligatory. Therefore, if such a constituent is missing, the sentence is ungrammatical or its acceptability is clearly and heavily dependent upon the existence of an appropriate discourse context. However, this test, which probably provides the strongest, if not sufficient, evidence for subcategorization, is not significantly relevant for our purpose in studying verb subcategorization in Chinese, where, as mentioned before, verbs allow great freedom for their subcategorized elements to be optional. For instance, in the following simple dialogue, the entirely acceptable response can be a "naked" verb stripped of all arguments discoursally recoverable.

1. Q: Ni3 ba3 qian2 cun2 zai4 yin2hang2-li3 le ma1?
you BA money deposit at bank inside LE MA
'Did you deposit the money in the bank?'

A: (Wo3) (ba3 qian2) cun2 (zai4 yin2hang2-li3) le.
I BA money deposit at bank inside LE
'I deposited the money in the bank.'

However, contrary to the common belief that in Mandarin Chinese there are no such cases of obligatory co-occurrence (e.g., Ernst 1989), certain Chinese

verbs, not necessarily small in number, do require their arguments, at least the ones in the predicate, to be overtly present, regardless of what discourse context they actually appear in. We observe that they include verbs like bei4 or zao1shou4 'suffer or incur (adversely)', ting1qi3lai3 'sound', and several other verbs that require a pivotal construction, [V + NP + VP], or [V + OBJ + XCOMP] in LFG terms, such as shi3de2 'make, force'. All the following sentences 1a-d below are thus ungrammatical without the arguments in parentheses, given any discourse context.

2. a. Wo3 bei4 *(ni3 da3).
I BEI you hit
'I was hit by you.'

b. Ta1 kan4qi3lai3 *(hen3 jian4kang1).
he look very healthy
'He looks very healthy.'

c. Wo3 po4shi3 *(ta1 *(qu4)).
I force he go
'I force him to go.'

d. *(Ni3 de gu4shi4) shi3de2 *(ta1 *(hen3 gan3dong4)).
you DE story make he very moved
'Your story makes him very moved.'

Note that the notation ...*(α)... means that the string is unacceptable when α is optional; in other words, α is required for the string to be well-formed. Evidence of obligatory co-occurrence in a grammar of Chinese therefore seems to be a sufficient condition for an element to be considered subcategorized-for, but it is definitely not a necessary one, for most of verbs do not require obligatory co-occurrence of their arguments.

3.1.2 Ontological Necessity

Elements that are subcategorized for by a verb are usually associated with semantic roles that are ontologically necessary for a complete interpretation of

the sentence. Thus, if a verb presupposes a certain semantic role normally expressed by a subcategorized constituent, even when the constituent is missing, such a constituent is still considered subcategorized for by the verb. Such sentences when in isolation thus leave one feeling odd due to their semantic incompleteness. This test is also relatively difficult to enforce in Chinese, again due to the freedom of optional, discoursally recoverable arguments. Furthermore, the ontological nature of this criterion also makes it notoriously difficult to apply. For example, since all events must occur at certain place and time, could we thus assume that all verbs presuppose a location and therefore subcategorize for a locative constituent syntactically? To allow assumptions like this would make the notion of subcategorization useless, so this criterion has to be applied in conjunction with other syntactic principles, which will be described below. Thus, while ontological necessity is a necessary condition for a constituent to be considered as subcategorized-for, like most other criteria below, it is not a sufficient one.

3.1.3 Constancy of Semantic Content

Non-subcategorized, adjunctive constituents can generally occur with a wide, relatively unrestricted range of head verbs while making a rather uniform contribution of semantic content to that of the sentence; subcategorized constituents, however, often may only co-occur with a semantically and/or syntactically or even arbitrarily restricted class of head verbs while their contribution to the semantic content is often idiosyncratically dependent upon the particular verb. Let's examine the following examples.

3. a. Wo3 *gen1 ta1* pao3bu4/wan2shua3/you2yong3/he1jiu3.
I with he jog play swim drink
'I jog/play/swim/drink with him.'

b. Wo3 *gen1 ta1* wei2xiao4.
I at he smile
'I smile at him.'

c. Wo3 *gen1 ta1* hen3 ke4qi4.
I to he very cordial
'I am very cordial to him.'

d. Wo3 *gen1 ta1* zu1 fang2zi.
I from he rent house
'I rent a house from him.'

In 3a, the constituent of *gen1 ta1* is a comitative adjunctive PP and its comitative semantic content is constant regardless of what head verb it co-occurs with. In contrast, in 3b-d, where the same italicized constituent is considered as subcategorized for by the individual verbs, precisely what semantic contribution to the sentence this PP makes is dependent upon the particular verb of that sentence. As a matter of fact, 3b-d all have a comitative reading as well, though such a reading is weak. This of course also comports with this current criterion and indicates that [gen1 NP] is a comitative adjunctive PP, with a constant semantic content, which may appear with a wide range of head verbs; yet, it may also be subcategorized for by a certain subclass of verbs and in this case its semantic content is dependent on the individual verb and thus the more idiosyncratic readings of 3b-d.

3.1.4 Distributional Restrictions

It is widely assumed that subcategorized constituents are usually restricted in their distribution in the constituent structure; non-subcategorized adjuncts, on the other hand, generally have a freer distribution. We shall illustrate with the following examples.

4. a. *Zai4 Tai2bei3*, ta1 chang2 chi1-guan3zi.
at Taipei he often eat restaurant
'In Taipei, he often eats out.'

b. Ta1 *zai4 Tai2bei3* chang2 chi1-guan3zi.
he at Taipei often eat restaurant
'In Taipei, he often eats out.'

c. *Yin1wei4 ta1 bu4 lai2*, wo3 sheng1qi4.
because he not come I angry
'Because he is not coming, I am angry.'

d. Wo3 sheng1qi4 *yin1wei4 ta1 bu4 lai2*.
I angry because he not come
'I am angry because he is not coming.'

a' Wo3 *ba3 ta1* da3 le.
I BA he hit LE
'I hit him.'

b'**Ba3 ta1* wo3 da3 le.
BA he I hit LE
'I hit him.'

c' Wo3 da3du3 *ta1 bu4 lai2*.
I bet he not come
'I bet that he is not coming.'

d'**Ta1 bu4 lai2* wo3 da3du3.
he not come I bet
'I bet that he is not coming.'

The italicized constituents in 4a-b, a PP, and in 4c-d, an S, are both adjuncts to the head verbs and thus show a greater freedom in their syntactic distribution. In comparison, the subcategorized elements in 4a'-d' are rather restricted in their syntactic distribution and violations of such restrictions lead to ungrammaticality. However, this condition is neither necessary nor sufficient for a constituent to be considered subcategorized-for. It merely describes a tendency.

3.1.5 One per Sentence

More than one instance of the same adjunct type may co-occur with the same head verb in a single clause; however, at most one instance of a

particular type of subcategorized constituent may appear with a head verb. Starosta (1988), in his Lexicase theory, distinguishes the adjunctive use of PP's as "outer case" and PP's that are subcategorized for as "inner case." To borrow his terminology, while theoretically there is no limit on the number of instances of an outer case in a clause, there should be at most one instance of each inner case; this is known as the "One-per-Sent" Principle in Lexicase. In LFG, the same effect is achieved by the fact that all subcategorizable grammatical functions have to obey the Functional Uniqueness Condition, while the class of ADJUNCTS does not. Accordingly, 5a-b below are acceptable because the italicized constituents are adjuncts while 5c-d are unacceptable because both of the italicized elements have to be taken as subcategorized-for constituents of the same type, or the same "inner case," and the "One-per-Sentence" condition is thus violated. Yet, 5e is good, for although there are three instances of locative PP's, only the postverbal one is the inner locative case subcategorized for by the verb shui4 'sleep', while the preverbal two are adjuncts.

5. a. *Zai4 Tai2wan1*, ta1 zai4 Tai2bei3 you3 fang2zi.
at Taiwan he at Taipei have house
'In Taiwan, he has houses in Taipei.'

b. Ta1 *wei4 guo2jia1 wei4 ren2min2* xi1sheng1 sheng1ming4.
he for nation for people sacrifice life
'He sacrificed his life for the nation for the people.'

c.*Wo3 *ba3 ta1 ba3 ni3* da3 le.
I BA he BA you hit LE
'*I hit you, him.'

d.*Wo3 shui4 *zai4 chuang2-shang4 zai4 wu1zi-li3*.
I sleep at bed top at house inside
'*I sleep on the bed, in the house.'

e. *Zai4 na4 shi2 zai4 wu1zi-li3* wo3 shui4 zai4 chuang2-shang4.
at that time at house inside I sleep at bed top
'At that time, in the house I slept on the bed.'

3.1.6 Proximity to the Head

The general assumption here is that a subcategorized element tends to appear closer to the verb in proximity than adjuncts. Thus, two useful tendencies can be deduced: a) an element is likely to be subcategorized if it appears closer to the verb than another element which is independently known to be subcategorized; b) an element tends to be adjunctive if it occurs further away from the verb than another constituent known to be an adjunct. These tests, as the way in which they are stated, describe a tendency, not an absolute distinction.

3.1.7 Semantic Selectional Restrictions

While the head verb usually imposes semantic selectional restrictions on its arguments, it rarely does so on adjunctive constituents. This test is closely related to the criterion of constancy of semantic content above.

3.1.8 Possible Internal Gaps

Adjuncts in general tend to disallow unbound gaps; sentential complements, by contrast, allow such long-distance gaps. In the following two noun phrases, 8a's ungrammaticality is thus caused by the relativized subject which is in an adjunctive clause of jia3ru2 'if'. The relativized subject in 8b, however, is in a sentence complement of the verb shuo1 'say' and thus causes no problem.

8. a.*[jia3ru2__bu4 lai2] wo3 hui4 sheng1qi4 de ren2
if not come I will angry de person
'*The person who [if__is not coming] I will be angry.'

b. Ni3 shuo1 [__bu4 lai2] de ren2.
you say not come DE person
'The person who you say [__ is not coming].'

3.1.9 Ability of Being an Indirect Question

An adjunctive clause with a wh-element in it can never be interpreted as an indirect question; rather, the entire sentence has to be interpreted as a question. Yet, a subcategorized predicative complement with a wh-element may be interpreted as an indirect question, though it is also possible for the whole sentence to be interpreted as a question. Logically then, like obligatory co-occurrence, evidence from indirect question constitutes another sufficient but not necessary condition for a constituent to be considered subcategorized-for. Thus, since 9a is ambiguous in that its sentential complement may be a direct or indirect question and that 9b only allows the interpretation of an indirect question, both of the complements in 9a and 9b have to be subcategorized for by the verbs; while neither of the italicized clauses of 9c-d can be an indirect question and thus their status regarding subcategorization has to be determined by other criteria.

9. a. Ta1 xuan1bu4 *ming2tian1 shei2 hui4 lai2*./?
he announce tomorrow who will come
'He announces who is coming tomorrow.'
OR 'Who is coming tomorrow, that he announced?'

b. Ta1men yan2jiu4 *ru2he2 shi3 jing1ji4 qi3fei1*.
they study how make economy take off
'They study how to make the economy take off.'

c. *Ji4ran2 shei2 bu4 lai2*, ta1men dou1 bu4 lai2?
since who not come they all not come
'All of them won't come because who is not coming?'

d. Ni3 bu4 ren4wei2 *shei2 hui4 lai3*?
you not think who will come?
'Who don't you think will come?'

Cheng (1984) conducted a very thorough survey of Mandarin verbs that take statements as arguments, including verbs that may take or require indirect questions. According to this criterion, the embedded clauses of what he calls

verbs of locution, e.g., bao4gao4 'report' and gao4su4 'tell', verbs of cognition, e.g., zhi1dao4 'know' and wang4ji4 'forget' and ask-type verbs should be considered subcategorized for. As for other types of verbs cited by cheng (1984:141-143), their embedded clauses can still be judged as subcategorized complements according to other criteria, such as obligatory co-occurrence, ontological necessity, distributional restrictions, one per sentence, selectional restrictions, and possible internal gaps.

3.1.10 Identification of a Subcategory

Verbs, or more generally predicative elements, in a grammar may be further classified into different subcategories, according to the different kinds of constituents they subcategorize for. It follows logically that a subcategorized constituent should be able to be used to identify a distinctive subcategory of verbs, while adjuncts have no such functions. In the discussion of semantic presupposition of ontological necessity, we mentioned that although all events presuppose a location we do not want to say that therefore all verbs subcategorize for a locative element. Part of the reason is precisely that such a position serves no function in distinguishing a subcategory of verbs from others. Similarly, as discussed earlier in 3.1.3, adjuncts are relatively free to co-occur with a wide range of verbs and thus lack the ability to identify a subcategory of verbs; subcategorized arguments on the other hand only co-occur with a restricted subclass of verbs.

In the following discussion of Mandarin verb subcategorization, we will use evidence from the above tests in justifying our account for Mandarin verb subcategorization. Therefore, it may be worthwhile reiterating that while a couple of the criteria may serve as definitive tests or sufficient conditions for a subcategorized constituent, other conditions may be necessary but by no means sufficient, and still there are others that merely provide a general tendency. Hence, a stronger case can be made when a combination of these conditions is satisfied.

3.2 Mandarin Verb Subcategorization in vLFG

Most of the modern linguistic theories, such as GB, GPSG, and HPSG, express subcategorization requirements in categorial terms like NP, AP, PP,

VP, and S. In the tradition of Case Grammar where the notion of Case Frame, which is similar to that of subcategorization or predicate argument structure, is characterized in terms of the semantic nature of case roles such as agent, patient, source, location, etc. Lexicase is unique in this respect in that it describes the lexical head's "valency" of dependents in semantic (case relations), grammatical (case forms), as well as categorial (+N, +V, etc.) terms. Thus, a clear distinction between arguments and adjuncts can still be made: a required dependent is considered an argument while others are considered adjuncts. Furthermore, as we pointed out earlier, the Lexicase distinction between "inner case" and "outer case," which applies to dependents of verbs as well as nouns, can be directly translated into that between arguments and adjuncts.

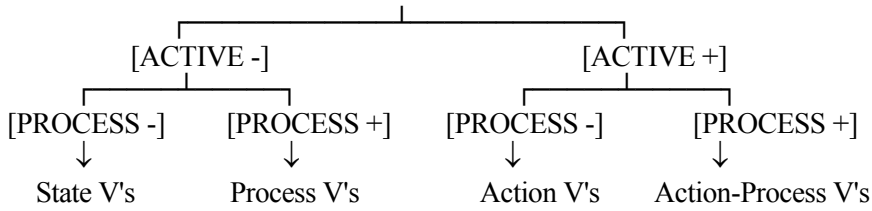
Unlike all the above-mentioned frameworks, LFG proposes that subcategorization be stated in terms of syntactic notions of grammatical functions such as subject, object, etc., in the functional structure. As we have discussed in the previous chapter, subcategorized grammatical functions of a lexical form may be assigned by thematic roles, and in fact they generally are. Perceivably then, a subcategorization scheme in LFG can be based on thematic roles as well. Nonetheless, as we have pointed out before, Lexical Mapping Theory does not account for non-thematically assigned functions. The existence of non-thematic grammatical functions (functions not assigned by, or not mapped with, thematic roles) thus ruins the otherwise one-to-one correspondence between thematic roles and subcategorized grammatical functions. Consequently, subcategorization requirements and restrictions have to be specified in terms of grammatical functions on lexical forms in the lexicon. Once again, in our vLFG formalism, the PRED attribute is reserved for the specification of subcategorization requirements of lexical forms only.

3.2.1 Activity and Stativity

Before we proceed to discuss each of the subcategories of verbs in Chinese, we should point out the importance of the traditional distinction of active versus stative verbs. Although this distinction does not fit in our scheme of the syntactic subcategorization of verbs in terms of grammatical functions, this distinction, which is semantic in nature, is very useful to account for verbs' gradability and ability to be marked with aspect particles, as we shall explicate

below. However, such a simple distinction between stativity and activity may not be adequate for Mandarin verbs and therefore the commonly held assertions that 1) active verbs may be marked with aspect while stative verbs may not and 2) stative verbs may take degree adverbs while active verbs may not are not entirely accurate (e.g. Chang et al 1988:416). We shall adopt a finer distinction using two binary features, ACTIVE and PROCESS, proposed by Her (1985-6), where ACTIVE is said to indicate the initialization of an action or event and PROCESS the change of state. Four subcategories of verbs are recognized: state, process, action, action-process. Note that the following table depicts the four subcategories and there is no theoretical significance attached to the higher and lower branching.

Table 3.1
Semantic Subcategories of Mandarin Verbs



gao1	bing4	pao3	sha1si3
'tall'	'sick'	'run'	'kill'
cong1ming2	si3	qu4	zhang3da4
'smart'	'die'	'go'	'grow up'
que4ding4	chen2 (intran.)	gong1zuo4	fa1fu2
'sure'	'sink'	'work'	'become fat'
xiang3nian4	kai1	shuo1	da3kai1
'miss'	'open' (intran.)	'say'	'open' (tran)

We thus set up four feature inheritance (FI) entries for these four classes:

FI-V-STATE:
[FS [ACTIVE -
PROCESS -
]
]

FI-V-PROCESS:
[FS [ACTIVE -
PROCESS +
]
]

FI-V-ACTION:
[FS [ACTIVE +
PROCESS -
]
]

FI-V-ACTPRO:
[FS [ACTIVE +
PROCESS +
]
]

In terms of gradability, only a state verb, [ACTIVE - PROCESS -], may be gradable, or to be modified by adverbial degree elements such as hen3 'very', comparative [bi3 NP] and geng4 'more', and superlative zui4 'most'. As we can see in the following examples, only state verbs are gradable.

1. a. Ta1 hen3 xiang3nian4 ni3. `state V
he very miss you
'He misses you very much.'

b.*Ta1 hen3 bing4-le. `process V
he very sick LE
'He is very ill.'

c.*Ta1 hen3 qu4 Tai2wan1. `action
he very go Tai2wan1.
'*He very goes to Taiwan.'

d.*Ta hen3 da3kai1 men2. `action-process
he very open door
'*He very opens the door wide.'

A certain degree of idiosyncracy however still exists among state verbs regarding their ability to be modified by degree adverbs. It seems that not all state verbs are gradable. Yao4 'want', for instance, is semantically very similar to xiang3 'want' in a pivotal construction and yet the former can never co-occur with degree adverbs. Likewise, although ren4wei2 'think' and xiang3 'think' are semantically equivalent when they take a sentential complement, the latter cannot be modified by degree adverbs. Such idiosyncracies of course still need to be accounted for in the lexical entries of individual verbs.

2. a. Wo3 hen3 xiang3 qu4.
I very want go
'I want to go very much.'

b.*Wo3 hen3 yao4 qu4.
I very want go
'I want to go very much.'

c. Wo3 hen3 bu4 ren4wei2 ta1 hui4 lai2.
I very not think he will come
'I really don't think he will come.'

d.*Wo3 hen3 bu4 xiang3 ta1 hui4 lai2.
I very not think he will come
'I really don't think he will come.'

While most verbs may be marked with aspect markers perfective le (i.e., postverbal le, not the sentence-final le) and experiential guo4, only verbs with

[ACTIVE +] or [PROCESS +] may be marked with the progressive aspect zhe, or the adverb zheng4zai4 'progressively'. In other words, only state verbs may not be marked with progress aspect, as illustrated with the examples below.

3. a. Ta1 xian4zai4 cong1ming2-le. `state V
he now smart LE
'He is now smart.'

a' Ta1 mei2you3 cong1ming2-guo4
he not smart GUO
'He has never been smart.'

a"*Ta1 cong1ming2-zhe.
he smart ZHE
'He is being smart.'

b. Chuan2 chen2-le. `process V
ship sink LE
'The ship has sunk.'

b' Chuan2 chen2-guo4.
ship sink GUO
'The ship has sunk before.'

b" Chuan2 chen2-zhe.
ship sink ZHE
'The ship is sinking.'

c. Ta1 shuo1-le ni3 hui4 lai2. `action V
he say LE you will come
'He has said that you would come.'

c' Ta1 shuo1-guo4 ni3 hui4 lai2.
he say GUO you will come
'He has said before that you would come.'

c" Ta1 shuo1-zhe ni3 hui4 lai2.
he say ZHE you will come
'He is saying that you will come.'

d. Ta1 fa1fu2 le. `action-process
he become-fat LE.
'He has become fat.'

d' Ta1 yi3qian2 fa1fu2 guo4.
he before become-fat GUO.
'He has become fat before.'

d" Ta1 na4 shi2hou4 zheng4zai4 fa1fu2 zhe.
he that time progressively become-fat GUO.
'At that time he was becoming fat.'

This semantic classification may also be relevant to the description of the morpholexical process of resultative compounding, where an action verb, [ACTIVE + PROCESS -], which may be either transitive or intransitive, is joined by an [ACTIVE -] verb, i.e., either a state verb or a process verb, to form an action-process verb, [ACTIVE + PROCESS +]. For instance,

[V1]	+	[V2]	=	Resultative Compound
da3 'hit'		kai1 'open'		da3kai1
ti1 'kick'		po4 'broken'		ti1po4
zha4 'bomb'		chen2 'sink'		zha4chen2
Action verbs		Process Verbs		Action-Process Verbs

[V1]	+	[V2]	=	Resultative Compound
pao3 'run'		lei4 'tired'		pao3lei4
zhang3 'grow'		gao1 'tall'		zhang3gao1
kan4 'see'		qing1chu3 'clear'		kan4qing1chu3
Action verbs		State Verbs		Action-Process Verbs

Similarly, in the syntactic construction of a verb joined by an adverbial phrase maker such as wan2 'finish' and qi3 and qi3lai2 'start' (Cheng 1990 and personal communication), the verb has to be an action verb, thus [ACTIVE + PROCESS -], as shown in the following examples.

3. a. Ta1 kan4 wan2 le shu1.
he read finish LE book
'He finished reading the book.'

b. Ta1 kan4 qi3 le shu1.
he read start LE book
'He started reading the book.'

Among the four verb classes here, no doubt finer distinctions can still be made and subclasses identified according to other semantic features, and derivational relations may exist among these semantic verb classes (Her 1985-6). The detailed study of such derivational relations and the co-occurrence restrictions and requirements between the semantic verb subcategories and elements of other syntactic and semantic categories would be beyond the scope of this study; however, such a study as well as the further exploration of the interaction between the semantic classification and the syntactic subcategorization of grammatical functions should be an essential part of the description of Mandarin verbs.

3.2.2 Syntactic Subcategorization

We have established in the previous chapter that the following subcategorizable grammatical functions are recognized in Chinese: SUBJ, OBJ, OBJ2, NCOMP, XCOMP, SCOMP, OBL_{THME}, OBL_{BNFC}, OBL_{LOCT}, and

OBL_{GOAL}; in other words, they are the only allowable values for PRED, which takes a list of the above functions as its value. We will first list below all the subcategories of verbs we have observed in Mandarin Chinese in terms of the grammatical functions they subcategorize for. Note that the order of the function name is random and thus insignificant; yet, for the sake of convenience we will follow the linear order of their appearance in c-structures.

1. <SUBJ>
2. <SUBJ , OBL_{LOCT}>
3. <SUBJ , OBL_{GOAL}>
4. <SUBJ , OBJ>
5. <SUBJ , OBL_{THME}>
6. <SUBJ , OBL_{THME} , OBJ>
7. <SUBJ , OBL_{GOAL} , OBJ>
8. <SUBJ , OBJ , OBL_{LOCT}>
9. <SUBJ , OBL_{THME} , OBL_{LOCT}>
10. <SUBJ , OBL_{THME} , OBJ , OBL_{LOCT}>
11. <SUBJ , NCOMP>
12. <SUBJ , XCOMP>
13. <SUBJ , SCOMP>
14. <SUBJ , OBJ , OBJ2>
15. <SUBJ , OBJ , OBL_{BNFC}>
16. <SUBJ , OBL_{THME} , OBJ , OBJ2>
17. <SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}>
18. <SUBJ , OBL_{THME} , OBL_{BNFC}>
19. <SUBJ , OBL_{GOAL} , OBL_{THME}>
20. <SUBJ , OBL_{GOAL} , OBL_{THME} , OBJ>
21. <SUBJ , OBJ , XCOMP>
22. <SUBJ , OBJ , SCOMP>
23. <SUBJ , OBL_{GOAL} , XCOMP>
24. <SUBJ , OBL_{GOAL} , SCOMP>
25. <SUBJ , OBJ , OBJ2 , SCOMP>
26. <SUBJ , OBL_{GOAL} , OBJ , SCOMP>

Many verbs are not limited to one subcategorization; they seem to have "polyvalency," to borrow the term from Pollard and Sag (1987). In our analysis, we do not allow optional subcategorized functions; thus, if a verb has polyvalency, we treat each as a distinctive lexical entity. For example, the verb shuo1 'say, speak' may have at least five subcategorizations as shown below, and therefore in the lexicon it is treated as five distinctive lexical forms, each with its own PRED value.

1. Wo3 shuo1 ying1wen2 he2 fa4wen2. `<SUBJ , OBJ>
   I speak English and French
   'I speak English and French.'

2. Wo3 gen1 ta1 shuo1 gu4shi4. `<SUBJ , OBL_{GOAL} , OBJ>
   I with he say story
   'I tell him stories.'

3. Wo3 shuo1 Ma3li4 hui4 lai2. `<SUBJ , SCOMP>
   I say Mary will come
   'I say that Mary will come.'

4. Wo3 gen1 ta1 shuo1 bie2 qu4.
   I with him say not go
   'I say to him not to go.'

5. Wo3 dui4 ta1 shuo1 ni3 hao3. `<SUBJ, OBL_{GOAL} , SCOMP>
   I to him say you good
   'I say to him that you are good.'

shuo1:
{ [CAT V `1
FS [FORM 'shuo1'
PRED <SUBJ , OBJ>
]]

[CAT V `2
FS [FORM 'shuo1'
PRED <SUBJ , OBL_{GOAL} , OBJ>
]]

[CAT V `3
FS [FORM 'shuo1'
PRED <SUBJ , SCOMP>
]]

[CAT V `4
FS [FORM 'shuo1'
PRED <SUBJ , OBL_{GOAL} , XCOMP>
]]

[CAT V `5
FS [FORM 'shuo1'
PRED <SUBJ , OBL_{GOAL} , SCOMP>
]]
}

3.2.3 Optionality in Subcategorization

There are several reasons why we have decided not to allow optionality in stating subcategorization, i.e., the value of PRED. First of all, given the newly developed Lexical Mapping Theory and the Argument-Function Biuniqueness Condition, a grammatical function in PRED has to either be assigned by a thematic role or be a syntactically required non-thematic function. To allow optional functions in PRED would have to mean optional thematic roles in the predicate argument structure. Consequently, the

Argument-Function Biuniqueness Condition will be compromised. Secondly, although it would seem that to allow optionality would cut down the number of subcategorization patterns, this may not always be true. For instance, given a verb like kai1 'open' which may be both transitive or intransitive, if we allow optionality, we will have to allow a pattern of <SUBJ , (OBJ)>. That is one more pattern in addition to the two patterns, <SUBJ> and <SUBJ , OBJ>, which are needed independently. Thus, to allow optionality we need to have three patterns; without optionality we can get by nicely with simply two. Furthermore, in terms of computational efficiency, optionality poses no clear advantages. Finally, to allow optionality in PRED's value would compromise the integrity of the Completeness and Coherence Conditions on f-structures.

3.2.4 Derivational Relations between Verb Classes

When a class of verbs of certain subcategorization can also take another pattern of subcategorization, often there is a derivational relation between the two classes. In LFG, the derivational relatedness between two verb classes, e.g., active forms of verbs and passive forms, is accounted for by morpholexical rules in Lexical Mapping Theory. Actually, the "lexical" part of the name of LFG is to indicate this commitment to characterize processes that alter the "valency," or subcategorization, of predicates in the lexicon, while in a transformational framework such processes are often accounted for by transformations, such as Move-α in GB. Although it is beyond the scope of this book to account for derivational relations among verb subcategories, we will point out the likely existence of such relationship in our discussion, and here we will give one specific example to illustrate how such relationship may be accounted for in LFG. Following the model presented in Bresnan and Kanerva (1989), we will account for the derivational relation between verbs in 1 and the so-called "locative inversion" verbs in 2.

1. a. Zhang1san1 zuo4 zai4 qi4che1- li3.
Zhangsan sit at car inside
'Zhangsan sits in the car.'

b. Hen3duo1 du1cha2 lai2 dao4 xue2xiao4.
many inspector come to school
'Many inspectors came to the school.'

c. Ta1 de hua4 gua4 zai4 qiang2-shang4.
he DE painting hang at wall top
'His painting is hanging on the wall.'

2. a. Qi4che1-li3 zuo4 zhe Zhang1san1.
car inside sit ZHE Zhangsan
'In the car there sits Zhangsan.'

b. Xue2xiao4 lai2 le hen3duo1 du1cha2.
school come LE many inspector
'The school had many inspectors arrived.'

c. Qiang2-shang4 gua4 zhe ta1 de hua4.
wall top hang ZHE he DE painting
'One the wall, there hangs his painting.'

Presumably, the verbs in both sets require a theme role and a location role. However, while verbs in 1a-c subcategorize SUBJ and OBL_{LOCT} (see our discussion in III.2. of this chapter), in 2a-c they subcategorize SUBJ and OBJ (see III.4.C). We will see how Lexical Mapping Theory may account for this derivational process of "locative inversion."

Four components comprise Lexical Mapping Theory: (a) a thematic hierarchy of semantic roles, (b) a classification of grammatical functions along two binary features, (c) principles of lexical mapping from semantic roles to the feature specification of grammatical functions, and (d) well-formedness conditions.

The following universal thematic hierarchy is assumed: agent > ben > goal/exp > inst > theme/pat > loc. Grammatical functions are classified by binary features: $\pm$ r (restricted, thematically) and $\pm$ o (object-like), as shown in the following chart.

	-r	+r
+o	OBJ	OBJ2
-o	SUBJ	OBL

Figure 3.1 Classification of Grammatical Functions

The lexical mapping principles are of three types: (a) intrinsic assignments, (b) morpholexical operations, and (c) default assignment. First of all, the following three intrinsic classifications are assumed:

3. th/pat → [-r]
ag → [-o]
loc → [-o]

Morpholexical rules are language-specific operations that affect feature assignment by adding features or suppressing thematic roles. To account for locative inversion in Chinese, we pose a morpholexical rule that allows the locative role to be optionally assigned [-r] when in the predicate argument structure of the lexical item the other role is theme.

4. Locative Inversion (optional): <th loc>
↓
[-r]

The default classification assigns the feature [-r] to the highest role in the hierarchy and [+r] to all others. Note that any feature assignment has to be feature-preserving. Thus, later assignment may not conflict with existing specifications. Finally there are the following well-formedness conditions on lexical forms.

5. a. Subject Condition:
every lexical form must have a SUBJ function.

b. Function-Argument Biuniqueness:
every argument must have a unique grammatical function and vice versa.

Now, let's use the verb zuo4 'sit' as an example and see how the morpholexical rule of "lexical inversion" relate its two 'valencies': <SUBJ , OBL_{LOCT}> and <SUBJ , OBJ>. We will step through the process of mapping thematic roles to grammatical functions through feature assignment by intrinsic classification, morpholexical rules, and default classification.

6.	zuo4 'sit' < theme	loc >	
Intrinsic	[-r]	[-o]	
Default		[+r]	
Class. of functions	S/O	OBL	
Subject Cond.	SUBJ	OBL	`in 1a

7.	zuo4 'sit' < theme	loc >	
Intrinsic	[-r]	[-o]	
Locative Inversion		[-r]	
Class. of functions	S/O	SUBJ	
Biuniqueness Cond.	OBJ	SUBJ	`in 2a

In 6, the optional locative inversion rule did not apply and what we get is the subcategorization of <SUBJ , OBL> of 1a. After the morpholexical rule applies in 7, the default classification can no longer apply without violating the feature-preserving principle. The morpholexical rule of locative inversion thus reveals the derivational relationship between verbs of 1a-c and 2a-c.

For more examples, refer to Bresnan and Kanerva (1989) or Huang (1989). Our study of verb subcategorization of grammatical functions presented in the following sections and Huang's preliminary application of Lexical Mapping Theory to Chinese provide a good starting point for the investigation and formal description of the intricate network of derivational processes among verb subcategories. For readers interested in such an undertaking, we urge that Starosta's criticism regarding the lack of linguistic relevance in the situation-based identification procedure of thematic relations in the Chomskyan tradition based on Gruber (1965) or case roles in the Fillmorean tradition (Fillmore 1968) be taken seriously and that his proposal of the

perception-based concept of thematic relations be taken into consideration (Starosta 1988).

3.3 Subcategorization of Grammatical Functions

For each of the verb subcategories we have identified, we will give an ample number of examples of verbs belonging to this subcategory, along with several illustrative sentences, and we will explain the grammatical characteristics of verbs in the subcategory.

3.3.1 <SUBJ>

3.3.1.1 Intransitive Verbs

Examples: qian1dao4 'sign in', fei1 'fly', san4bu4 'stroll', ge1chang4 'sing', yong4gong1 'hard-working', hao3 'good', 'jian4kang1 'healthy', he2ai3ke3qin1 'kind', tou2teng2 'worried', xin1suan1 'sad', liang4 'lit', 'kai1 'open', diao4 'lost', chen2 'sunk', qi1 'paint', xi3 'wash', kan4 'read', si1po4 'torn', da3kai1 'open', liang4gan1 'line-dry'

1. a. Niao3 fei1-le.
   bird fly LE
   'The bird has flown away.'

   b. Ta1 hen3 tou2teng2.
   he very headache
   'He is very worried.'

   c. Chuan2 chen2 le.
   ship sunk LE
   'The ship has sunk.'

   d. Zuo4ye4 zuo4hao3 le.
   homework done LE
   'The homework is done.'

Verbs of this class are commonly known as intransitive verbs that subcategorize a single function, SUBJ. An intransitive verb typically requires an actor-like subject, such as 1a above. However, SUBJ, being a semantically non-restricted function, may be assigned by other semantic roles as well. This is evident in the subjects of 1a-d above.

The majority of state verbs in Chinese also belong to this subcategory. The subject of an intransitive state verb is more likely to be a theme-like NP. Mandarin Chinese is clearly a adjectival-verb language in that it offers no consistent basis for distinguishing verbs with adjectival meanings from other verbs (e.g. Chao 1968, Schachter 1985). However, there are some state verbs that also function like genuine adjectives in that they may modify a noun attributively in a non-relative construction; most of the adjectives are mono-syllabic, such as da4 fei2 zhu1 'a big fat pig' and huang2 che1 'yellow cars' but not *fen3hong2 che1 'pink cars'. However, not all mono-syllabic state verbs may appear in such a construction, for example *gui4 che1 'expensive cars'. On the other hand, some bi-syllabic state verbs do appear in such construction, for example cong1ming2 ren2 'smart people'. These idiosyncratic facts have to be accounted for in the lexicon. Clearly then, we treat lexical items with the same phonological representation and similar meaning but with different syntactic distribution as separate, distinct lexical forms. Recall that {} indicates a set of homographs.

```
FI-V-1:
        [ CAT V
          FS [ PRED < SUBJ > ]
        ]

cong1ming2:
{       [ FS [ FORM 'cong1ming2' ]    `verb
          FI-V-STATE
          FI-V-1
        ]
        [ CAT A                       `adjective, attributive
          FS [ FORM 'cong1ming2' ]
        ]
 }
```

gui4:
[FS [FORM 'gui4' `verb
]
FI-V-STATE
FI-V-1
]

The use of these attributive adjectives, though very common in classical Chinese, is extremely limited in modern Mandarin; these adjectives do not allow any kind of modification, for instance, *hen3 da4 zhu1 'very big pigs' or *you4 da4 you4 fei2 zhu1 'pigs big and fat'. It seems they have to be used in isolation and once they are modified in any way they would have to be in a relative construction headed by de to be well-formed: hen3 da4 de_zhu1 and you4 da4 you4 fei2 de zhu1. The many restrictions of the use of adjectives may prove that it is a residue from a historical change and that it is probably on its way out.

The class of process verbs also largely belongs to this subcategory, as exemplified by 1c-d. These verbs often take a patient-like NP as subject and are sometimes referred to as "unaccusative" verbs, according to the use by Perlmutter (1978). Noticeably, many, though not all, of these unaccusative verbs also have accusative usages, which indicates that the former is derivationally related to the later. We will illustrate with examples below. In the current theory of LFG, such derivational relations, such as between the accusative and unaccusative verbs and between the active and passive verbs, are accounted for by lexical rules in the component of the Lexical Mapping Theory.

1. e. Chuan2 chen2 le.
ship sunk LE
'The ship has sunk.'

f. Ta1 ba3 chuan2 chen2 LE.
he BA ship sunk LE
'He sank the ship.'

g. Zuo4ye4 zuo4hao3 le.
homework done LE
'The homework is done.'

h. Ta zuo4hao3 le zuo4ye4.
he done LE homework
'He has done the homework.'

3.3.1.2 Weather Verbs

Examples: di4zhen4 'have earthquake', xia4yu3 'rain', qi3feng1 '(wind) blow', da3lei2 'thunder'.

2. a. Nei4 ge di4fang1 chang2chang2 di4zhen4.
that CLS place often earthquake
'That place often has earthquakes.'

b. Hai3-shang4 qi3feng1 le.
sea top wind-blow LE
'The wind is blowing on the sea.'

c.*Hai3 qi3feng1 le.
sea wind-blow LE
'The wind is blowing on the sea.'

d. You4 xia4yu3 le.
again rain LE
'It's raining again.'

Most weather verbs belong to this class; they may either take an overt SUBJ that has to be a place word, thus with the feature-value pair [PLACE +], or they may allow no overt SUBJ, thus 2d above does not leave one feeling the sentence is incomplete. Therefore, in effect we are saying that this type of verb subcategorizes, optionally, a locative constituent. Since we take the position that all verbs must subcategorize at least SUBJ, this locative constituent, when overt, has to be the subject. In the lexical entries we set up a

"dummy," or default, SUBJ to account for the sentences where there is no overt subject. Note that the value OPT is a placeholder: it always succeeds in unification. Thus, when there is an overt subject such as in 2a-b above, the overt value of its FORM will replace OPT. If the subject is not a place word, for example 2c, and thus with [PLACE -], it would not unify with the verb's functional structure. When there is no overt subject, e.g., 2d, the f-structure will always have SUBJ [FORM OPT PLACE +] and therefore will not be judged as incomplete.

```
FI-V-1-2:
        [ FS [ SUBJ [ FORM OPT
                      PLACE +
                    ]
             ]
        ]

di4zhen4:
        [ FS [ FORM 'di4zhen4' ]
          FI-V-1-2
          FI-V-1
        ]
```

2d-f. you4 xia4yu3 le. 'It is raining again.'

```
        [ SUBJ [ FORM OPT
                   PLACE +
                 ]
          FORM 'xia4yu3'
          PRED <SUBJ>
          ADJ [ FORM 'you4' ]
          ASPECT PERFECT
        ]
```

3.3.2 <SUBJ , OBL_{LOCT}>

Examples: zhu4 'live', tiao4 'jump', pao3 'run', zhan4 'stand', ban1 'move, relocate', chuang3 'barge', zhuang4 'hit (intran.)', fei1 'fly', tang3 'lie', kao4

'lean', zuo4 'sit', qu4 'go', hui2 'return', gua4 'hang', tie1 'paste', cha1 'stick', zhong4 'plant'.

1. a. Ta1 zhu4 zai4/dao4 shan1-shang4.
he lives at /to hill top
'He lives on the hill.'

b. Ta1 ban1 *zai4/dao4 Tai2bei3.
he move at /to Taipei
'He moved to Taipei.'

c. Ta1 tiao4 zai4/dao4 shui3-li3.
he jump at /to water inside
'He jumps in the water.'

d. Ta1 hui2 *zai4/dao4 shui3-li3.
he return at /to water inside
'He returned into the water.'

Postverbal subcategorized locative PPs in Chinese are well-accepted in the current literature (e.g., Huang 1982, Huang 1989, Ernst 1989). First of all, some verbs in this class, such as zhu4 'live' and chuang3 'barge', do require the postverbal locative PP obligatorily. There is only one postverbal locative PP allowed and it always appears immediately after the verb. Semantically, verbs of this subcategory denote an action which crucially depends on the manner of spatial orientation (Ernst 1989:118); thus, most of these verbs denote some sort of locomotion. The oblique grammatical function OBL_{LOCT} may be fulfilled by locational prepositional phrases such as [zai4 NP] and [dao4 NP]. Both prepositions have [PCASE LOCT] specified in their entries but with their own unique PFORM. Individual verbs in this class may have arbitrary requirements in selecting the particular prepositions they may co-occur with. Such information must be specified in the individual verb entry.

FI-V-2:

[CAT V
  FS [PRED <SUBJ , OBL_{LOCT}>]
  (↑ OBL_{LOCT} PCASE) $=_c$ LOCT
]

zhu4:

[FS [FORM 'zhu4'
     ]
  FI-V-2
]

ban1:

[FS [FORM 'ban1'
     ]
  (↑ OBL_{LOCT} PFORM) $=_c$ 'dao4'
  FI-V-2
]

zai4:

[CAT P
  FS [PFORM 'zai4'
       PCASE LOCT
     ]
]

dao4:

[CAT P
  FS [PFORM 'dao4'
       PCASE LOCT
     ]
]

However, unlike 1a-d where the subject of the verbs is an actor-like NP, other verbs in this subcategory are unaccusative verbs that subcategorize a patient-like subject and a locative PP. These verbs, such as the ones in 1e and

1f below, are derivationally related to verbs that subcategorize <SUBJ , OBJ , OBL_{LOCT}>, such as verbs in the primed sentences below.

1. e. Yi1 fu2 hua4 gua4 zai4 mei3shu4guan3-li3.
one CLS painting hang at museum inside
'A painting hangs in the museum.'

e' Wo3 gua4 yi1 fu2 hua4 zai4 mei3shu4guan3-li3.
I hang one CLS painting at museum inside
'I hang a painting in the museum.'

f. Da4 shu4 zhong4 zai4 Tai2bei3.
big tree plant at Taipei
'The big tree was planted in Taipei.'

f' Wo3 zhong4 da4 shu4 zai4 Tai2bei3.
I plant big tree at Taipei
'I plant a big tree in Taipei.'

3.3.3 <SUBJ , OBL_{GOAL}>

Examples: ren2ci2 'kind', hao3 'good', zun1jing4 'respectful', ke4qi4 'polite', guan1xin1 'care' ; xiao4 'laugh', wei2xiao4 'smile', ge1chang4 'sing', zuo4ai4 'make love', shi4ai4 'show affection', da3jia4 'fist-fight', hu2shuo1ba1dao4 'talk nonsense'.

1. a. Ta1 gen1/he2/tong2/xiang4/dui4 ni3 wei2xiao4.
he with/with/with /toward /to you smile
'He smiles at you.'

b. Ta1 gen1/*he2/tong2/*xiang4/dui4 ni3 ren2ci2.
he with/with/with /toward/to you kind
'He is kind to you.'

c. Ta1 gen1/*he2/tong2/*xiang4/dui4 ni3 ke4qi4.
he with/with/with /toward/to you polite
'He is polite to you.'

d. Ta1 gen1/he2/tong2/xiang4/dui4 ni3 shi4ai4.
he with/with/with /toward/to you show-affection
'He shows affection to you.'

It is sometimes assumed that Chinese is a "right-branching" language where arguments appear postverbally (e.g., Huang 1982). However, we agree with Ernst (1989) where he argues that there is evidence that certain preverbal PPs in Chinese should be considered subcategorized as well. First of all, semantically, verbs here usually denote an emotion or action that crucially depends on the existence of a correspondent. We assign this correspondent constituent to the grammatical function of OBL_{GOAL}. Secondly, in our discussion of constancy of semantic content of non-subcategorized dependent constituents, we have pointed out that unlike the adjunctive use of [gen1 NP] which contributes a constant reading as a comitative PP, in sentences 2a-d [gen1 NP] has a variety of semantic content, other than the comitative reading. Furthermore, in sentences 1a-d above, it is also obvious that verbs in this subcategory have to arbitrarily select particular prepositions for this subcategorized constituent. We will present another piece of evidence from criterion of proximity to the head, in relation to the [ba3 NP], which we argue is a subcategorized constituent, in a later section.

One of the following prepositional phrases: [gen1 NP], [dui4 NP], [xiang4 NP], [he2 NP], and [tong2 NP] may fulfill the OBL_{GOAL} function. Since different verbs in this class may require different prepositions rather arbitrarily, such information must be specified in the individual verb entry. All these above preposition have [PCASE GOAL] but each has its own unique PFORM.

FI-V-3:
[CAT V
FS [PRED <OBL_{GOAL} , SUBJ>]
(↑ OBL_{GOAL} PCASE) =$_c$ GOAL
]

shi4ai4:
[FS [FORM 'shi4ai4']
FI-V-3
]

xiang1tong2:
[FS [FORM 'xiang1tong2']
(↑ OBL_{GOAL} PFORM) =$_c$ { 'gen1' 'he2' 'tong2' }
FI-V-3
]

ke4qi4:
[FS [FORM 'ke4qi4']
(↑ OBL_{GOAL} PFORM) =$_c$ { 'gen1' 'tong2' 'dui4' }
FI-V-3
]

guan1xin1:
[FS [FORM 'guan1xin1']
(↑ OBL_{GOAL} PFORM) =$_c$ 'dui4'
FI-V-3
]

gen1:
[CAT P
FS [PFORM 'gen1'
PCASE GOAL
]
]

dui4:

```
[ CAT P
  FS [ PFORM 'dui4'
       PCASE GOAL
     ]
]
```

3.3.4 <SUBJ , OBJ>

3.3.4.1 Typical Transitive Verbs

Examples: jian4 'construct', fa1she4 'shoot', gou4mai3 'buy', fan4mai4 'sell', bai4fang3 'visit', tan2lun4 'discuss', ji1xiao4 'laugh at', fu4ze2 'be responsible for', xi3huan1 'like', ai4 'love', guan1xin1 'be concerned with', xiang3nian4 'miss', si3 'die', ge1 'cut', da3po4 'break', da3duan4 'break', ge1shang1 'cut', chi1 'eat', wei4 'feed', rong2na4 'accommodate', shui4 'accommodate (the sleeping)', zuo4 'accommodate (the sitting)', zhan4 'accommodate (the standing)', ji3man3 'crowded'.

1. a. Ta1 fa1she4 huo3jian4.
   he shoot rocket
   'He shoots rockets.'

   b. Ta1 da3po4 hua1ping2.
   he break vase
   'He broke the vase.'

Verbs belonging to this class are known as transitive verbs. A transitive verb typically requires an agent-like subject and a patient-like object, for example 1a-b above. However, again due to the semantic unrestrictive nature of SUBJ and OBJ, they may be linked with different thematic roles. For example, some state verbs may also function transitively with a rather goal-like object.

1. c. Ta1 guan1xin1 ren2min2.
he concerned people
'He is concerned about people.'

d. Ta1 dui4 ren2min2 guan1xin1.
he to people concerned
'He is concerned about people.'

FI-V-4:
[CAT V
FS [PRED <SUBJ , OBJ>]
]

guan1xin1:
{ [FS [FORM 'guan1xin1']
(↑ OBL_{GOAL} PFORM) $=_c$ 'dui4'
FI-V-3
]
[FS [FORM 'guan1xin1']
FI-V-4
]
}

Many unaccusative verbs may also function transitively in a rather peculiar construction, exemplified below in 1e-g, where their subjects are rather patient-like NPs; yet, they also have a direct object that is also clearly patient-like. The semantic relation between the subject and the object is largely that of an inalienable possession or intimate part.

1. e. Ta1 si3 le mu3qin1.
he die LE mother
'He had his mother died on him.'

e' Mu3qin1 si3 le.
mother die LE
'(His) mother died.'

e"*Wo3 ba3 ta1 si3 le mu3qin1.
I BA he die LE mother
'*I died his mother.'

f. Ta1 ge1po4 le shou3.
he cut LE hand
'His hand was cut.'

f' Shou3 ge1po4 le.
hand cut LE
'(His) hand was cut.'

f" Wo3 ba3 ta1 ge1po4 le shou3.
I BA he cut LE hand
'I cut his hand.'

g. Ping2zi zhuang1man3 le shui3.
bottle fill-full LE water.
'The bottle is filled with water.'

g' Ping2zi zhuang1man3 le.
bottle fill-full LE
'The bottle is filled.'

g" Wo3 ba3 ping2zi zhuang1man3 le shui3.
I BA bottle fill-full LE water
'I filled the bottle with water.'

Since the object is universally more a typical position for patient and furthermore in this construction it is indeed the object that directly receives the action or undergoes the process denoted by the verb, we consider the object a patient. The subject is indirectly affected by the action or process; we thus take the subject to be a theme-like element. This is also confirmed by comparing 1e-g with their intransitive unaccusative counterparts 1e'-f'. This also provides us some evidence to treat [ba3 NP] as assigning grammatical function related to theme-like role. Also, as shown with the double-primed

sentences above, most verbs here belong to the subcategorization of <SUBJ , OBL_{THME} , OBJ> as well, but si3 'die' clearly is not since it does not have a ba3 counterpart.

A small number of transitive verbs, such as chi1 'eat, feed', shui4 'accommodate (the sleeping)', seem to be able to have an unaccusative use where they take a patient-like subject and an agent-like object. We will look at some examples of these sentences.

1. h. Na4 fang2zi zhu4 le hen3duo1 ren2.
that house live LE many person
'That house accommodates many people.'
OR 'There are many people living in that house.'

i. Na4 ba3 yi3zi zuo4 le wu3 ge ren2.
that CLS chair sit LE five CLS person
'That chair accommodated the sitting of five people.'
OR 'On that chair there were five people sitting there.'

j. Yi1 bang4 rou4 chi1 san1 ge ren2.
one pound meat eat three CLS person
'One pound of meat feeds three people.'

No doubt this use of these verbs is more a marked one compared to their normal use and that they are derivationally related to each other. However, not all verbs in this subclass are so related, for example rong2na4 'accommodate', which do not have a counterpart that takes agent subject and patient object.

3.3.4.2 Verbs Requiring a Locative OBJ

Examples: qu4 'go', lai2 'come', hui2, hui2lai2, hui2qu4 'return', di3da2, dao4da2 'arrive', fei1 'fly', zhu4 'live', yi2ju1, qian1ju1 'move, relocate', yi1min2 'immigrate into'.

2. a. Ta1 lai2 xue2xiao4.
he come school
'He comes to the school.'

b.*Ta1 lai2 fang2zi.
he come house
'He comes to the house.'

c. Ta1 lai2 fang2zi li3.
he come house inside
'He comes into the house.'

d. Ta1 lai2 dao4 xue2xiao4.
he come to school
'He comes to school.'

Verbs in this category differ from the previous one in that the OBJ here must be place words, thus with [PLACE +]. Although, as we can see from the examples above, most of the verbs here may also have the subcategorization of <SUBJ , OBL_{GOAL}> (as seen in 2d), a few of them, such as di3da2 and dao4da2 'arrive', cannot appear with a locative PP and require a locative NP. The verb qu4 'go', for some speakers at least, cannot take a locative PP.

FI-V-4-2:
[(↑ OBJ PLACE) $=_c$ +
]

di3da2:
[FS [FORM 'di3da2']
FI-V-4
FI-V-4-2
]

qu4:
{ [FS [FORM 'qu4']
(↑ OBL_{LOCT} PFORM) $=_c$ 'dao4'
FI-V-3
]

```
        [ FS [ FORM 'guan1xin1' ]
          FI-V-4
          FI-V-4-2
        ]
}
```

3.3.4.3 Locative Inversion Verbs

Examples: fa1sheng1 'happen', chu1xian4 'appear', mao4chu1 'appear', lai2 'come', zuo4 'sit', zhan4 'stand', tang3 'lie', shui4 'sleep', tiao4chu1 'jump out of', zuo3jin4 'walk into', gua4 'hang', tie1 'paste', xie3 'write', hua4 'paint', zhong4 'plant', cha1 'stick'.

3. a. Jia1-li3 lai2-le ke4ren2.
home inside come LE guest
'There are some guests who came to my home.'

a' Ke4ren2 lai2 dao4 jia1-li3.
guest come to home inside
'Some guests came to my home.'

b. Qiang2-shang4 gua4-zhe yi1 fu2 hua4.
wall LOC-on hang ZHE one CLS painting
'There is a painting hanging on the wall.'

b' Yi1 fu2 hua4 gua4 zai4 qiang2-shang4.
one CLS painting hang at wall top
'There is a painting hanging on the wall.'

Verbs in this subcategory have undergone the "locative inversion" process (Bresnan and Kanerva 1989, Huang 1989); thus, they are derivationally related to the verbs with subcategorization of <SUBJ , OBL_{LOCT}>, as shown with the examples above. Here, their SUBJ must be place words, thus with [PLACE +]. Semantically, according to Chao (1968:530), these sentences often express the existence or the coming into existence of the entity expressed in the object. He included you3 as one of these verbs. Similarly, Li

and Thompson (1981:509-519) call these "presentative sentences" which they classify into two types: 1) existential verbs including you3 'exist' and other verbs they call "positional verbs" such as xie3 'write', gua4 'hang', for example sentence 3b, and 2), verbs of motion, for example 3a. However, we consider you3 as a unique verb denoting a general concept of existence, and we will have a lengthy discussion below. Also, you3 is different from the positional verbs and verbs of motion in that it does not have a derivational <SUBJ , OBL_{LOCT}> counterpart. Moreover, we believe the possessive use of you3 is part of its existential use and both can be accounted for in a single entry.

3.3.4.4 You3 'exist, have'

1. a. Wo3 you3 hen3 duo1 shu1.
I have very many book
'I have many books.'

b. Zai4 zhuo1-shang4 wo3 you3 hen3 duo1 shu1.
at desk top I have very many book
'On the desk I have many books.'

2. a. Zhuo1-shang4 you3 hen3 duo1 shu1.
desk- top exist very many books
'On the desk there are many books.'

b. Zai4 wu1zi-li3 zhuo1-shang4 you3 hen3 duo1 shu1.
at house inside desk- top exist very many books
'In the house on the desk there are many books.'

3. a. You3 hen3 duo1 shu1 zai4 zhuo1-shang4.
Exist very many book at desk- top
'There are many books on the desk.'

b. Zai4 zhuo1-shang4 you3 hen3 duo1 shu1.
at desk- top exist very many book
'On the desk there are many books.'

c. Zai4 wu1zi-li3 zai4 zhuo1-shang4 you3 hen3 duo1 shu1
at house inside at desk- top exist very many book
'In the house on the desk there are many books.'

In this section we will present a discussion on the verb you3 'have, exist' in Mandarin Chinese. We will argue that it belongs to this subcategory because it subcategorizes <SUBJ , OBJ>; however, its idiosyncracies have to be accounted for uniquely in its lexical entry. We will examine some previous accounts of this verb and then present evidence for our analysis. Generally, three usages of this verb can be recognized: 1) possessive, 2) existential and 3) (subjectless) locational. To facilitate discussion we first showed the above examples. Sentences in 1 above are of the first usage, possessive; sentences in 2 show the existential usage of you3 where its subject has to be a place word, and in 3 you3 may be preceded by a locational prepositional phrase and yet here it has no overt noun phrase subject.

First we will discuss you3 in these three verbal usages. We will find a common semantic interpretation of existence to explain all three usages and also discuss the implications on the possessive construction of [NP de NP]. Then we will examine another usage of you3 where it seems to be part of an adjectival element, such as you3 yi4si 'interesting'. Finally, we will give an explicit formulation of the lexical entry accounting for all the above usages of you3 in our vLFG formalism.

There were some transformational accounts of the verb you3 during the heyday of Transformational Grammar. For example, Li (1972) links verbs denoting 'be, exist, and have' (shi4, zai4, and you3) transformationally by posing a common deep structure. We can reject this analysis for several reasons. First of all, transformations are not allowed in LFG framework. Second, this analysis entails the grouping of shi4, zai4, and you3 into one single lexical entry and therefore the implication that the choice of shi4, zai4, or you3 is not lexical but purely phonological. This analysis contradicts the fact that these three verbs do manifest different syntactic behaviors and semantic content.

While we distinguish the preverbal locational NP, zhuo1-shang4, in 2a and the preverbal locational PP, zai4 zhuo1-shang4, in 3b, Huang (1966), along with others, such as Li and Thompson (1981:509-516), on the other hand, treats the two types of sentences as grammatically equivalent; in other words,

the preverbal PP in 3b is also taken to be the subject. This analysis, as Starosta (1985) protests, seems to be based on nothing more than the same English translations for these two types of sentences. Clearly, in 3a there is no preverbal subject since you3 occupies the sentence-initial position. Furthermore, note the fact that in Mandarin Chinese non-subcategorized locational prepositional phrases usually occur preverbally, exemplified in 1b, 2b and 3c. In order to capture this generalization, sentences 3b-c, just like 3a, would have to be considered as without overt subjects, and thus all preverbal locational PP's in 3b-c function adjunctively just like the preverbal PP's in 1b and 2b. In 2, however, as both Starosta (1985) and Ding et al (1979:78-79) have pointed out, the preverbal unmarked place words should be regarded as subjects, just like the NP subjects of 1a-b. Yet, both Chao (1968:530) and Li and Thompson (1981:511) treat them as topics, which is a very ambiguous position for they do not distinguish subjects from topics syntactically.

Starosta (1985), in his localistic Lexicase analysis, makes exactly the same three distinctions; namely, he recognizes the possessive and the non-possessive you3, and he further divides the non-possessive you3 into two subtypes: existential ($you3_1$) and locational ($you3_2$). Thus, this Lexicase analysis necessarily implies that in the lexicon there are three separate lexical entries: you3, $you3_1$, and $you3_2$. Since we do recognize that there are three different usages of the verb you, it would be convenient, and it is also perfectly acceptable within the LFG framework, to simply pose three different you3's. Nonetheless, we believe doing so fails to capture some important generalizations of the verb you3 as a single lexical unit.

As Li and Thompson (1981:516) have correctly pointed out, many languages of the world, including Mandarin Chinese, express both possession and existence with the same verb. To pose two (or three) different verbs for the possessive and existential usages is to render this phenomenon accidental and trivial, at least for Chinese. The fact that in many languages the same verb expresses possession and existence is, we believe, largely because the relation of possession between two entities is in fact a kind of relation of existence as well. That A possesses B is in effect equivalent to that B exists within the domain of A. To clarify this statement, first we have to point out that possession is NOT equivalent to ownership; rather, ownership is only a specific kind of, or an instance of, possession. That A owns B can be interpreted as B exists within the domain of A's ownership. In other words,

existence subsumes possession, which in turn subsumes ownership. Therefore, while it is possible to interpret sentences 1a-b in terms of ownership (i.e., "I own many books."), it is not always appropriate to do so. We will examine the following sentences of similar usage:

4. a. Wo3 you3 hen3 duo1 qin1qi1.
I have very many relative
'I have many relatives.'

b. Hua1 you3 hen3 duo1 zhong3lei4.
flower has very many variety
'Flowers have many varieties.'

c. Tai2wan1 you3 hao3 qi4hou4.
Taiwan have good climate
'Taiwan has good climate.'

d. Qi1yue4 you3 san1shi2yi1 tian1.
July have thirty-one day
'July has thirty-one days.'

All these sentences are better interpreted in terms of existence, namely that "A you3 B" is viewed as "B exists in the domain of A." This is precisely the kind of interpretation most suitable for sentences in 2, where the subjects are always place words that semantically function as the universe or the domain in which the object exists. Once we have shown that sentences of 1 and 2 are both syntactically and semantically equivalent in structure, we have eliminated the plausibility of posing two different you3's for 1 and 2. We thus also recognize that you3 requires two arguments: SUBJ and OBJ. Note that Cheng (1979) lists some you3 sentences like 4b and 4d above to demonstrate the "approximative" use of you3 in addition to the three uses we have listed above. However, it is clear that semantically our redefined concept of possessive and existential use of you3 also covers this "approximative" use, and syntactically sentences in 4 as well as in 1 and 2 are all equivalent in terms of their requirement of subject and object.

Next, we shall examine sentences in 3. We have already pointed out that sentences 2a and 3a are not syntactically equivalent because 2a has an overt subject while 3a does not. However, semantically, sentences of 3 are also best interpreted in terms of existence, namely that "you3 B" is viewed as "B exists, within an unspecified domain." The preverbal locational prepositional phrase is not a subcategorized element; rather it is an adjunctive element that denotes the general location of the existence. Compare the following sentences:

5. a. Zai4 wu1zi-li3 wo3 you3 hen3 duo1 shu1.
at house inside I have very many book
'In the house I have many books.'

b. Zai4 wu1zi-li3 zhuo1-shang4 you3 hen3 duo1 shu1.
at house inside desk- top exist very many books
'In the house on the desk there are many books.'

c. Zai4 wu1zi-li3 you3 hen3 duo3 shu1.
at house inside exist very many books
'In the house there are many books.'

The preverbal prepositional phrases in all three sentences are of the same adjunctive function. Now, to claim that the you3 in 3a-c and 5c is the same verb as in 2 and 3 is in effect to also claim that all of these sentences have the same syntactic structure, that is they all have a SUBJ and an OBJ.

you3:

```
[ FS      [ FORM 'you3'
            SUBJ OPT
          ]
  FI-V-4
]
```

Following the preliminary analysis of you3 of Her (1989a), we are therefore treating you3 as a single verb which subcategorizes two grammatical functions, SUBJ and OBJ, with a special default value on its SUBJ in the functional structure. Note that OPT is a placeholder and always

succeeds in unification. Thus, in this case, if the SUBJ of the you3 sentence is overt, then whatever the existing value of its SUBJ is, it will overwrite the default value OPT; only when its SUBJ is not overt and therefore with no feature SUBJ, will OPT remain to be the value of SUBJ.

By setting such a default value on you3's SUBJ we can account for the fact that it is possible for you3 not to have an overt subject, as seen in sentences of 3. In other words, the functional structures of sentences in 3 will not be considered incomplete, in LFG terminology. Note also that, although when used in existential sentences such as the ones in 2 the SUBJ must be a place word, we do not need such a constraint because in the possessive usage in 1 there is no such selectional restriction on the SUBJ. Finally, it may be worth mentioning that in all the you3 sentences thus far, negation is always marked by mei2, and in Mandarin Chinese you3 is the only verb that can be negated with mei2 regardless of what temporal aspect the verb denotes; since this phenomenon is not purely due to phonological factors (e.g., *mei2 you3shan4 'not friendly'), it certainly serves as another indication that you3 is a single lexical unit. This treatment of you3 is most economical and therefore presumably captures most generalizations of this verb.

Now we will discuss the implications of this unified treatment of you3 on the possessive construction of [NP1 de NP2]. While it is most common to interpret such a construction as "NP1 possesses NP2," we contend that, like the interpretation of [A you3 B], it is more appropriate to interpret the de possessive as "NP2 which exists in the domain of NP1." While it is absurd to interpret hua4xue2 de zhi1shi4 'the knowledge of chemistry' as "the knowledge possessed by chemistry," it is entirely appropriate to interpret it as "the knowledge which exists in the domain of chemistry." Thus, if we take all the sentences of [NP1 you3 NP2] in 4, each of them has a semantically equivalent counterpart of [NP1 de NP2] construction, as illustrated below.

4' a. Wo3 de qin1qi1.
I DE relative
'Relatives of mine.'

b. Hua1 de zhong3lei4.
flower DE variety
'Varieties of flowers.'

c. Tai2wan1 de qi4hou4.
Taiwan DE climate
'The climate of Taiwan.'

d. Qi1yue4 de san1shi2yi1 tian1.
July DE thirty-one day
'The thirty-one days of July.'

Next we shall proceed to examine another usage of the verb you3, one in which you3 seems to be part of a stative idiomatic expression. Sentences listed in 6 are some examples of this kind of [you3-NP] expressions. Although Cheng (1979) has hinted that this type of you3 constructions should be considered idiomatic, we are not aware of any previous detailed account of this usage.

6. a. Ta1 hen3 you3 yi4si.
he very have meaning
'He is very interesting.'

b. Ta1 hen3 you3 qian2.
he very have money
'He is very rich.'

c. Ta1 hen3 you3 jiao4yang3.
he very have education
'He is very educated.'

d. Ta1 hen3 you3 huo2li4.
he very have energy
'He is very energetic.'

The first noticeable peculiarity of sentences in 6 is that it is semantically odd for you3 which denotes the concept of 'existence' to be modified by hen3 'very' and therefore gradable. To look at this from a different perspective, however, this of course can serve as a strong indication to us that in 6 you3

has little to do with the concept of existence. More evidence comes from the fact that only when followed by certain abstract nouns can hen3 and you3 co-occur, as exemplified in 7, and that you3 cannot be modified by hen3 when followed by a concrete noun, as we demonstrate in 8.

7. a.*Ta1 hen3 you3 xue2shuo1.
he very have theory.
'He is full of theories.'

b.*Ta1 hen3 you3 jin1qian2.
he very have money
'He is very rich.'

c.*Ta1 hen3 you3 jiao4yu4.
he very have education
'He is very educated.'

d.*Ta1 hen3 you3 jian4kang1.
he very have health
'He is very healthy.'

8. a.*Ta1 hen3 you3 shu1.
he very have book
'He has many books.'

b.*Ta1 hen3 you3 xue2sheng1.
he very have student
'He has many students.'

c.*Ta1 hen3 you3 mao1.
he very have cats
'He has many cats.'

d.*Ta1 hen3 you3 gong1si1
he very have company
'He has many companies.'

It seems that no syntactic or semantic generalizations can be made as to what kind of abstract NP objects can be used for the you3 phrase to be gradable, since the ones in 7 are very similar to the ones in 6 in their semantic properties. Based upon this fact, clearly we can reject the analysis where sentences in 6 are to be treated just like other you3 sentences such as the ones in 1, 2 and 3. Sentences in 6 cannot be analyzed syntactically as regular verb phrase of [you3 NP]. We will examine the two other possible alternative analyses: one, to treat them as lexicalized elements generated through a lexical process; or two, to regard them as non-lexical, idiomatic expressions.

If [you3-NP] expressions in 6 are complex lexical items with incorporated objects, then the internal syntactic structure of the [you3-NP] expression is of no relevance to the syntactic structure of the sentence, and that they behave like basic, non-phrasal verb. This is how Bresnan (1982:51, 57) treats some fixed expressions, called complex verbs, in English, such as "go over," "look on," "make fun of," "catch sight of" and "look askance at." Some Chinese fixed expressions also clearly belong to this class, such as ba3feng1 'to be on the lookout' whose literal meaning of "to hold the wind" no longer exists. Each expression of complex verbs, for example the [you3-NP] expressions in 6, is listed as a separate entry in the lexicon. However, examples in 9 refute this lexical incorporation hypothesis.

9. a. Ta1 hen3 you3 hua4hua4 de tian1cai2.
he very have paint de talent
'He is very talented in painting.'

b. Ta1 hen3 you3 xie3 bao4gao4 de ji4qiao3.
he very have write report de skill
'He is very skillful in writing reports.'

c. Ta1 hen3 you3 xue2shu4 de jiao4yang3.
he very have academic de education
'He is very educated academically.'

d. Ta1 hen3 you3 qing1chun1 de huo2li4.
he very have youth de energy
'He is full of youthful energy.'

Since syntactically adjacency is assumed to be a necessary condition for incorporation or any other morpholexical process such as compounding (Bresnan 1982:51), we would have to stipulate that in 9a-d the entire [you3-NP] expression is a single lexical unit. The fact that in 9a-d all the NP's following you3 are syntactically complex, and in fact may be potentially infinitely so, indicates that it is impossible in this case for the entire [you3-NP] to be analyzed as a single lexical item, for that will make our lexicon infinitely large and thus unformulatable. Furthermore, the fact that all these stative [you3-NP] expressions, when negated, still have to be marked with mei2, but never with bu4 like other state verbs in Chinese, also indicates that they are not lexicalized state verbs. This is true with all the stative [you3-NP] expressions, including the ones whose NP can never be separated from you3 such as you3-qian2 'rich' and you3-yi4si 'interesting'. Therefore, we are left with our final option: to analyze the you3 phrases in 6 and 9 as idiomatic expressions whose internal syntactic structure is relevant to that of the sentence. Again, this is how Bresnan (1982:45-50) treats another type of "idiom chunks" in English, idiomatic expressions such as "keep tabs on," "pull one's leg" and "kick the bucket."

None of the facts we have discussed thus far contradicts this analysis. Now, compare sentences in 6 and 7. We see that although the nouns following you3 in 6 are very similar to those in 7 syntactically and semantically, the you3 phrases in 7 cannot be modified by hen3. For instance, hen3 you3 qian2 'very rich' is good, while *hen3 you3 jin1qian2 is not, and while hen3 you3 jiao4yang3 'very educated' is good, *hen3 you3_jiao4yu4 is not. This observation clearly points to an idiomatic solution. Further facts indicate the same solution.

10. a. Ta1 hen3 you3 qian2.
he very have money.
'He is very rich.'

b.*Ta1 hen3 you3 mai3 shu1 de qian2.
he very have buy book de money.
'He has lots of money for buying books.'

11. a. Ta1 hen3 you3 kan4fa3.
he very have opinion
'He is very insightful.'

b.*Ta1 hen3 you3 bu4tong2 de kan4fa3.
he very have different de opinion
'He is full of different opinions.'

Although, as we have shown with examples in 9, the NP following you3 may be syntactically complex, whether this is allowed seems to be arbitrarily restricted, depending on each you3 expression of this sort. Thus, within the specification of the idiom you3 qian2 'rich', it has to be specified that qian2 cannot take any modifier, while the idiom you3 ji4qiao3 'skillful' has no such restriction. Only an idiomatic analysis is compatible with this finding. A similar fact can be observed in English idioms. For "kick the bucket" to have its idiomatic reading no modifier is allowed on the noun "bucket." For the same reason "he is pulling my right leg" has no idiomatic reading. Yet idiomatic expressions like "take a look" does allow certain modifiers on the noun, e.g., "take a long look" and "take a hard look." Similarly all sentences in 9 are good.

Another piece of evidence in support of the idiomatic analysis comes from topicalized and preposed positions.

12. a. Wo3 you3 qian2.
I have money
'Money, I have.'
OR 'I am rich.'

b. Qian2, wo3 you3.
money I have
'Money, I have.'

c.*Qian2, wo3 hen3 you3.
money I very have
'I am very rich.'

d.*wo3 qian2 hen3 you3.
I money very have
'I am very rich.'

13. a. Wo3 you3 ji4qiao3.
I have skill
'Skills, I have.'
OR 'I am skillful.'

b. Ji4qiao3, wo3 you3.
skill I have
'Skills, I have.'

c.*Ji4qiao3, wo3 hen3 you3.
skill I very have
'I am very skillful.'

d.*Wo3 ji4qiao3 hen3 you3.
I skill very have
'I am very skillful.'

Notice that while it is possible for the object of you3 in a non-idiomatic usage to be topicalized or preposed (thus only the literal reading for 12b and 13b), the idiomatic usage does not allow that (thus 12c-d and 13c-d are not acceptable). Again a similarity can be observed with certain English idiomatic expressions. The idiomatic analysis accounts for the fact that 14b is ungrammatical and 14d has no idiomatic reading.

14. a. Money, I take.
b.*A look, I take.
c. The basket, he kicked.
d. The bucket, he kicked. (only literal reading)

We therefore conclude that for all the [you3-NP] expressions allowed to be modified by hen3, a non-lexical idiomatic treatment is most compatible with our findings on their syntactically idiosyncratic behaviors. The two other alternative analyses are ruled out because while a lexical treatment is far too rigid and would create an infinitely large lexicon impossible to formulate, a general syntactic analysis is excessively non-restrictive and thus constitutes a serious problem of over-generation.

Finally, incorporating the idiomatic expressions, we will reformulate the lexical entry of you3 in the vLFG formalism. First, however let's look at Bresnan's (1982:46-7) formulation of an idiomatic expression such as "keep tabs on" in its (partial) lexical entry.

15. keep V
(↑ PRED) = 'keep-tabs-on <(↑SUBJ)(↑ON OBJ)
(↑OBJ FORM) =$_c$ TABS>'

The lexical form of this entry shows that "keep-tabs-on" requires three grammatical functions: SUBJ, ON OBJ, and OBJ. However, since semantically "keep-tabs-on" is very similar to individual verbs like "watch" or "investigate," thematically it should only require two arguments, agent and theme. Likewise, in the case of [you3-NP] idioms, for example you3-qian2 'rich', semantically they function like state verbs, such as fu4yu4 'rich', and thus require only one thematic role, theme. However, syntactically they require two grammatical functions, SUBJ and OBJ.

We should, at this point, explain again that in LFG theory grammatical relations are lexically encoded by mapping the predicate argument structure (PAS) to grammatical functions; however, although the Function-Argument Biuniqueness Principle ensures that a unique grammatical function is mapped with each thematic role and a unique thematic role to each function associated with the PAS, the relation between thematic roles and grammatical functions may not always be one-to-one because of the possibility of non-thematic grammatical functions, e.g. raised subjects. The treatment of idioms like "keep-tabs-on" presents another case of non-thematic grammatical functions. "Keep-tabs-on" is thematically like verbs such as "watch" and "investigate"

that require two arguments, and the assignment of grammatical functions is the following (Bresnan 1982:46):

16. keep-tabs-on (agent, theme)
↓ ↓
<SUBJ ON-OBJ> OBJ

Clearly the grammatical function of OBJ, whose FORM has to be TABS, is non-thematically assigned. For the same reason the OBJ function in the lexical form of you3-qian2 is also non-thematical. Yet, the notation in 16 that Bresnan has adopted does not show this characteristic. We should follow the notation used in Chapter 5 of Bresnan (1982:289) where the non-thematical grammatical function of "seem-to" is placed outside of the angled brackets.

17. John seems sick to Mary.
seem-to
predicate argument structure: seem (1,2)
grammatical function assignment: <(XCOMP) (OBL_{go})>
lexical form: 'seem-to <(XCOMP)(OBL_{go})> (SUBJ)'

We will now illustrate the relationship of the thematic argument structure, required grammatical functions, and the lexical form of idioms like you3-qian2 'rich'.

you3-qian2
predicate argument structure: you3-qian2 (theme)
grammatical function assignment: <(SUBJ)>
lexical form: 'you3-qian2 <(SUBJ)> (OBJ)'

The entire lexical entry of you3 should account for its regular use as a non-stative verb, thus not allowed to be modified by degree adverbs, and also the stative use of idiomatic expressions which are gradable by the modification of adverbs such as hen3. It should also account for the fact that the object of certain idioms can be modified and others cannot. The following is our formulation of the lexical entry. No doubt, the lexical entry of you3 will be rather lengthy due to the large number of [you3-NP] idioms; therefore, we

will simply show some examples of the possible idioms and note that the following is still an incomplete, partial entry of you3.

```
FI-V-4-ID:                                  `FI entry for VO idioms
      [ CAT V
        FS [ PRED <SUBJ , OBJ @>            `non-thematical OBJ
             OBJ [ BACKGROUND -             `OBJ cannot be preposed
                   FORM ANY                 `must have overt value
                 ]      ]                   `to overwrite ANY
        FI-V-STATE                          `stative, gradable
      ]

you3:
{       [ FS [ FORM 'you3'                  `regular use
               SUBJ OPT                     `default value for SUBJ
             ]
        FI-V-ACTION                         `non-gradable
        FI-V-4                              `subcategory information
       ]
       [ FS [ FORM 'you3-ji4qiao3'          `idiomatic use
              OBJ   [ FORM 'ji4qiao3' ]     `skillful
            ]
        FI-V-4-ID
       ]
       [ FS [ FORM 'you3-tian1cai2'          `talented
              OBJ   [ FORM 'tian1cai2' ]
            ]
        FI-V-4-ID
       ]
       .
       .
       [ FS [ FORM 'you3-yi4si'              `interesting
              OBJ   [ FORM 'yi4si'
                      ADJ NONE              `cannot take
                    ]                       `any modifier
            ]
```

```
        FI-V-4-ID
      ]
      [ FS [ FORM 'you3-qian2'              `wealthy
             OBJ   [ FORM 'qian2'
                     ADJ NONE
                   ]
           ]
        FI-V-4-ID
      ]
    .
    .
    .
}
```

Two special features might need to be explained again. When a grammatical function is followed by an @ sign, such grammatical function is non-thematical. NONE is a special value, which always fails when unifying with any other value. We impose the NONE value on the attribute ADJ in the OBJ of idioms like you3-qian2, where the OBJ cannot take a modifier and still maintain the idiomatic reading. As we have stated earlier, because TOPIC is usually a placeholder of old, or background, information, we designate a attribute-value pair of [BACKGROUND +] in TOPIC. Since none of the OBJ of these expressions can be preposed, we impose [BACKGROUND -] in the OBJ so that its f-structure will fail if topicalized.

Finally, we note that several of these [you3-NP] expressions as stative verbs may also function as genuine adjectives (not verbs) in that they appear in a non-relative construction modifying a noun attributively, such as you3qian2 ren2 'rich people', you3qing2 ren2 'people in love' and you3xin1 ren2 'people with premeditated motives'. Given the highly restrictive nature of adjectives in Chinese, we will have to treat expressions like you3qian2, you3qing2 and you3xin1 as separate lexical entries as adjectives as well.
Take you3qian2 for example.

18. You3qian2 ren2 xi3huan1 kai1 pao3 che1.
    rich people like drive race car
    'Rich people like to drive race cars.'

you3qian2:
[CAT A
FS [FORM 'you3qian2']
]

To summarize, we recognize the idiomatic use of you3 and the regular non-idiomatic use. We have rejected the notion that the different usages of the non-idiomatic you3 are to be attributed to more than one lexical entry of you3; also, for the stative use of certain [you3-NP] expressions, we have established arguments for an idiomatic analysis where the internal syntactic structure of the idiom is relevant to that of the sentence. And, finally we formulated within the vLFG formalism a unified lexical entry accounting for all the usages of this verb discussed in this section.

3.3.4.5 Verbs Requiring TOPIC

Examples: na2shou3 'be good at', zuo4zhu3 'take charge', dao3luan4 'meddle', guo4mu4 'browse'

5. a. Zhei4 jian4 shi4, ni3 zuo4zhu3.
this CLS matter you make-master
'You'll take charge of this matter.'

b.*Ni3 zuo4zhu3.
you make-master
'You'll take charge.'

c.*Ni3 zuo4zhu3 zhei4 jian4 shi4.
you make-master this CLS matter
'You'll take charge of this matter.'

In section 2.9, we discussed in detail that there are some two dozen verbs in Chinese that seem to require their OBJ to be missing and their TOPIC to be present at least in the discourse if not within the same sentence. We will now give them a revised formulation incorporating the feature inheritance entries.

```
FI-V-4-5:
      [ FS [ OBJ [ BACKGROUND +     `OBJ cannot be overt
                   FORM ANY         `requires overt TOPIC
                 ]                  `to unify with OBJ to
        ]                           `overwrite ANY
      ]

zuo4zhu3:                           `take charge of
      [ FS [ FORM   'zuo4zhu3' ]
        FI-V-4
        FI-V-4-5
      ]

na2shou3:                           `be good at
      [ FS [ FORM   'na2shou3' ]
        FI-V-4
        FI-V-4-5
      ]
```

3.3.5 <SUBJ , OBL_{THME}>

Examples: da3 'hit', sha1 'kill', ti1 'kick', bang3 'tie', xie3 'write', hui3 'destroy', kan4 'read', mai4 'sell', liang4gan1 'line-dry', zhuang1man3 'fill', da3po4 'break', ti1kai1 'kick .. open' or 'kick .. away', da3duan4 'hit-break'.

1. a. Wo3 ba3 shu1 hui3 le.
I BA book destroy LE
'I destroyed the book.'

b. Wo3 ba3 men1 ti1kai1 le.
I BA door kick-open LE
'I kicked the door open.'

Ba3 is probably the single element that has been discussed most frequently in Chinese linguistics. Its actual function seems obscure and has been a

long-standing topic of debate. The first essential issue we should be concerned with here is whether [ba3 NP] should be assigned to the grammatical function of direct object or an oblique function. Although as observed by Her (1985-6) the greatest majority of Chinese linguists consider [ba3 NP] a preposed object and thus ba3 sentences an important OV sentence type in Mandarin, we will point out several inadequacies in this conventional analysis and argue that it is more appropriate to treat it as an oblique function. Thus, more specifically in LFG terminology, we will attempt to establish that [ba3 NP] is of the PP category that encodes the grammatical oblique function of OBL_{THME}, not OBJ.

3.3.5.1 Syntactic Category of ba3

First of all it is important to note that historically ba3 was originally a verb, meaning 'take', 'hold', etc., but its verbal function is nearly lost in Modern Mandarin, except in some lexicalized expressions (Her 1989d). The question that needs to be resolved first is whether ba3 should be considered a preposition or a so-called "object marker" similar to the Japanese [-o] that marks the accusative case. Obviously, the syntactic category of ba3 has direct implications on whether [ba3 NP] should be assigned the object function or an oblique function.

Wang (1947) initially observed the "disposal" meaning of ba3. Since then, that ba3, unlike a pure grammatical case marker, has semantic content has been recognized by some Chinese linguists (e.g., Li 1974 and Li and Thompson 1981:468), who analyze ba3 as having the meaning of "dispose" or "process." Thus, ba3 resembles other prepositions in Mandarin such as gen1 'with', dao4 'to', and gei3 'to', in that they all have semantic content directly derived from their verbal meaning. Chu (1984), recognizing that ba3 is a lexical item clearly with semantic content and that its presence or absence not only makes a difference in terms of the meaning but also the grammaticality of a sentence, first correctly rejects the conventional analysis where ba3 is an "object marker." Consequently, he refuses to accept that ba3 sentences are of OV order; instead, he analyzes ba3 as a preposition and the ba3 construction as having a structure of [S [P O] V]. It is important to note that here the O is considered the object of the preposition ba3, not the direct object of the verb. We fully agree with Chu's conclusion and will support it by providing further

evidence below. We thus also have to object to the rather ambiguous position in which ba3 is recognized as a preposition and yet at the mean time also termed as an "object marker" (e.g., CKIP 1989:43). The same argument also applies to refute the transformational account (e.g. Teng 1977:35-36) where ba3, with semantic content, is transformationally introduced when the postverbal object is preposed.

3.3.5.2 The Double Direct Object Fallacy

In one type of ba3 sentences, it is possible for the preverbal [ba3 NP] to co-occur with a postverbal unmarked NP, as exemplified in 2a-e below. These indisputable data show that the object of ba3 cannot be syntactically the direct object of the main verb within a framework where only one direct object is allowed per verb (Her 1985-6), unless, of course, one is ready to accept the validity of a "double direct object" construction in a linguistic theory, and thus in a Chinese grammar.

2. a. Ta1 ba3 fang2zi qi1 le xin1 qi1.
He BA house paint LE new paint
'He painted the old house new.'

b. Ta1 ba3 hai2zi chuan1hao3 le yi1fu2.
She BA child dress-finish LE clothes
'She finished dressing the child.'

c. Gong1ren2 ba3 da4men2 jia1 le suo3.
Worker BA gate add LE lock
'The worker locked the gate.'

d. Mao1 ba3 yu2 yao3diao4 le tou2.
Cat BA fish bite-off LE head
'The cat bit off the head of the fish.'

e. Ta1 ba3 wo3 qiang3 le qian2.
he BA I rob LE money
'He robbed me of my money.'

Li and Thompson (1981:470-471), considering ba3 as an "object maker" and thus the NP following ba3 as the direct object, treat the NP following the verb in the above examples also as the direct object. They schematize the syntactic pattern of such sentences like the following:

[subject ba3 $object_1$ verb $object_2$]

Since Li and Thompson adopt a rather unconstrained informal "functional" grammar, such claims bear little theoretical consequence. Similarly, in a very loosely-defined Case Grammar, taking ba3 to be an object marker, CKIP (1989:43,67) also seems to assume the double object solution. In the formal theory of LFG, subcategorizable grammatical functions, such as SUBJ and OBJ, may not violate the Consistency Condition. They must have no more than one unique value; only adjunctive functions such as ADJ may have more than one value in a conglomerated list. Thus, to allow a double direct object construction in LFG violates this universal generalization. On the other hand, we can avoid this violation by simply accepting the analysis where [ba3 NP] is assigned to a subtype of the oblique function, OBL_{THME}. Thus, we have to consider the double direct object approach invalid in the LFG theory. (By the same token, the so-called "double subject" analysis of constructions like wo3 tou2 teng2 'I have a headache' is also unacceptable in LFG.) Our analysis, which utilizes existing categories allowable in a theory, without resorting to the creation of a new category such as double direct object, is more revealing and thus should be preferred.

A similar argument can be made from a typological point of view. Andrews (1985:120-121), in a survey of the major functions of noun phrases in the world's languages, although he accepts Li and Thompson's analysis in stating that Mandarin Chinese might be an example of a language with two direct object grammatical relations instead of just one, offers the following caution: "one would want to investigate further before accepting this conclusion," and vaguely concludes that "In any event, Mandarin is an extreme example of a language with multiple forms of expression for o" (underline mine and o is short for direct object). Extremities and exceptions should always be examined with extra caution and scrutiny in the study of language universals. Although Andrews clearly recognizes that the NP marked by ba3 and the

unmarked postverbal NP have different grammatical functions, which is "supported by the fact that they differ in NP-marking and position," he unfortunately does not follow through. The fact that it is possible to have both these NPs in one clause is an even stronger indication that they represent two different grammatical relations. However, after first accepting Li and Thompson's premise that ba3 marks direct object in Mandarin, he really has little choice but to accept their analysis of double direct object while recognizing the extremity of such a claim. Our analysis indeed recognizes [ba3 NP] and postverbal unmarked NP as assigned two different grammatical functions: oblique function and the direct object function. Therefore, we maintain, without exception, the universal that in all languages a verb may subcategorize at most one grammatical function of direct object.

3.3.5.3 Semantic Restrictions of [ba3 NP]

The universal characteristic of OBJ as a semantically unrestricted grammatical function and OBL-Θ as a semantically restricted one in the LFG theory provides another test for us to decide which grammatical function [ba3 NP] should be assigned to. Recall that semantically unrestricted functions, such as SUBJ and OBJ, may be linked to any thematic role in the thematic structure of a verb. Semantically restricted functions are more intrinsically related to their semantic content and thus may only be linked to thematic roles that are semantically compatible. In the following examples in 3, all post-verbal noun phrases are no doubt direct objects. Note that these OBJs are associated with several different thematic roles: theme, goal, location, and instrument.

3. a. Wo3 da3 le ta1.
I hit LE he
'I hit him.' `theme

b. Wo3 yong4 le yan2.
I use LE salt
'I consumed the salt.' `theme,
OR 'I utilized the salt.' `instrument

c. Wo3 jing4 le ta1. `goal
I toast LE he
'I toasted him.'

d. Wo3 qu4 le Tai2bei3. `location
I go LE Taipei
'I went to Taipei.'

However, only 3a-b, whose OBJs are linked with the thematic role theme, have grammatically acceptable ba3 counterparts, as shown in 4 below. Notice also that while 3b is ambiguous in that yong4 could be interpreted as "utilize" which requires an instrument role for its object or as "consume" which requires a theme role, its ba3 counterpart 4b can only have the reading where [ba3 yan2] is linked with the theme role. Sentences in 4 thus show that the grammatical function of [ba3 NP] prepositional phrases can only be linked with the theme role.

4. a. Wo3 ba3 ta1 da3 le.
I BA he hit LE
'I hit him.' `theme

b. Wo3 ba3 yan2 yong4 le.
I BA salt use LE
'I consumed the salt.' `theme
'*I utilized the salt.' `*instrument

c.*Wo3 ba3 ta1 jing4 le.
I BA he toast LE
'I toasted him.' `*goal

d.*Wo3 ba3 Tai2bei3 qu4 le.
I BA Taipei go LE
'I went to Taipei.' `*location

Actually, given the semantic content of "dispose" or "process" of ba3 as a preposition, it is rather natural for the [ba3 NP] prepositional phrase to be

associated with only the thematic role of theme or patient. Another type of evidence for the semantic restrictiveness of [ba3 NP] comes from certain transitive verbs, such as jing4 'toast' in 4c above, which do not normally have acceptable ba3 counterparts, but when they are modified by an "extended" expression, they may indeed co-occur with [ba3 NP]. See the following example.

5. a. Wo3 ba3 ta1 jing4 de ta1 bu4 neng2 zai4 he1 le.
I BA he toast DE he not can again drink LE
'I toasted him so much that he could not drink any more.'

In this sentence, the added expression indicating the result of the action greatly exaggerates the degree of my toasting him. As Li and Thompson (1981:469) put it, "It is as if one cannot help thinking" that he was affected in some way when he was toasted so much by me that he could not drink any more. Therefore, it seems that ba3 forces its NP to be interpreted as having the theme role in such constructions, although normally, as shown in 3c, jing4 requires a goal role in its predicate.

To sum up, all the evidence provided thus far clearly indicates that ba3 should be classified as a preposition and that [ba3 NP] should be assigned to a subtype of the semantically restricted oblique function, not to the semantically unrestricted direct object. Since the semantic content of ba3 and the semantic restriction on [ba3 NP] seem to require an interpretation of theme role of the NP, we assign OBL_{THME} to be its grammatical function. Incidentally, it is a well-known fact that [ba3 NP], being always preverbal, has a strong tendency for its noun to have a definite reading; this fact is accounted for by our functionally annotated phrase structure rule that creates VP, where OBL_{THME} gets a default value of [DEFINITE +] unless it is otherwise marked.

All verbs in this subcategory that subcategorize <SUBJ , OBL_{THME}> also belong to the previous subcategory of 3.3.4.1 that subcategorize <SUBJ , OBJ>. However, as we have mentioned before, the postverbal object, when unmarked, has a strong tendency to be indefinite. For example,

6. a. Wo3 ba3 ji1 sha1 le.
I BA chicken kill LE
'I killed the chicken.'

a' Wo3 sha1 le ji1.
I kill LE chicken
'I killed a chicken.'

b. Wo3 ba3 men2 ti1huai4 le.
I BA door kick-bad LE
'I kicked and broke the door.'

b' Wo3 ti1huai4 le men2.
I kick-bad LE door
'I kicked and broke a door.'

FI-V-5:
[CAT V
FS [PRED <SUBJ , OBL_{THME}>]
($\uparrow$ OBL_{THME} PCASE) $=_c$ THME
]

sha1:
{ [FS [FORM 'sha1']
FI-V-4
]
[FS [FORM 'sha1']
FI-V-5
]
}

ba3:
[CAT P
FS [PFORM 'ba3'
PCASE THME
]
]

Most resultative compound verbs, such as 6b above, belong to this subcategory and thus all have non-ba3 derivational counterparts. However, an interesting observation is that when a resultative compound takes the form with a potential infix, affirmative de or negative bu2, it cannot subcategorize a [ba3 NP] constituent. The following sentences illustrate this observation. Again, this fact is accounted for in the individual lexical entries.

7. a. Ming2tian1 wo3 hui4 ba3 men2 da3po4.
tomorrow I will BA door break
'Tomorrow I will break that door.'

a' Ming2tian1 wo3 hui4 da3depo4 na4 shan4 men2.
tomorrow I will break that CLS door
'Tomorrow I will break that door.'

a"*Ming2tian1 wo3 hui4 ba3 men2 da3depo4.
tomorrow I will BA door break
'Tomorrow I will break that door.'

b. Wo3 bu4 neng2 ba3 men2 da3po4.
I not can BA door break
'I cannot break that door.'

b' Wo3 da34bu2po4 na4 shan4 men2.
I break-not-open that CLS door
'I cannot break that door.'

b"*Wo3 ba3 men2 da3bu2po4.
I BA door break-not-open
'I cannot break that door.'

da3po4:
{ [FS [FORM 'da3po4']
FI-V-4
]

```
          [ FS [ FORM 'da3po4' ]
            FI-V-5
          ]
}

da3depo4:
        [ FS [ FORM 'da3depo4' ]
          FI-V-4
        ]

da3bu2po4:
        [ FS [ FORM 'da3bu2po4' ]
          FI-V-4
        ]
```

3.3.6 <SUBJ , OBL_{THME} , OBJ>

Examples: bo1 'peel', shang4 'lock', shua1shang4 'paint', zhuang1man3 'fill', zhuang1hao3 'fix', chi1 'eat', da3puo4 'break', fa1she4 'shoot';; gei3 'give', gong1ji3 'provide', song4 'give', jie4 'lend', jiao1 'teach', zu1 'rent', jie4 'lend', ti2gong1 'provide', gao4su4 'tell', dang1 'consider', kan4cheng2 'mistake...as', jiao4zuo4 'call'

1. a. Wo3 ba3 ta1men pian4 le liang3 ge.
   I BA they cheat LE two CLS
   'I cheated two of them.'

   b. Ta1 ba3 men2 shang4 le liang3 dao4 suo3.
   he BA door lock LE two CLS lock
   'He locked the door with two locks.'

   c. Ta1 ba3 hu2li2 bo1 le pi2.
   he BA fox peel LE skin
   'He peeled the skin off the fox.'

d. Ta1 ba3 hu2li2 bo1 le san1 zhi1.
he BA fox peel LE three CLS
'He peeled three of the foxes.'

FI-V-6:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ>]
(↑ OBL_{THME} PCASE) $=_c$ THME
]

As we have argued in the previous section, we reject the double object analysis for sentences in 1 above. We treat the [ba3 NP] PP as having the OBL_{THME} oblique function. Another possible preposition that may assign this OBL_{THME} function is jiang1, which is used mostly in a more formal context such as the written text. The OBJ here can either be a full NP with a lexical head, thus with the head feature FORM, such as 1b-d above, or a headless NP construction without the head feature FORM, such as 1a. Note also that semantically when the OBJ is linked with a patient role, there is an interesting relation between the entities in OBL_{THME} and patient OBJ: the entity in OBJ is a closely related, often inalienable, part of the entity in OBL_{THME}, either before or after the action denoted by the verb. Thus, it is always the patient OBJ that directly receives the action while the theme OBL_{THME} is indirectly affected by it. However, notice that the relationship has to be that of whole-part or possession but cannot be that of class-member. This semantic restriction applies without exception to all ba3 sentences with a postverbal patient OBJ, which directly receives the action or undergoes the process denoted by the verb. However, this semantic restriction does not apply to verbs that are derivationally related to ditransitive verbs, such as gei3, whose OBJ is not linked to a patient but more likely to a thematic role of beneficiary, maleficiary, or goal. Our distinction of theme and patient in ba3 sentences is also supported by the studies of Chinese resultative verb compounds and verb-copying by Chang (1989 and 1990a).

This semantic restriction may provide partial explanation for the interesting observation given by Her (1985-6) that a postverbal patient OBJ and a preverbal [ba3 NP] OBL_{THME} are in complementary distribution.

1. e. Mao1 chi1 le yu2.
Cat eat LE fish
'The cat ate the fish.'

f. Mao1 ba3 yu2 chi1 le.
Cat BA fish eat LE
'The cat ate the fish.'

g.*Mao1 ba3 $yu2_i$ chi1 le $yu2_i$.
Cat ba3 fish eat LE fish
'The cat ate the fish.'

No doubt the sentence 1g has a well-formed c-structure and f-structure since chi1 does belong to this subcategory and thus subcategorizes <SUBJ , OBL_{THME} , OBJ>. As a matter of fact, it seems that all formal features are satisfied in this sentence. Therefore, we know for sure that this sentence is ill-formed not because there is any violation in c- or f-structure. As we have seen in sentences 1a-d, it is entirely acceptable for the verb with preverbal ba3 to have a postverbal patient object, provided that this object and the object of ba3 are of a part-whole or possession relationship. In other words, the presence of an overt direct object with the identical form with the noun in OBL_{THME} is not the cause of the ill-formedness. The culprit is rather the difficulty or impossibility for a part-whole relationship to exist between the direct object and the NP in the ba3 phrase. We could therefore assume that sentence 1g is syntactically well-formed, but it is semantically or pragmatically ill-formed due to the violation of a semantic constraint.

The principles of anaphoric binding may provide another part of the solution to the unacceptability of sentences like 1g. Since full-NPs (or R-expressions in GB terminology) must be free, in 1g, it violates this principle for the NP in OBJ to be bound by the NP in OBL_{THME}. Thus, the only possible grammatical reading would force the two NPs, yu2 'fish', not to be co-referential. And when such reading is possible within the semantic constraint of whole-part or possession relationship either before or after the action, sentences like 1g will be acceptable, as predicted by our account. Thus, all the following sentences are acceptable, but only if OBJ and OBL_{THME} are not co-referential and a sensible reading is still possible. Yet, the reading

where the OBL and OBL_{THME} are co-referential is always anomalous or nonsensical and also violates the binding principle of full NPs.

1. h. Ta1 ba3 $tu3_i$ sha3shang4 le $tu3_j/*_i$.
He BA dirt spread-on LE dirt
'He spread dirt on the dirt.'

i. Ta1 ba3 $yi1fu2_i$ gai4shang4 le $yi1fu2_j/*_i$.
He BA clothes cover-on LE clothes
'He covered the clothes with clothes.'

j. Ta1 ba3 $pi2_i$ buo1 le $pi2_j/*_i$.
He BA skin peel LE skin
'He peeled the skin off the skin.'

k. Ta1 ba3 $suo3_i$ jia1 le $suo3_j/*_i$.
He BA lock add LE lock
'He added a lock on the lock.'

However, Her (1989a) also assumes that for sentences like 1a where the object is a headless NP the head noun of the OBL_{THME} must be unifiable with the entire OBJ. Thus, the unacceptability of the following sentence is said to be caused by the fact that the tou2 as a classifier is in conflict or not in agreement with the grammatical classifier of xiao3ji1 'chick', which is zhi1.

1. l.*Hu2li2 ba3 xiao3ji1 chi1-le liang3 tou2.
fox BA chick eat two CLS
'The fox ate two of the chicks.'

Such an analysis may be incorrect; instead, we would suggest that the whole-part relationship between the OBL_{THME} and the OBJ is a semantic and discoursal one only and that there is definitely not a control relation where unification takes place between the two functions. This is evident from the following example of a simple discourse.

1. m. Q: Ni3 ba3 Lao3wang3 chi1 le ji3 zhi1 ji1?
you BA Old Wang eat LE how-many CLS chicken
'How many of Old Wang's chicken did you eat?'

n. A: Wo3 ba3 Lao3wang3 chi1 le liang3 zhi1.
I BA Old Wang eat LE two CLS
'I ate two of Old Wang's (chickens).'

In the case of 1n where zhi1 cannot be the grammatical classifier of Lao3wang3 and thus the unification between OBL_{THME}, Lao3wang3, and the OBJ, liang3 zhi1, would definitely fail. The fact that 1n is an entirely acceptable sentence proves that there is no control relation between OBL_{THME} and OBJ under any circumstance. Thus, our account predicts that 1l is in fact syntactically well-formed and attributes its unacceptability to its semantic anomaly. Therefore, if we can conjure up some discourse context in which such a sentence makes some sense, it will be judged as well-formed. Indeed we can. Imagine someone telling some children a fairy-tale where some smart little fairy chicks have a farm with some cows and there is also this bad giant fox trying to steal and devour some of the cows. The fox finally succeeds. One of our young listeners might ask the following question to which 1l is a syntactically grammatical as well as semantically acceptable response.

1. o. Q: Hu2li2 ba3 xiao3ji1 chi1-le ji3 tou2 niu2?
fox BA chick eat LE how-many CLS cow
'How many of the chicks' cows did the fox eat?'

l. A: Hu2li2 ba3 xiao3ji1 chi1-le liang3 tou2.
fox BA chick eat LE two CLS
'The fox ate two of the chicks' (cows).'

This observation confirms our previous statements that 1) the relationship between OBL_{THME} and a patient OBJ in a single clause is that of whole-part or possession, but cannot be that of class-member, 2) this semantic restriction applies without exception to all ba3 sentences with a postverbal patient OBJ, and 3) there is no control relation between them. Our analysis thus accounts

for the anomaly of the following sentence nicely, while an agreement account does not.

1. p.?Ta1 ba3 yi1 zhi1 ji1 chi1-le wu3 zhi1.
he BA one CLS chicken eat LE five CLS
'?He ate five of the chicken.'

All the ditransitive verbs, such as gei3 'give', song4 'give', gong1ji3 'supply', that subcategorize <SUBJ , OBJ , OBJ2>, also have counterparts in this subcategory. However, for these verbs, their OBJ is linked to a beneficiary, maleficiary or goal role and therefore there does not exist a part-whole or possession relation between the OBL_{THME} and the OBJ.

1. q. Wo3 ba3 shu1 gei3 le ta1.
I BA book give LE he
'I gave him the books.'

r. Ta1 ba3 qing2bao4 gong1ji3 jun1fang1.
he BA information supply military
'He supplies information to the military.'

Incidentally, though we will not get into the analysis of it, we note that verbs in the following sentences do not belong to this subcategory; rather, they subcategorize <SUBJ , OBL_{THME}> only and the postverbal NPs yi1 quan2 and liang2 xia4 are not OBJ. They function entirely like the frequency phrases such as yi1 ci4 'once' and belong to the non-subcategorizable ADJ, quantifying the extent or the frequency of the action.

1. s. Ta1 ba3 ni3 da3-le yi1 quan2.
he BA you hit ASP one fist
'He hit you with his fist once.

t. Ta1 ba3 ni3 da3-le yi1 xia4.
he BA you hit ASP one time
'He hit you once.'

3.3.7 <SUBJ , OBL_{GOAL} , OBJ>

Examples: jie3shi4 'explain', bao4gao4 'report', shuo1ming2 'illustrate', du3, da3du3 'bet', biao3shi4, biao3da2 'express', shuo1 'say', cheng2ching1, biao3ming2 'clarify', mai3 'buy', mai4 'sell', tui1xiao1 'promote', kua1zhang1 'exaggerate', zu1 'rent', jie4 'borrow', xun2wen4 'ask'

1. a. Ta1 gen1/tong2/dui4/xiang4/he2 ni3 jie3shi4 Yi4jing1.
   He with you explain Yi Ching
   'He explains Yi Ching to you.'

   b. Ta1 gen1/tong2/*dui4/xiang4/*he2 ni3 zu1 fang2zi
   He with you rent house
   'He rents a house from you.'

   c. Ta1 gen1/tong2/dui4/xiang4/*he2 ni3 tui1xiao1 fang2zi
   He with you promote house
   'He promotes the house to you.'

FI-V-7:

[CAT V
  FS [PRED <SUBJ , OBL_{GOAL} , OBJ>]
  (↑ OBL_{GOAL} PCASE) $=_c$ GOAL
]

jie3shi4:

[FS [FORM 'jie3shi4']
  FI-V-7
]

zu1:

[FS [FORM 'zu1']
(↑ OBL_{GOAL} PFORM) $=_c$ { 'gen1' 'xiang4' 'tong2' }
  FI-V-7
]

tui1xiao1:
[FS [FORM 'tui1xiao1']
~(↑ OBL_{GOAL} PFORM) =$_c$ 'he2' `PFORM cannot be 'he'
FI-V-7 `~ negates the schemata
]

Again, in the default case all prepositions that have [PCASE GOAL] may assign the OBL_{GOAL} function; however, individual verbs within this subcategory that subcategorizes <SUBJ, OBL_{GOAL} , OBJ> may require different prepositions rather arbitrarily as shown in the above sentences 1a-c, and such idiosyncratic information must be specified in the individual verb entry, as exemplified in the above entries.

3.3.8 <SUBJ , OBJ , OBL_{LOCT}>

Examples: xie3 'write', gua4 'hang', diu1 'throw', ban1 'move', ji4 'mail', pai4 'sent', tui1 'push', tiao1 'carry', dai4 'carry', mai4 'sell', kai1 'drive', fang4 'put', ti1 'kick'

1. a. Ta1 xie3 le yi1 ge zi4 zai4/*dao4 zhi3-shang4.
he write LE one CLS character at to paper top
'He wrote a character on the paper.'

b. Ta1 diu1 le yi1 jian4 yi1fu2 dao4/zai4 chuang2-xia4.
he throw LE one CLS dress to at bed under
'He threw a dress under the bed.'

c. Ta1 ban1 le yi1 kuai4 shi2tou2 dao4/*zai4 fang2-li
he move LE one piece rock to at house inside
'He moved a rock into the house.'

As we can see from the above examples, although most verbs in this subcategory may appear with either zai4 or dao4, both of which have PCASE LOCT, some verbs may occur only with one of them and the choice seems to be related the meaning of the verb.

FI-V-8:
[CAT V
FS [PRED <SUBJ , OBJ , OBL_{LOCT}>]
(↑ OBL_{LOCT} PCASE) =$_c$ LOCT
]

diu1:
[FS [FORM 'diu1']
FI-V-8
]

xie3:
[FS [FORM 'xie3']
(↑ OBL_{LOCT} PFORM) =$_c$ 'zai4'
FI-V-8
]

ban1:
[FS [FORM 'ban1']
(↑ OBL_{LOCT} PFORM) =$_c$ 'dao4'
FI-V-8
]

3.3.9 <SUBJ , OBL_{THME} , OBL_{LOCT}>

Examples: xie3 'write', gua4 'hang', diu1 'throw', ban1 'move', ji4 'mail', pai4 'sent', tui1 'push', tiao1 'carry', dai4 'carry', mai4 'sell', kai1 'drive', fang4 'put', ti1 'kick'

1. a. Ta1 ba3 zi4 xie3 zai4 zhi3- shang4.
he BA character write at paper top
'He wrote the character on the paper.'

b. Ta1 ba3 yi1fu2 diu1 dao4 chuang2-xia4.
he BA dress throw to bed under
'He threw a dress under the bed.'

FI-V-9:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{LOCT}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{LOCT} PCASE) =$_c$ LOCT
]

Verbs in this subcategory seem to coincide with the previous subcategory of 3.3.8; thus, we assume that there is a derivational relation between verbs of the two subcategories.

3.3.10 <SUBJ , OBL_{THME} , OBJ , OBL_{LOCT}>

Examples: xie3 'write', gua4 'hang', diu1 'throw', ban1 'move', ji4 'mail', pai4 'sent', tui1 'push', tiao1 'carry', dai4 'carry', mai4 'sell', kai1 'drive', fang4 'put', ti1 'kick'

1. a. Ta1 ba3 hua4 gua4 le yi1 jian4 zai4 qiang2-shang4.
he BA painting hang LE one CLS at wall top
'He hung one of the paintings on the wall.'

b.*Ta1 ba3 hu2li2 gua4 le pi2 zai4 qiang2-shang4.
he BA fox hang LE skin at wall top
'He hung the fox's skin on the wall.'

FI-V-10:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ , OBL_{LOCT}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{LOCT} PCASE) =$_c$ LOCT
(↑ OBJ FORM) =$_c$ NONE
]

Since the OBJ here is always linked to a patient role and it must be headless and thus without a head FORM, as shown with the above two sentences, the

general semantic restriction of the whole-part relationship between the OBL_{THME} and OBJ applies here as well. Also, the OBJ here cannot be a full NP and therefore must not have a FORM attribute. The constraint (↑ OBJ FORM) $=_c$ NONE can be satisfied either if FORM does not exist in OBJ or FORM actually has the value NONE. Incidentally, verbs in this subcategory are derivationally related to verbs in both 3.3.8 and 3.3.9 above.

3.3.11 <SUBJ , NCOMP>

3.3.11.1 Equational Verbs

Examples: shi4 'be', cheng2wei2 'become', jiao4 'be called', jiao4zuo4 'be called', hao4cheng1 'be known as'.

1. a. Ma3li4 shi4 ta1 de lao3shi1.
Mary be s/he DE teacher
'Mary is his/her teacher.'

b. Ma3li4 shi4 zi4ji3 de lao3shi1.
Mary is self DE teacher
'Mary is her own teacher.'

c. Ma3li4 da3 ta1 de lao3shi1.
Mary hit s/he DE teacher
'Mary hit his/her teacher.'

All the so-called equational verbs belong to this class. They subcategorize <SUBJ , NCOMP>. An NCOMP, although assigned by an NP, is a predicative complement, which therefore in turn requires an internal SUBJ. Sentences 1a-b above provide one kind of evidence for the analysis of NCOMP. Notice that it is impossible for Ma3li4 'Mary' and ta1 's/he' to be co-referential in 1a. Thus, in order for the sentence to have the possible reading "Mary is her own teacher," the possessive NP has to be a reflective pronoun, such as in 1b. Yet, in other sentences of transitive verbs with an identical c-structure, these two elements are free to be co-referential. Thus, in

1c, Ma3li4 can be co-referential with ta1 and thus the sentence may be interpreted as "Mary hit her own teacher."

*$NP1_i$ shi4 $NP2_i$ de NP3 (1a) versus
$NP1_i$ da3 $NP2_i$ de NP3 (1c)

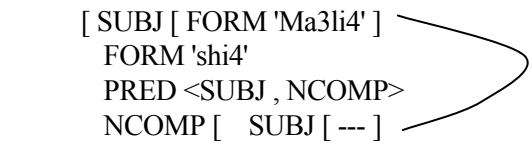
If we assume that the two sentences of 1a and 1c also have similar f-structures as well, then we really cannot explain the difference of co-referentiality in the two sentences. We thus assume that 1a actually has a different f-structure from 1c. We analyze these equational verbs as subcategorizing an NCOMP whose SUBJ is functionally controlled by the SUBJ of the matrix verb. Since an extremely wide range of nouns may appear in the NCOMP, it is non-economical to treat all the nouns as predicative and subcategorizing a SUBJ. Instead we posit a non-thematically assigned <SUBJ @> to the PRED attribute in the NCOMP through the matrix verb so that the NCOMP will not be incoherent. Also note that since we analyze this SUBJ as a non-thematic function, it has no correspondence of a thematic role.

FI-V-11:
[CAT V
FS [PRED <SUBJ , NCOMP>]
(↑ NCOMP SUBJ) = (↑ SUBJ)
]

FI-V-11-1:
[(↑ NCOMP PRED) = <SUBJ @>]

shi4:
[FS [FORM 'shi4']
FI-V-11
FI-V-11-1
]

1a-f.



1c-f.

```
[ SUBJ [ FORM 'Ma3li4' ]
  FORM 'da3'
  PRED <SUBJ , OBJ>
  OBJ [ POSS [ FORM 'ta1' ]
        FORM 'lao3shi1'
] ]
```

Given the above two distinctive different f-structures, we can now account for their difference in co-referentiality. "Nucleus" is defined as "the f-structure domain that contains a PRED attribute" (recall that in our vLFG formalism only lexical forms have the PRED feature). Furthermore, we will assume the following anaphoric binding principles for Chinese nominals:

Anaphoric Binding Principles in vLFG for Chinese:

A: Reflexive pronouns must be bound by an antecedent within the minimal nucleus containing the reflexive pronoun and a subjective function (i.e., SUBJ or POSS).

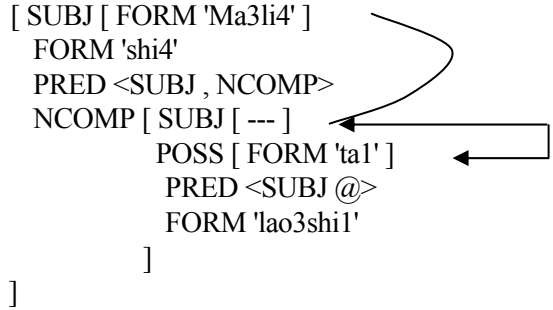
B: Non-reflexive pronouns must not be bound by an antecedent within the minimal f-structure containing the pronoun and a subjective function.

C: Full-NPs (or R-expressions in GB terminology) must be free.

Based on the above f-structures and the anaphoric binding principles, we can now account for why in 1a ta1 cannot have Ma3li4 as its antecedent: given ta1 as a non-reflexive pronoun, it has to be free within the minimal

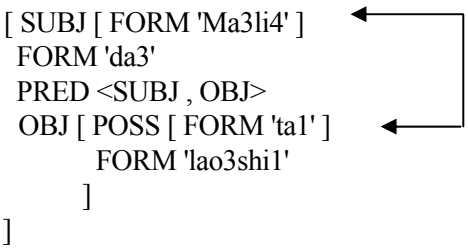
f-structure containing it and a subjective function, and since now the NCOMP is that minimal f-structure, ta1 has to be free in the NCOMP and thus cannot be bound by NCOMP's SUBJ which is now identified as Ma3li4.

1a-f.



In contrast, the minimal f-structure that contains ta1 and a subjective function is the OBJ and ta1 is indeed free in the OBJ. Thus, ta1 may be bound by an antecedent outside of OBJ, so it is possible for Ma3li4 and ta1 to be co-referential in 1c.

1c-f.



Note that in this analysis we still recognize that nouns do have subcategorization requirements in terms of grammatical functions. For example, nouns like yao2yan2 'rumor' subcategorize an SCOMP.

3.3.11.2 Verbs Denoting Inherent Quality

Examples: mai4 'sell for', zhong4 'weigh', chang2 'length is', kuan1 'width is', gao1 'height is', zhi2 'worth'.

2. a. Na4 ben3 shu1 zhi2 wu3 yuan2.
that CLS book worth five dollar
'That book is worth five dollars.'

a' Na4 ben3 shu1 wu3 yuan2.
that CLS book five dollar
'That book is five dollars.'

b. Na4 ge xi1gua1 zhong4 liang3 gong1jin1.
that CLS watermelon weigh two kilogram
'That watermelon weighs two kilograms.'

b' Na4 ge xi1gua1 liang3 gong1jin1.
that CLS watermelon two kilogram
'That watermelon weighs two kilograms.'

Other than the typical equational verbs such as shi4 'be' and cheng2wei2 'become', there are Chinese verbs that denote certain qualities such as worth, weight, length, and height that may also belong to this subcategory. As we can see from the above examples, we analyze the postverbal NP's as assigning the NCOMP function because they seem to be predicative in that they may appear alone without the verbs. Since the nouns allowed in the NCOMP here constitute a rather limited and closed class and thus have to be identified as a subcategory of nouns which may function predicatively and subcategorize <SUBJ>, we do not need to impose a PRED in the NCOMP for it to be coherent. We will now give the FI entry of this subcategory, the lexical entry of zhong4, and the f-structure of 1a.

```
FI-V-11:
        [ CAT V
          FS [ PRED <SUBJ , NCOMP> ]
          (↑ NCOMP SUBJ) = (↑ SUBJ)
        ]

zhong4:
{       [ FS [ FORM 'zhong4' ]
           FI-V-11
         ]
          [ FS [ FORM 'zhong4' ]
           FI-V-1
           FI-V-STATE
          ]
}

2a-f.
        [ SUBJ [ FORM 'shu1'
                 DET   'na4'
                 CLS   'ben3'
               ]
          FORM 'zhi2'
          PRED <SUBJ , NCOMP>
          NCOMP [   SUBJ [ --- ]
                    ADJ   { [ FORM 'wu5' ] }
                    PRED <SUBJ>
                    FORM 'yuan2'
                  ]
        ]
```

3.3.12 <SUBJ , XCOMP>

3.3.12.1 Regular Type

Examples: ji4hua4 'plan', ting2zhi3 'cease', kai1shi3 'commence', zhun3bei4 'prepare', yao4 'want', xiang3, xiang3yao4 'want', xi3huan1 'like', xi1wang4 'hope', tao3yan4 'dislike', ke3wang4 'hope', ju4jue2 'refuse', rong2yi4 'easy', nan2 'difficult'

1. a. Ta1 qi4tu2 qiang3 yin2hang2.
he attempt rob bank
He attempts to rob a bank.

b. Ta1 hen3 ke3wang4 jian4 ni3.
he very hope see you
He very much hopes to see you.

c. Ta1 ju4jue2 ting2zhi3 gong1zuo4.
he refuse stop work
He refuses to stop working.

d. Ta1 hen3 rong2yi4 gan3mao4.
he very easy catch cold
It is very easy for him to catch cold.

Verbs in this class must take an embedded open complement XCOMP, which is a non-finite clause without an overt SUBJ. In addition, the SUBJ of the matrix verb must functionally control the lower SUBJ of the XCOMP. In effect, then, the SUBJ of the matrix verb also has to satisfy the semantic restrictions of the verb in the XCOMP. In terms of finiteness, due to the lack of systematic morphological marking, the distinction between a finite clause and a non-finite clause in Mandarin Chinese is not always clear. However, there are several tests one may apply to detect the finiteness or non-finiteness of a clause. Huang (1982) suggests that a finite clause must have the following three characteristics.

Three Characteristics of a Finite Clause:
A. the potential occurrence of a postverbal aspect marker, such as le ([ASPECT PERFECTIVE]), guo4 ([ASPECT EXPERIENTIAL]), and zhe ([ASPECT PROGRESSIVE]).
B. the potential occurrence of a preverbal modal verb, such as hui4 'will', neng2 'can', etc.
C. the occurrence of an overt lexical subject.

To put the last characteristic more precisely in LFG terms, a finite clause cannot have its subject functionally controlled by a controller. An aspect particle contributes a feature ASPECT to the f-structure of a VP exocentrically, as shown in the following example.

1. e. Wo3 qu4 guo4 Tai2bei3.
I go GUO Taipei
I have been to Taipei.

1e-f.

```
[ FORM 'qu4'
  ASPECT EXPERIENTIAL
  PRED   <SUBJ , OBJ>
  SUBJ   [ FORM 'wo3' ]
  OBJ    [ FORM 'Tai2bei3' ]
]
```

Given the assumption that the matrix verb of a sentence is always finite, we only need to be concerned about the finiteness of embedded clauses. These tests thus provide the means to confirm what we may intuitively determine as to whether an embedded clause is a finite SCOMP or a non-finite XCOMP. Take 1a for example: the embedded VP has to be considered a non-finite clause for it does not pass any of the above three tests for finiteness, as we will illustrate below.

1. f.*Ta1 qi4tu2 qiang3 le ying2hang2.
he attempt rob LE bank
*He attempts to have robbed a bank.

g.*Ta1 qi4tu2 neng2 qiang3 ying2hang2.
he attempt can rob bank
*He attempts to can rob a bank.

h.*Ta1 qi4tu2 ni3 qiang3 ying2hang2.
he attempt you rob bank
He attempts for you to rob a bank.

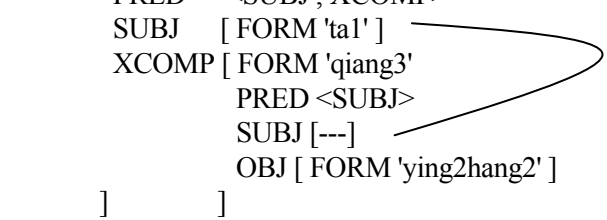
We will account for the control relation and the fact that the embedded clause has to be non-finite (i.e., no modality or aspect) in the default FI entry of all verbs in this subcategory. Recall that NONE is a special value, which conflicts with any other overt value in unification. When there is a constraint on an f-structure for a certain feature's value to be NONE, only under two circumstances that f-structure can satisfy this constraint: 1) there is indeed this feature with the value NONE, or 2) this feature does not exist.

FI-V-12:
[CAT V
FS [PRED <SUBJ , XCOMP>]
]

FI-V-12-1:
[(↑ SUBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODALITY) $=_c$ NONE
(↑ XCOMP ASPECT) $=_c$ NONE
]

qi4tu2: `attempt
[FS [FORM 'qi4tu2']
FI-V-12
FI-V-12-1
]

1a-f. `Ta1 qi4tu2 qiang3 ying2hang2.



3.3.12.2 Tough Construction

Examples: rong2yi4 'easy', nan2 'difficult', zhi2de2 'worthwhile'

2. a. Gan3mao4 hen3 rong2yi4 yi1zhi4.
cold very easy cure
Colds are easy to cure.

This is the so-called "tough" construction. It seems that all verbs in this subcategory are gradable state verbs and they require an embedded clause which lacks the potential occurrence of an overt subject, aspect markers, and modal verbs. We thus determine that this clause is a non-finite clause, assigning the function XCOMP. The difference of the XCOMP in a "tough" construction and elsewhere is that here it not only has an unfulfilled SUBJ, but also an unfulfilled OBJ as well. In addition, the matrix verb must assign a value of OPT to the XCOMP's SUBJ for the XCOMP to be complete since here it can neither be overt nor be functionally controlled. The following sentence is therefore ungrammatical.

2. b.*Gan3mao4 hen3 rong2yi4 yi1sheng1 yi1zhi4.
cold very easy doctor cure
Colds are easy for doctors to cure.

Although the matrix SUBJ functionally controls the XCOMP's OBJ, the control relation between the two grammatical functions is potentially

unbounded, as there could be more non-finite clauses embedded within the XCOMP.

2. c. Gan3mao4 hen3 rong2yi4 she4fa3 yi1zhi4.
cold very easy try cure
Colds are easy to try to cure.

d. Na4 ge gu4shi4 zhi2de2 gao4su4 ta1men.
that CLS story worth tell they
That story worth telling them.

In 2c, the matrix SUBJ controls the unbounded OBJ embedded in the XCOMP's XCOMP, but in 2d it controls the unbounded OBJ2 of gao4su4 'tell'. We adopt the regular language of describing paths of unbounded (or long distance) dependency developed by Kaplan and Zaenen (1987). Recall that X+ indicates "one or more instances of X" and {X Y} indicates "either X or Y." The expression (↑ SUBJ) = (↑ XCOMP+ {OBJ OBJ2}) thus specifies that the matrix SUBJ, going through the path of one or more XCOMPs, controls either an OBJ or OBJ2, which is the end of the path. Furthermore, semantically verbs in this subcategory require only one thematic role which links to the XCOMP function; consequently the SUBJ function required in PRED is non-thematically assigned. The following FI entry unique to verbs in this subcategory captures these characteristics of the "tough" construction in Chinese. We will also give the f-structure of 2c as an example.

FI-V-12-2:

```
[ CAT V
  FS   [ PRED     <SUBJ @ , XCOMP>
         XCOMP [ SUBJ OPT ]
       ]
  (↑ SUBJ) = (↑ XCOMP+ {OBJ OBJ2})
  (↑ XCOMP MODALITY) =c NONE
  (↑ XCOMP ASPECT) =c NONE
]
```

rong2yi4: `easy

```
[ FS [ FORM 'rong2yi4' ]
  FI-V-12
  FI-V-12-2
  FI-V-STATE
]
```

2c-f. `Gan3mao4 hen3 rong2yi4 she4fa3 yi1zhi4.

```
[ FORM    'rong2yi4'
  PRED    <SUBJ , XCOMP>
  SUBJ    [ FORM 'gan3mao4' ]
  XCOMP [ FORM 'she4fa3'
            PRED <SUBJ , XCOMP>
            SUBJ OPT
            XCOMP [ FORM 'yi1zhi4'
                      PRED   <SUBJ , OBJ>
                      SUBJ    [ --- ]
                      OBJ     [ --- ]
          ]
  ACTIVITY  -
  PROCESS   -
  ADJ     { [ FORM 'hen3' ] }
]
```

Note however that rong2yi4 'easy' and nan2 'difficult' also take an XCOMP which is complete, except the usual missing SUBJ, and thus does not have an unbounded gap of OBJ, for example sentence 1d of 3.3.12.1, repeated below. Therefore, to account for the two different constructions, we have to recognize two separate lexical forms for rong2yi4 'easy'.

1. d. Ta1 hen3 rong2yi4 gan3mao4.
he very easy catch cold
It is very easy for him to catch cold.

```
rong2yi4:                                              `easy
{       [ FS [ FORM 'rong2yi4' ]                  `for 2a, 2c
          FI-V-12
          FI-V-12-2
          FI-V-STATE
        ]
        [ FS [ FORM 'rong2yi4' ]                  `for 1d of 3.3.12.1
          FI-V-12
          FI-V-12-1
          FI-V-STATE
        ]
}
```

According to our account of Mandarin "tough" construction, the movement account that Hou (1979) provides misses several important characteristics of this syntactic construction. First of all, he uses ke3neng2 'possible' as an example; unfortunately, while "possible" and "impossible" in English do appear in the "tough" construction, in Chinese ke3neng2 'possible', either positive or negative, does not appear in the "tough" construction.

2. e. He is impossible to teach.
   f. He is possible to beat.

   g. Ta1 bu4 ke3neng2 jiao1.
      he not possible teach
      For him to teach is impossible.

The Chinese sentence 2g therefore does not have the reading of the "tough" construction of 2e. Likewise, the example sentence that Hou (ibid.:65) provides (repeated here as 2h below) is not of this "tough" construction.

2. h. Nei5 jian4 shi4 ta1 bu4 ke3neng2 zhi1dao4.
      that CLS matter he not possible know
      That matter is impossible for him to know. (Hou's translation)
      For him to know that matter is impossible. (My translation)

i. [Ta1 zhi1dao4 nei5 jian4 shi4] bu4 ke3neng2.
he know that CLS matter not possible
It is impossible that he knows that matter.

Hou considers that 2h, before all the movement takes place, has the "original" shape of 2i. This is rather misconstrued, even within the movement account, for 2h simply has the lowest object, nei5 jian4 shi4, topicalized. Otherwise, it is equivalent to the following sentence.

2. j. Ta1 bu4 ke3neng2 zhi1dao4 nei5 jian4 shi4.
he not possible know that CLS matter
For him to know that matter is impossible.

Thus, ke3neng2 is a modal verb in 2h and 2k, just like bi4xu1 'must' and neng2gou4 'can' in the following sentence 2k. As we will see in the next section, although modal verbs do subcategorize SUBJ and XCOMP, it is the matrix SUBJ that controls the XCOMP's SUBJ. They do not function as verbs like rong2yi4 'easy' that appear in the "tough" construction.

2. k. Ta1 bi4xu1/neng2gou4 jie3shi4 nei5 jian4 shi4.
he must /can explain that CLS matter
He must/can explain that matter.

Furthermore, Hou (ibid.:64-65) characterizes the "tough" movement as the following: "Tough Movement in Chinese is a rule which raises the object of an embedded clause so that it becomes the subject of a higher clause, while the original subject and verb are extraposed to the right of the predicate." His account thus misses three important characteristics of this construction: first, the missing (or raised) object in the embedded clause may be of unbounded dependency, second, the subject of the embedded clause cannot be overt nor can it be controlled within the sentence, and third, the embedded clause has to be non-finite.

3.3.12.3 Modal Verbs

Examples: hui4 'will', neng2 'can', ying1gai 'should', bi4xu1 'must', ke3yi3 'may', gan3 'dare', ken3 'willing', ke3neng2 'may'.

3. a. Ta1 ming2tian1 hui4 lai2.
he tomorrow will come
He will come tomorrow.

b.*Ta1 hui4 kan4 le shu1.
he will read LE book
He will have read the book.

c.*Ta1 hui4 chi2 guo4 shi2tou2.
he will eat GUO rock
*He will have eaten rocks before.

In terms of syntactic categories or parts of speech, there is little justification for a separate category for the so-called auxiliaries in a grammar of Chinese, for syntactically they behave very similarly to other main verbs in a sentence. They can be negated, can stand alone as a short reply and can form a question with the A-not-A construction, three characteristics associated with Chinese verbs (Chao 1968). Thus, they should be treated just like verbs. We will simply identify them as modal verbs. Like other verbs in 3.3.12, they subcategorize an XCOMP, where no aspect is allowed, as shown in 3b-c above. Like verbs of 3.3.12.1, the SUBJ of the matrix verb functionally controls the XCOMP's SUBJ. Another characteristic of the modal verbs is that they do not impose any semantic selectional restrictions on their SUBJ. However, the fact that the SUBJ of a modal verb functionally controls its XCOMP's SUBJ means that in effect the verb in the XCOMP imposes its selectional restrictions on the matrix SUBJ.

Li and Thompson (1981:173) state that one of the characteristics of modal verbs is that they cannot be modified by intensifiers, such as hen3 'very'. We dispute their observation: although the majority of the modal verbs are not gradable, others such as ying1gai 'should', neng2 'can', hui4 'can', gan3 'dare' seem to have the characteristic of state verbs. The following example sentence

they claim is ungrammatical is actually perfectly acceptable to most native speakers. Zhang (1983:101-102) has made the same observation.

3.c. Ta1 hen3 hui4 chang4ge1.
he very can sing
He is very able to sing.

Furthermore, according to Chao's (1968:665) classification of verbs, what he calls auxiliaries such as hui4 'can' are able to be modified by degree adverbs like hen3 'very'. However, we believe that each modal verb has to be individually marked in terms of its gradability, just like verbs of other subcategories. Take xiang3 'miss, think' and ren4wei2 'think' for example: while xiang3 'miss, think' is gradable when it subcategorizes an object, it is not gradable when subcategorizing a sentential complement. Yet, ren4wei2 'think', though similar in meaning to xiang3 in the latter use, is gradable when subcategorizing an SCOMP.

3. d. Ta1 hen3 xiang3 ni3.
he very miss you
He misses you very much.

e.*Ta1 hen3 xiang3 ni3 ying1gai1 nu3li4.
he very think you should work-hard
He very much thinks that you should work hard.

f. Ta1hen3 ren4wei2 ni3 ying1gai1 nu3li4.
he very think you should work-hard
He very much thinks that you should work hard.

Similarly, while most of the modal verbs are gradable, some are not, e.g., bi4xu1 and dei3 'must' and hui4 'will' (when it indicates possibility, not ability). Very correctly, however, Li and Thompson point out that, unlike most other verbs, modal verbs can never take aspect markers, such as le, guo and zhe. This fact certainly has to be accounted for. We thus posit in the FI entry of modal verbs a feature-value pair of [ASPECT NONE] to make sure that in case an overt aspect marker occurs with a modal verb, the value of the

feature ASPECT in the FS of the aspect marker necessarily conflicts with the NONE value in the modal verb. We also posit [MODALITY ANY] as default for modal verbs to ensure that each of them has a real value of modality.

FI-V-12-3:
[FS [ASPECT NONE
MODALITY ANY
]
]

gan3: `dare
[FS [FORM 'gan3'
MODALITY ABILITY
]
FI-V-12
FI-V-12-1
FI-V-12-3
]

However, unlike non-modal verbs that subcategorize XCOMP, it is possible for some of the modal verbs to take anther modal verb as its complement and thus allow modality in their XCOMP.

3. g.Ta1 ke3neng2 neng2 ge1chang4.
he may can sing
He may be able to sing.

h.Ta1 bi4xu1 yuan4yi4 ge1chang4.
he must willing sing
He must be willing to sing.

i. Ta1 hui4 gan3 ge1chang4.
he will dare sing
He will dare to sing.

In some dialects of English, similar phenomenon exists, and expressions like "He might can come" is acceptable. According to Cheng (1990 and personal communication), modal verbs are of two types: speaker-oriented and subject-oriented. Speaker oriented modal verbs, such as ying1gai1 'should (possibility)', ke3neng2 'may (possibility), bi4xu1 'must (obligation)',and hui4 'will (possibility)', which express the speaker's characterization of or attitude towards the entire situation depicted, seem to have the ability to take modal complements. However, other modal verbs like gan3 'dare', yuan4yi4 'willing', ken3 'willing', hui4, neng2 and neng2gou4 'can (ability)' that are subject-oriented in that they are more intrinsically related to the subject's perspective and do not reflect the speaker's judgement on the situation. Subject-oriented modal verbs do not allow modal complements.

3. j.*Ta1 neng2 ke3neng2 ge1chang4.
he can may sing
He can possibly sing.

k.*Ta1 yuan4yi4 bi4xu1 ge1chang4.
he willing must sing

l.*Ta1 gan3 hui4 ge1chang4.
he dare will sing

m.*Ta1 neng2gou4 ying1gai1 ge1chang4.
he can should sing

Since subject-oriented modal verbs do observe the constraints posed in the FI entry of 3.3.12.1, we can simply impose the same constraints on them. We will give two examples.

```
yuan4yi4:                                   `subject oriented
        [ FS [ FORM 'yuan4yi4'              `willing
               MODALITY ABILITY
             ]
          FI-V-12
          FI-V-12-1
          FI-V-12-3
        ]

ken3:                                       `subject oriented
        [ FS [ FORM 'ken3'                  `willing
               MODALITY ABILITY
             ]
          FI-V-12
          FI-V-12-1
          FI-V-12-3
        ]
```

However, since speaker-oriented modal verbs do allow modal complements, FI-V-12-1 does not apply. We will set up another entry where such constraint is removed.

```
FI-V-12-3-SPEAKER:
        [ (↑ SUBJ) = (↑ XCOMP SUBJ)
          (↑ XCOMP ASPECT) =c NONE
        ]

ke3neng2:                                   `speaker-oriented
        [ FS [ FORM 'ke3neng2'              `may (possibility)
               MODALITY POSSIBILITY
             ]
          FI-V-12
          FI-V-12-3
          FI-V-12-3-SPEAKER
        ]
```

```
bi4xu1:                                          `speaker-oriented
        [ FS [ FORM 'bi4xu1'                     `must (obligation)
               MODALITY OBLIGATION
             ]
          FI-V-12
          FI-V-12-3
          FI-V-12-3-SPEAKER
        ]
```

Chao (1968) includes verbs like xiang3 'want', zhi2de2 'worth', xiang3yao4 'want', gao1xing4 'happy', pa4 'afraid', and others as modal auxiliaries. In fact, none of these verbs should be considered modal verbs, as both CKIP and Li and Thompson have noticed. As stated earlier, no modal verbs may appear in a non-finite clause after subject-oriented modal verbs, for example 3o below. The fact that all these verbs above can appear in a non-finite clause subcategorized by a subject-oriented modal verb such as gan3 'dare' indicates that they are not modal verbs.

3. o.*Ta1 gan3 ying1gai1 ge1chang4.
He gan3 should sing
*He dares should sing.

p. Ta1 hui4 xiang3 qu4.
he will want go
He will want to go.

q. Ta1 hui4 zhi2de2 ai4.
he will worth love
He will be worth loving.

r. Ta1 hui4 gao1xing4 jian4 ni3.
he will happy see you
He will be happy to see you.

In the above examples, we clearly see that only ying1gai1 'should', which can never appear in an embedded non-finite clause of a subject-oriented

modal verb, passes the test as a modal verb; all the others are not. Therefore, all these non-modal verbs should be excluded from Chao's list.

It would seem that the number of Chinese modal verbs should be rather limited; yet, a complete inventory of these verbs is still rather controversial. Several extensive, but varying, lists are available in the literature: e.g., Chao (1968), Li and Thompson (1981), Lu (1984), and CKIP (1989). The CKIP list is noticeably more complete than the others. There are in general several types of modality recognized in Chinese. Lu (1984) and CKIP (1989) recognize four, while Chao (1968) recognizes three. We will assign four possible values to the feature MODALITY: OBLIGATION, PERMISSION, ABILITY, and POSSIBILITY. Noting that there are likely correlations between the semantics of speaker/subject-orientation and the semantics of modality to be further explored, we will give the following classification of modal verbs. However, neither the list nor the classification is intended to be comprehensive. We will use [sp] to indicate that the verb is speaker-oriented and [su] subject-oriented. Gradability is marked by *, and possible modality by +. For example, ying1gai1 'should' has two possible uses, one under obligation and the other possibility, and while both uses are speaker-oriented, only when used as having modality obligation can ying1gai1 'should' be modified by degree adverbs like hen3 'very'.

Table 3.2
Classification of Mandarin Modal Verbs

	OBLIGATION	PERMISSION	ABILITY	POSSIBILITY
bi4xu1	+ [sp]			
xu1yao4	+ [sp] *			
dei3	+ [sp]			
bu2yong4	+ [sp] *			
ying1gai1	+ [sp] *			+ [sp]
bu4zhun3		+ [sp]		
neng2		+ [sp]	+ [su] *	
ke3yi3		+ [sp]	+ [su] *	
neng2gou4		+ [sp]	+ [su] *	
gan3			+ [su] *	
ken3			+ [su] *	

yuan4yi4			+ [sp] *	
qing2yuan4			+ [sp] *	
hui4			+ [su] *	+ [sp]
ke3neng2				+ [sp] *

+ = possible use
* = gradable by degree adverbs
[sp] = speaker-oriented
[su] = subject-oriented

3.3.12.4 Clause Union Verbs

Examples: she4fa3 'try', tou2zi1 'invest', gao1xing4 'happy', nu3li4 'hard-working', sui2bian4 'casual', fei4xin1 'devoted', zhuan1xin1 'attentive', pin1ming4 'struggle', qi4tu2 'attempt'

4. a. Ta1 she4fa3 qiang3 le yin2hang2.
he tried rob LE bank
He tried to rob a bank.

b. Ta1 she4fa3 qiang3 guo4 yin2hang2.
he tried rob GUO bank
He has tried to rob a bank before.

c.*Ta1 she4fa2 neng2 qiang3 yin2hang2.
he try can rob bank
*He try to can rob a bank.

d.*Ta1 she4fa3 ni3 qiang3 yin2hang2.
he try you rob bank
He tries for you to rob a bank.

The important difference between verbs of this group and verbs in 3.3.12.1 is that the embedded clause, XCOMP, of verbs here such as she4fa3 'try' may contain an aspect particle, e.g., 4a-b. These are the so-called "clause union verbs" observed by Li (1985), where she gives a GB analysis. Due to the fact

that the embedded clause here still cannot have an overt subject or modal verbs, e.g., 4c-d, we believe the clause is still a non-finite one. However, in the case of aspect, the matrix verb and the embedded verb fuse together and the aspect marking occurs with the embedded verb, not the matrix verb.

FI-V-12-4:

[(↑ SUBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODALITY) $=_c$ NONE
(↑ ASPECT) $=_c$ NONE
(↑ XCOMP ASPECT) = (↑ ASPECT)
]

she4fa3: `attempt

[FS [FORM 'she4fa3']
FI-V-12
FI-V-12-4
]

We will compare a verb from 3.3.12.1, ji4hua4 'plan', with a clause union verb form here, she4fa3 'try', in the following examples.

4. e. Ta1 she4fa3 qiang3 guo4 yin2hang2.
he tried rob GUO bank
He has tried to rob a bank before.

f. Ta1 mei2 she4fa2 qiang3 guo4 yin2hang2.
he not try rob GUO bank
He has never tried to rob a bank before.

e' Ta1 ji4hua4 guo4 qiang3 yin2hang2.
he plan GUO rob bank
He has planned to rob a bank before.

f' Ta1 mei2 ji4hua4 guo4 qiang3 yin2hang2.
he not plan GUO rob bank
He has never planned to rob a bank before.

In both 4e and 4e', although the "trying" and "planning" have taken place, whether the action of "robbing a bank" has ever happened is not indicated; yet, the aspect marker guo4 has to appear with the matrix verb ji4hua4 'plan' in 4e' but with the verb in XCOMP, qiang3 'rob', in 4e. In addition, that the negation of 4e and 4e' is the same, as shown in 4f and 4f', again indicates that in 4e the scope of the aspect marker is not limited to the verb in XCOMP and rather its scope covers the matrix verb sh4fa3 'try' as well.

There might be some doubt regarding whether the state verbs cited in this subcategory, such as nu3li4 'hard-working' and zhuan1xin1 'attentive', are the verbs or adverbs. We contend that they are main verbs because, first of all, unlike the majority of manner adverbs they cannot appear before the subject.

4. g. Ta1 zhuan1xin1 kan4wan2 le na4 ben3 shu1.
he attentive read-finish LE that CLS book
He concentrated his attention and finished that book.

h.*Zhuna1xin1 ta1 kan4wan2 le na4 ben3 shu1.
attentive he read-finish LE that CLS book
Attentively, he finished that book.

i. Ta1zhuan1xin1de kan4wan2 le na4 ben3 shu1.
he attentively read-finish LE that CLS book
He attentively finished that book.

j. Zhuna1xin1de ta1 kan4wan2 le na4 ben3 shu1.
attentively he read-finish LE that CLS book
Attentively, he finished that book.

Secondly, most of the state verbs do not have this complementation pattern; for example, none of the following state verbs has this subcategorization: jin3shen4 'careful', cu1xin1 'careless', da4fang1 'generous', kuai4le4 'happy', shang1xin1 'sad', etc. The fact that only a seemingly arbitrarily restricted class

of state verbs may require this complementation pattern provides another support for our analysis: while the complementation patterns of lexical forms may often be idiosyncratic among verbs of similar meanings, the syntactic distribution in the c-structure of lexical items in the same syntactic category is usually consistent and less restrictive.

Finally, all three characteristics of verbs cited by Chao (1968) can be found with these verbs: (a) they can be negated, (b) they can form A-not-A questions, and (c) they can stand alone in a short answer.

4. k. Ta1 bu4 zhuan1xin1 kan4 shu1.
he not attentive read book
He is not concentrated in reading.

l. Q: Ta1 zhuan1(xin1)-bu2-zhuan1xin1 kan4shu1?
he attentive not attentive read
Is he concentrated in reading or not?

m. A: Zhuan1xin1.
attentive
Yes, he is concentrated.

3.3.13 <SUBJ , SCOMP>

3.3.13.1 Regular Type

Examples: shuo1 'say', fa1xian4 'discover', bao3zheng4 'guarantee', zhi1dao4 'know', xiang3 'think', jian1chi2 'insist', xiang1xin4 'believe', shi1wang4 'disappointed', que4ding4 'sure', pan4wang4 'hope', ren4wei2 'think', xi1wang4 'hope', huai2yi2 'doubt'

1. a. Ma3li4$_i$ que4ding4 ta1$_{i/j}$ bu2 hui4 lai2.
Mary sure s/he not will come
Mary is sure that s/he won't come.

b. Ma3li4$_i$ xiang1xin4 ta1$_{i/j}$ lai2 guo4 Tai2bei3.
Mary believe s/he come GUO Taipei
Mary believes that s/he has been to Taipei before.

The verbs in this class subcategorize a SUBJ and SCOMP. An SCOMP is an embedded finite clause with all of its subcategorized grammatical functions, thus including SUBJ, completely fulfilled. Also, as shown in the above two examples, an SCOMP, being finite, can take modal verbs or aspect particles. In addition, the fact that the pronoun following the verb may be co-referential with the matrix SUBJ indicates that this pronoun has to be the SUBJ of the SCOMP and it cannot be the OBJ of the matrix verb.

FI-V-13:
[CAT V
FS [PRED <SUBJ , SCOMP>]
]

que4ding4: `sure
[FS [FORM 'que4ding4']
FI-V-13
]

Note that verbs here do not require an indirect question as its complement; verbs that subcategorize <SUBJ , SCOMP> and require the SCOMP to be in the form of an indirect question are separately grouped under 3.3.13.2 below due to this requirement.

3.3.13.2 Verbs Requiring an Interrogative SCOMP

Examples: diao4cha2 'investigate', yan2jiu4 'study', tan4tao3 'examine', wen4 'ask', kao3cha2 'research', kan4kan4 'take a look', kao3lu4 'deliberate'

2. a. Jing3cha2 hui4 diao4cha2 shei2 sha1 le ta1.
police will investigate who kill LE he
The police will investigate who killed him.

b.*Jing3cha2 hui4 diao4cha2 ni3 sha1 le ta1.
police will investigate you kill LE he
*The police will investigate you killed him.

c. Wo3men zai4 yan2jiu4 ni3 hui4 bu2 hui4 sha1 ta1.
we now study you will not will kill he
We are studying whether you will kill him or not.

d.*Wo3men zai4 yan2jiu4 ni3 hui4 sha1 ta1.
we now study you will kill he
*We are studying that you will kill him.

The verbs in this subcategory, although they subcategorize a SUBJ and SCOMP as well, require its SCOMP to be interrogative. The SCOMP thus must have the Q attribute with the value +, and such an attribute-value pair is supplied only by wh-words, wh-phrases, or A-not-A form of question. In other words, the embedded clause that verbs here subcategorize must be able to be interpreted as an indirect question.

FI-V-13-2:
[(↑ SCOMP Q) $=_c$ +]

diao4cha2: `investigate
[FS [FORM 'diao4cha2']
FI-V-13
FI-V-13-2
]

Cheng (1984:141-143) provides a fairly comprehensive classification of Mandarin verbs that subcategorize sentential complements according to their meanings. We will cite his classification with slight modification and some of his examples below. These semantic types will be related to our verb subcategories. Note that only the last type, inquisitive verbs, requires an indirect question as complement.

1) Verbs of Locution
A. Say-type verbs. `3.3.13.1 and 3.3.24.1

tou4lu4	'reveal'	ti2dao4	'mention'
ti2qi3	'bring up'	gao4bai2	'confess'

B. Tell-type verbs. `3.3.22.1, 3.3.24.1

hui2da2	'answer'	jing3gao4	'warn'
tong1zhi	'notify'	ti2xing3	'remind'

2) Verbs of Cognition. `3.3.13.1

dong2de2	'know'	liao4dao4	'predict'
wang4ji4	'forget'	qing1chu3	'understand'

3) Verbs of Discovery. `3.3.13.1

fa1jue2	'discover'	kan4chu1	'perceive'
meng4dao4	'dream of'	kan4	'see'

4) Verbs of Judgement. `3.3.13.1

gu1ji4	'estimate'	ken3ding4	'ascertain'

5) Verbs of Imagination. `3.3.13.1

A. Assumption verbs.

ren4wei2	'regard'	hai2yi2	'doubt'

B. Approval verbs.

pi1ping2	'criticize'	fan3dui4	'oppose'

C. Thinking verbs

gan3jue2	'feel'	xiang3	'think'

D. Hope verbs.

que4xing4	'convinced'	zhi3wang4	'hope'

6) Inquisitive Verbs.

A. Ask-type verbs. `3.3.13.2, 3.3.22.2, 3.3.24.2

wen4	'ask' `3.3.13.2, 3.3.22.2
wen4dao4	'ask' `3.3.13.2, 3.3.24.2
zhi2wen4	'interrogate' `3.3.13.2, 3.3.22.2, 3.3.24.2
pan2wen4	'interrogate' `3.3.22.2
qing3wen4	'politely ask' `3.3.13.2, 3.3.22.2, 3.3.24.2
zhui1jiu4	'seek a final answer' `3.3.13.2, 3.3.24.2
qing3shi4	'ask for direction' `3.3.13.2, 3.3.24.2

B. Test-type verbs. `3.3.13.2

diao4cha2	'investigate'	yan2jiu4	'study'
tao3lun4	'discuss'	guan1xin1	'concerned'
tan4tao3	'inquire'	shi4yan4	'test'

3.3.14 <SUBJ , OBJ , OBJ2>

3.3.14.1 Regular Type

Examples: gei3 'give', gong1ji3 'provide', song4 'give', jie4 'lend', jiao1 'teach', ti2gong1 'provide', zu1 'rent', chi1 'eat', pian4 'cheat', qiang3 'rob', tou1 'steal', hua1 'cost', yong4 'use', gao4su4 'tell', tong1zhi1 'notify', jing2gao4 'warn', da1ying4 'promise', chuan4gao4 'advise', ti2xing3 'remind', jiao4 'call', cheng1hu1 'address'

1. a. Wo3men ti2gong1 le ta1 xu3duo1 yuan2zhu4.
we provide LE he much help
We provide him lots of help.

b. Ta1 qiang3 le ni3 hen3duo1 qian2.
he rob LE you much money
He robbed you of much money.

c. Ta1 tong1zh11 le ni3 san1 jian4 shi4.
he notify LE you three CLS thing
He notified you three things.

These are the so-called ditransitive verbs. It seems that none of the verbs in this class is a state verb. They subcategorize a direct object and an indirect object, or OBJ and OBJ2. Recall that we have argued in the previous chapter that, within the theory of LFG, it is more appropriate to treat the NP immediately following the verb as having the OBJ function, a semantically unrestricted function which in this current construction may be linked to the thematic roles of beneficiary, e.g., 1a, maleficiary, e.g., 1b, or goal, e.g., 1c. The second NP assigns the grammatical function of OBJ2, which is semantically restricted to be the theme role.

FI-V-14:

```
[ CAT V
  FS   [ PRED <SUBJ , OBJ , OBJ2> ]
]
```

1b-f. `Ta1 qiang3 le ni3 hen3duo1 qian2.

```
[ FORM 'qiang3'
  PRED   <SUBJ , OBJ , OBJ2>
  SUBJ   [ FORM 'ta1' ]
  OBJ    [ FORM 'ni3' ]
  OBJ2   [ FORM 'qian2'
           ADJ { [ FORM 'hen3duo1' ] }
         ]
  ASPECT PERFECTIVE
]
```

3.3.14.2 Idiomatic Expressions

Examples: bang1...mang2 'help'; chi1...dou4fu3 'take advantage of, tease', kai1...wan2xiao4 'joke', fang4...yi1ma3 'forgive'

2. a. Wo3 bang1 le ta1 hen3duo1 mang2.
we help LE he many busy
We helped him a lot.

b. Ta1 xi3huan1 chi1 ni3 dou4fu3.
he likes eat you tofu
He likes to take advantage of you.

We have also observed that the OBJ2, as a subcategorizable function, may be a non-thematic argument. The above two sentences both require a subcategorization of <SUBJ , OBJ , OBJ2>; however, due to the idiomatic nature of the expressions bang1-mang2 'help' and chi1-dou4fu3 'take advantage of, tease', we argue that thematically they require two roles only. Therefore, in terms of syntax, we recognize that bang1 and chi1 here do

subcategorize three grammatical functions, but in terms of semantics they, like verbs of similar meanings bang1zhu4 'help' and tiao2xi4 'tease, flirt with', require only two thematic roles. Since the noun that is required as part of the idiomatic expression appears as the second postverbal NP we thus determine that OBJ2 is non- thematically assigned.

```
FI-V-14-2:
        [ CAT V
          FS    [ PRED <SUBJ , OBJ , OBJ2 @> ]
        ]

bang1:                                          `help
{       [ FS [ FORM 'bang1' ]
          FI-V-4
        ]
        [ FS [ FORM 'bang1-mang2'
               OBJ2    [ FORM 'mang2'
                         BACKGROUND -
                       ]
             ]
          FI-V-14-2
        ]
}
```

It is necessary to allow verbs such as fang4 and bang1 to subcategorize <SUBJ , OBJ , OBJ2 @> with OBJ2's FORM strictly constrained because this subcategorization is not part of their regular syntactic behavior. Incidentally, this analysis in fact also provides another support for our assertion that the first postverbal NP assigns OBJ and the second one assigns OBJ2. Recall that, in LFG theory, there is a hierarchical constraint on the assignment of grammatical functions between OBJ and OBJ2 which states that OBJ2 is assigned only in the case that OBJ has been assigned (Bresnan 1982b:294). Consequently, if we were to consider the first postverbal NP as having the function OBJ2, in the subcategorization of idiomatic verbs of this subcategory we would have to leave OBJ as the function that is non-thematically assigned. That, of course, would violate the hierarchical

constraint on the assignment of OBJ and OBJ2. Nonetheless, there is another piece of evidence against our analysis. Recall that in LFG only semantically unrestricted functions can be non-thematically assigned. This constraint indicates that the second postverbal NP would be assigned OBJ. The status of OBJ2 in Chinese thus remains controversial.

3.3.15 <SUBJ , OBJ , OBL_{BNFC}>

Examples: song4 'give', ji4 'mail', xie3 'write', chuan2zhen1 'telex, fax', gong1ji3 'provide', jie4 'lend', jiao1 'teach', zu1 'rent', ti2gong1 'provide', jia4 'marry', huan2 'return', jie4shao4 'introduce', chuan2 'pass'

1. a. Wo3men hui4 ti2gong1 xu3duo1 yuan2zhu4 gei3 ni3.
we will provide much help to you
We will provide you with lots of help.

b. Ta1 zu1 le yi1 ge fang2jian1 gei3 ni3.
he rent LE one CLS room to you
He rented a room to you.

Except for a few verbs like gei3 'give', most verbs of subcategory 3.3.14 above that subcategorize a beneficiary OBJ also belong to this class. However, we observe that other verbs of 14 which subcategorize a non-beneficiary OBJ do not have derivational counterparts that subcategorize <SUBJ , OBJ , OBL_{BNFC}>. This observation confirms nicely the semantically restricted nature of the function OBL_{BNFC}, which is assigned by the preposition gei3 'to'. Notice also that many of the verbs here do not have counterparts in 3.3.14.1, such as jia4 'marry', xie3 'write', chuan2zhen1 'fax', etc.

FI-V-15:

[CAT V
FS [PRED <SUBJ , OBJ , OBL_{BNFC}>]
($\uparrow$ OBL_{BNFC} PCASE) $=_c$ BNFC
]

```
gei3:                                          `give, to
{       [ FS [ PFORM 'gei3' ]
         FI-V-14
        ]
        [ CAT P
          FS    [ PFORM 'gei3'
                  PCASE BNFC
                ]
        ]
}

jia4:                                          `marry
          [ FS [ FORM 'jia4' ]
            FI-V-15
          ]
```

3.3.16 <SUBJ , OBL$_{THME}$, OBJ , OBJ2>

Examples: gei3 'give', gong1ji3 'provide', song4 'give', jie4 'lend', jiao1 'teach', zu1 'rent', ti2gong1 'provide'

1. a. Ta1 ba3 shu1 gei3 le wo3 san1 ben3.
   he BA book give LE I three CLS
   He gave me three of the books.

   b. Ta1 ba3 mi4mi4 gao4su4 le wo3 hen3duo1.
   he BA secret tell LE I many
   He told many of the secrets.

   c.*Ta1 ba3 shu1 gei3 le wo3 sheng4jing1.
   he BA book give LE I Bible
   Among the books, he gave me the Bible.

   d.*Ta1 ba3 shu1 gei3 le wo3 san1 tou2.
   he BA book give LE I three CLS
   He gave me three of the books.

e.?Ta1 ba3 shu1 gei3 le wo3 san1 pian4.
he BA book give LE I three CLS
He gave me three pieces of the book.

Notice that not all the ditransitive verbs that subcategorize <SUBJ , OBJ , OBJ2> have derivational counterparts here that subcategorize <SUBJ , OBL_{THME} , OBJ , OBJ2>; it seems only ditransitive verbs that require a beneficiary or a goal OBJ have counterparts in this subcategory, so verbs like qiang3 'rob' that take a maleficiary OBJ do not have counterparts in the current subcategory.

The semantic constraint of a whole-part relationship applies between OBL_{THME} and the OBJ2. From the example sentences above we can tell that the OBJ2 here has to be headless and thus cannot be a full NP, e.g., 1c. More specifically, the OBJ2 has to be quantity expressions such as classifiers, measure words and quantifiers. In terms of f-structure, then, OBJ2 cannot have a head feature FORM. We therefore impose a constraint on OBJ2's FORM. (See the FI entry below.) Furthermore, since semantically OBJ2 has to be able to be interpreted as a part of the entity in OBL_{THME}, when such an interpretation is impossible, the sentence will be unacceptable. Hence, sentence 1d above is ill-formed because, due to the restriction of the classifier tou2, which has to refer to animals of large size, what san1 tou3 'three' is referred to cannot be possibly interpreted as part of the definite shu1 'books' in OBL_{THME}. Similarly, 1d is questionable because such an interpretation is difficult; yet, if the book in question is understood as a Chinese antique book made of many pieces of bamboo slips, then the sentence should be acceptable to most native speakers. Thus, with this understanding, we find the following sentence definitely acceptable.

1. f. Ta1 ba3 zhu2jian3 Yi4jing1 gei3 le wo3 san1 pian4.
he BA bamboo-piece Yi-Ching give LE I three CLS
He gave me three pieces of the Yi-Ching made of bamboo slips.

FI-V-16:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ , OBJ2>]

($\uparrow$ OBL_{THME} PCASE) $=_c$ THME
($\uparrow$ OBJ2 FORM) $=_c$ NONE
]

3.3.17 <SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}>

Examples: chuan2zhen1 'telex, fax', song4 'give', ji4 'mail', xie3 'write', gong1ji3 'provide', song4 'give', jie4 'lend', jiao1 'teach', zu1 'rent', ti2gong1 'provide', jia4 'marry', huan2 'return', jie4shao4 'introduce'

1. a. Ta1 ba3 wen2jian4 chuan2zhen1 le yi1 fen4 gei3 wo3.
he BA document fax LE one CLS to I
He faxed to me one of the documents.

b.*Ta1 ba3 wen2jian4 chuan2zhen1 le feng1mian4 gei3 wo3.
he BA document fax LE cover to I
He faxed to me the cover of the documents.

c.*Ta1 ba3 wen2jian4 chuan2zhen1 le yi1 tou2 gei3 wo3.
he BA document fax LE one CLS to I
He faxed to me one of the documents.

Verbs of subcategory 15 that subcategorize <SUBJ , OBJ , OBL_{BNFC}> all have derivational counterparts in this subcategory. As we have discussed previously, semantically, the entity in OBL_{THME} has a theme-like role and that of the OBJ in this construction has a patient-like role. Between the patient that receives directly the action of the verb and the theme that is indirectly affected by the action, the patient has to be interpreted as part of the theme. Thus, all the restrictions that apply between OBL_{THME} and OBJ2 in the previous subcategory of <SUBJ , OBL_{THME} , OBJ , OBJ2> apply here between OBL_{THME} and OBJ. The following is the FI entry for this subcategory. We will also give the f-structure of the sentence 1a.

FI-V-17:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{BNFC} PCASE) =$_c$ BNFC
(↑ OBJ FORM) =$_c$ NONE
]

1a-f. `Ta1 ba3 wen2jian4 chuan2zhen1 le yi2 fen4 gei3 wo3.

[FORM 'chuan2zhen1'
PRED <SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}>
SUBJ [FORM 'ta1']
OBL_{THME} [PFORM 'ba3'
PCASE THME
FORM 'wen2jian4'
DEFINITE +
]
OBJ [CLASS 'fen4'
ADJ { [FORM 'yi1'] }
]
OBL_{BNFC} [PFORM 'gei3'
PCASE BNFC
FORM 'wo3'
]
]

3.3.18 <SUBJ , OBL_{THME} , OBL_{BNFC}>

Examples: chuan2zhen1 'telex, fax', song4 'give', ji4 'mail', xie3 'write', gong1ji3 'provide', song4 'give', jie4 'lend', jiao1 'teach', zu1 'rent', ti2gong1 'provide', jia4 'marry', huan2 'return', jie4shao4 'introduce'

1. a. Ta1 ba3 wen2jian4 chuan2zhen1 le gei3 wo3.
he BA document fax LE to I
He faxed to me the documents.

b. Ta1 ba3 mei4mei4 jie4shao4 le gei3 wo3.
he BA sister introduce LE to I
He introduced his sister to me.

All verbs in this subcategory seem to be derivationally related to the verbs in the previous subcategory. As for verbs subcategorizing <SUBJ , OBJ , OBJ2>, only the ones that require a beneficiary OBJ have counterparts in this current subcategory.

FI-V-18:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{BNFC}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{BNFC} PCASE) =$_c$ BNFC
]

3.3.19 <SUBJ , OBL_{GOAL} , OBL_{THME}>

Examples: jie3shi 'explain', bao4gao4 'report', shuo1ming2 'illustrate', biao3shi4, biao3da2 'express', shuo1 'say', cheng2ching1, biao3ming2 'clarify', tui1xiao1 'sell', kua1zhang1 'exaggerate', xie4lou4 'reveal'

1. a. Ta1 ba3 mi4mi4 dui4 wo3 xie4lou4 le.
he BA secret to I reveal LE
He revealed to me the secret.

b. Ta1 dui4 wo3 ba3 mi4mi4 xie4lou4 le.
he to I BA secret reveal LE
He revealed to me the secret.

First of all, we have observed that in the above two sentences of identical meaning in their c-structures the ordering between the OBL_{THME} PP and the OBL_{GOAL} PP is interchangeable. Since we have already independently established that [ba3 NP] should be considered a subcategorized element, the fact that [gen1/dui4 NP] may appear between [ba3 NP] and the head verb provides another support for considering PP's such as dui4 wo3 'to me' in 1b

above a subcategorized element because of its proximity to the head verb. Also, it is observed that all verbs in the current subcategory all have counterparts in 3.3.7 that subcategorize <SUBJ , OBL_{GOAL} , OBJ>.

FI-V-19:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{GOAL}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{GOAL} PCASE) =$_c$ GOAL
]

3.3.20 <SUBJ , OBL_{GOAL} , OBL_{THME} , OBJ>

Examples: jie3shi4 'explain', bao4gao4 'report', shuo1ming2 'illustrate', biao3shi4, biao3da2 'express', shuo1 'say', cheng2ching1, biao3ming2 'clarify', tui1xiao1 'sell', kua1zhang1 'exaggerate', xie4lou4 'reveal'

1. a. Ta1 ba3 mi4mi4 dui4 wo3 xie4lou4 le hen3duo1.
he BA secret to I reveal LE many
He revealed to me many of the secrets.

b. Ta1 dui4 wo3 ba3 mi4mi4 xie4lou4 le yi1 ge.
he to I BA secret reveal LE one CLS
He revealed to me one of the secrets.

Again, noticeably all the verbs here have derivational counterparts in the previous class 3.3.19 and that the ordering between OBL_{GOAL} and OBL_{THME} in the c-structure is free. The OBJ has to be a quantifying phrase and cannot be a full NP. Moreover, semantically the OBJ has to be interpreted as part of the NP in the ba3 phrase.

FI-V-20:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{GOAL} , OBJ>]
(↑ OBL_{THME} PCASE) $=_c$ THME
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
(↑ OBJ FORM) $=_c$ NONE
]

3.3.21 <SUBJ , OBJ , XCOMP>

3.3.21.1 Regular Type

Examples: jian4yi4 'recommend', pi1zhun3 'approve', ning2yuan4, ning2ke3 'prefer', xuan3 'choose, elect', xuan3ze2 'select', xi3huan1 'like', gu3li4 'encourage', tao3yan4 'dislike', yao4, xiang3yao4 'want', jiao1 'teach'

1. a. $Ta1_i$ pi1zhun3 $ta1_{j/*i}$ qu4 Mei3guo2.
he approve he go U.S.
He approves for him to go to the U.S.

b. $Ta1_i$ xi3huan1 $ta1_{j/*i}$ de2 guan4jun1.
he like he get championship
He likes for him to get the championship.

a'*Ta1 pi1zhun3 ta1 qu4 le Mei3guo2.
he approve he go LE U.S.
He has approved for him to go to the U.S.

b'*Ta1 xi3huan1 ta1 hui4 de2 guan4jun1.
he like he will get championship
He likes for him to get the championship.

These verbs appear in a syntactic construction often known as the "pivotal" construction in the literature of Chinese linguistics, a term first used by Chao (1968) and popularized by Li and Thompson (1981). They subcategorize a SUBJ, an OBJ, and an XCOMP. The XCOMP is a non-finite clause without

an overt subject, and in a pivotal construction, the OBJ of the matrix verb functionally controls the SUBJ of the XCOMP. In other words, the matrix OBJ unifies with, and thus functionally becomes, the SUBJ of the XCOMP.

FI-V-21:

```
[ CAT V
  FS    [ PRED <SUBJ , OBJ , XCOMP> ]
]
```

FI-V-21-1:

```
[ (↑ OBJ) = (↑ XCOMP SUBJ)
  (↑ XCOMP ASPECT) =c NONE
  (↑ XCOMP MODAL) =c NONE
]
```

ning2yuan4:

```
[ FS [ FORM 'ning2yuan4' ]
  FI-V-21
  FI-V-21-1
]
```

First of all, since in Chinese there is no systematic marking of finiteness nor subject-verb agreement, it is not always clear whether the embedded clause is indeed a finite SCOMP or if it is a non-finite XCOMP with its preceding NP as the OBJ also subcategorized by the matrix verb. Take 1a above for example. It is important to determine which of the following two f-structures depicts the correct subcategorization.

1a-f1.

```
[ FORM 'pi1zhun3'                          `allow
  PRED     <SUBJ , SCOMP>
  SUBJ     [ FORM 'ta1' ]
  SCOMP [ FORM 'qu4'
            PRED   <SUBJ , OBJ>
            SUBJ   [ FORM 'ta1' ]
            OBJ    [ FORM 'mei3guo2' ] ]
]
```

1a-f2.

```
[ FORM 'pi1zhun3'
  PRED      <SUBJ , OBJ , XCOMP>
  SUBJ      [ FORM 'ta1' ]
  OBJ       [ FORM 'ta1' ]
  XCOMP   [ FORM 'qu4'
              PRED   <SUBJ , OBJ>
              SUBJ   [---]
              OBJ    [ FORM 'mei3guo2' ]
            ]
]
```

The Anaphoric Binding Principles in vLFG for Chinese we have presented in 3.3.11 of this chapter provide a solution. Recall that a non-reflexive pronoun cannot have an antecedent within the minimal f-structure containing the pronoun and a SUBJ function. The fact that in 1a-b the second pronoun ta1 's/he' cannot be co-referential with the matrix subject clearly indicates that they must be contained with the minimal f-structure that contains a PRED feature and both the matrix SUBJ and the second pronoun ta1. In other words, in no way can the second pronoun ta1 be considered the uncontrolled SUBJ of the embedded clause, so we know that 1a-f1 cannot be the correct f-structure. Rather, the second pronoun ta1 should be considered an OBJ subcategorized by the matrix verb. Furthermore, the fact that the embedded clause can take neither aspect nor modality indicates the clause is a non-finite XCOMP. If we compare 1a-b with the following sentences, using the

Anaphoric Binding Principles, we can determine that the embedded clauses in 1c-d' are finite SCOMPs.

1. c. Ta1$_i$ shuo1 ta1$_{j/i}$ qu4 Mei3guo2.
he say he go U.S.
He says that he goes to the U.S.

d. Ta1$_i$ xi1wang4 ta1$_{j/i}$ de2 guan4jun1.
he hope he get championship
He hopes that he gets the championship.

c' Ta1 shuo1 ta1 qu4 le Mei3guo2.
he say he go LE U.S.
He says that he has gone to the U.S.

d' Ta1 xi1wang4 ta1 hui4 de2 guan4jun1.
he hope he will get championship
He hopes that he will get the championship.

The fact that the second pronoun ta1 may be co-referential with the matrix SUBJ shows that they are not in the same f-structure that contains them both. The second pronoun ta1 thus is considered the uncontrolled SUBJ of the embedded clause. Furthermore, the potential occurrence of aspect particles and modal verbs in the embedded clauses of xi1wang4 'hope' and shuo1 'say' in the above examples indicates the finiteness of the clauses.

3.3.21.2 Clause Union Verbs

Examples: shi3, ling4 'make', yin3, ling3, yin3ling3, dai4ling3 'lead', qing3 'invite, treat', qiang2po4, bi1po4, po4shi3, bi1 'force', rang4 'let', pei2 'accompany', ya1 'escort', jiao4, ming4ling4 'order', yin3you4 'induce, seduce', shou4chuan2 'authorize', yuan2liang4 'forgive'

2. a. Ta1 qing3 wo3 chi1 guo4 yi1 dun4 fan4.
he treat I eat GUO4 one CLS meal
He has treated me a meal before.

b. Ta1 dai4ling3 ta1men chu1 le Ai1ji2.
he lead they leave LE Egypt
He led them to leave Egypt.

c. Ta1 qiang2po4 ta1men chi1 zhe4 ku3gua1.
he force they eat ZHE bitter-melon
He is forcing them to eat bitter melons.

Similar to verbs of the subcategory 3.3.12.4, verbs here exhibit the characteristic of "clause union." The difference between verbs here and the previous subcategory 3.3.21.1 is therefore that the non-finite embedded clause, XCOMP, here may contain aspect markers, as shown in the above sentences 2a-c. The matrix verb and the embedded verb fuse together and it is the embedded verb that takes the aspect marking, rather than the matrix verb. All the verbs in this subcategory are causative verbs with a pivotal construction.

FI-V-21-2:
[(↑ OBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODAL) = NONE
(↑ ASPECT) $=_c$ NONE
(↑ XCOMP ASPECT) = (↑ ASPECT)
]

bi1: `force
[FS [FORM 'bi1']
FI-V-21
FI-V-21-2
]

Again, we will compare a clause union verb from here, qiang2po4 'force', with a verb gu3li4 'encourage' from the previous subcategory 3.3.21.1.

2. d. Ta1 qiang2po4 wo3 qiang3 guo4 yin2hang2.
he force I rob GUO bank
He has forced me to rob a bank before.

e. Ta1 mei2 qiang2po4 wo3 qiang3 guo4 yin2hang2.
he not force I rob GUO bank
He has never forced me to rob a bank before.

d' Ta1 gu3li4 guo4 wo3 qiang3 yin2hang2.
he encourage GUO I rob bank
He has encouraged me to rob a bank before.

e' Ta1 mei2 gu3li4 guo4 wo3 qiang3 yin2hang2.
he not encourage GUO I rob bank
He has never encouraged me to rob a bank before.

Notice that although in 2d and 2d' the actions of "forcing" and "encouraging" have taken place, whether the action of "I robbing a bank" has or not is not clear. Therefore, the two sentences below are equally sound.

2.f. Ta1 qiang2po4 wo3 qiang3 guo4 yin2hang2, ke3shi4 wo3 mei2
he force I rob GUO bank but I not
qiang3.
rob
He has forced me to rob a bank before, but I never did.

f' Ta1 gu3li4 guo4 wo3 qiang3 yin2hang2, ke3shi4 wo3 mei2
he encourage GUO I rob bank but I not
qiang3.
rob
He has encouraged me to rob a bank before, but I never did.

The syntactic difference between gu3li4 'encourage' and qiang2po4 'force' is that the aspect marker guo4 has to appear with the matrix verb in 2d' but with the verb in XCOMP, qiang3 'rob', in 2d. Moreover, since 2d and 2d' have the same negative form, as shown in 2e and 2e', the scope of the aspect marker in 2d should be equivalent to that of the aspect marker in 2d'. Consequently, this indicates that the aspect marker of 2d, though in the embedded XCOMP, covers the aspect of matrix verb qiang2po4 'force' as well.

3.3.21.3 "Promise" Verbs

Examples: da1ying4 'promise', ying4xu3 'promise', dai4biao3 'represent'

3. a. Ta1 da1ying4 wo3 ming2tian1 huan2 wo3 qian2.
he promise I tomorrow return I money
He promises me to return the money to me tomorrow.

b. Ta1 dai4biao3 wo3 chu1xi2 le hui4yi4.
he represent I attend LE meeting
He represented me to attend the meeting.

The important difference between verbs here and the two previous subcategories is that the functional controller of XCOMP's SUBJ has to be the matrix SUBJ, not OBJ. In other words, verbs of this subcategory, though they subcategorize <SUBJ , OBJ , XCOMP>, do not appear in the pivotal construction where the matrix OBJ controls the XCOMP's SUBJ. We suspect the number of verbs that require this control relation is very limited. Even so, we find that da1ying4 'promise' and dai4biao3 'represent' manifest different behavior in terms of the non-finiteness of the XCOMP they subcategorize: while da1ying4's XCOMP is strictly non-finite, dai4biao3 'represent' shows the characteristic of "clause union verbs" in that its XCOMP may take aspect particles. Therefore, while da1ying4 'promise' may take aspect markers, dai4biao3 'represent' does not.

3. c. Ta1 da1ying4 guo4 wo3 ming2tian1 huan2 wo3 qian2.
he promise GUO I tomorrow return I money
He has promised me before to return the money to me the next day.

d.*Ta1 dai4biao3 guo4 wo3 chu1xi2 hui4yi4.
he represent GUO I attend meeting
He has represented me before to attend the meeting.

e. Ta1 dai4biao3 wo3 chu1xi2 guo4 hui4yi4.
he represent I attend GUO meeting
He has represented me before to attend the meeting.

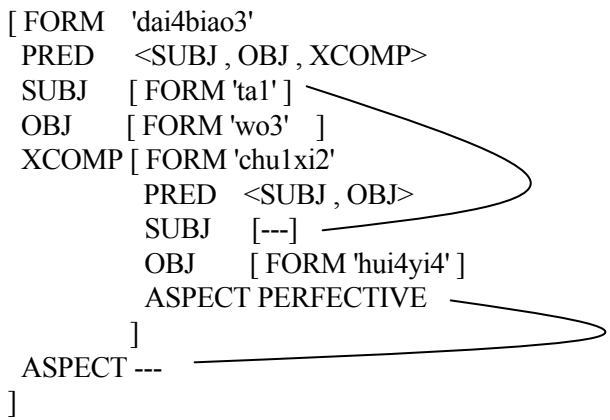
Due to the limited members in this subcategory, we will account for the difference in the individual lexical entry, rather than in the FI entry of the subcategory.

FI-V-21-3:
[(↑ SUBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODALITY) $=_c$ NONE
]

da1ying4: `promise
[FS [FORM 'da2ying4']
FI-V-21
FI-V-21-3
(↑ XCOMP ASPECT) $=_c$ NONE
]

dai4biao3: `represent
[FS [FORM 'dai4biao3']
FI-V-21
FI-V-21-3
(↑ ASPECT) $=_c$ NONE
(↑ XCOMP ASPECT) = (↑ ASPECT)
]

4b-f.



3.3.21.4 Tough Construction

Examples: zhi2de2 'worth, worthwhile', rang4, jiao4 'let', lun2dao4, gui1 'alternate'

4. a. Ta1 zhi2de2 wo3men bang1zhu4.
he worth we help
He is worth us helping.

b. Che1zi jin1tian1 lun2dao4 le ni3 xi3.
car today alternate LE you wash
Today it is your turn to wash the car.

c. Ta1 zhi2de2 wo3men she4fa3 bang1zhu4.
he worth we try help
He is worth us trying to help.

d. Na4 ge gu4shi4 lun2dao4 le ni3 gao4su4 ta1men.
that CLS story alternate LE you tell they
It is your turn to tell them the story.

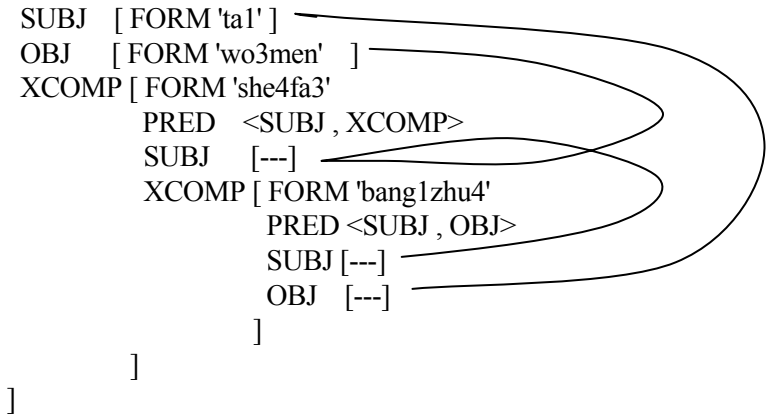
Verbs here do appear in the so-called "pivotal" construction, although like verbs of the "tough" construction of subcategory 3.3.12.2, the embedded non-finite XCOMP not only has an unfulfilled SUBJ, but also an unfulfilled OBJ, which is potentially unbounded. While it is the matrix OBJ that functionally controls XCOMP's SUBJ, the matrix SUBJ is the functional controller of the potentially unbounded OBJ. These characteristics of verbs in this subcategory are captured in the following FI entry. We will also provide the f-structure of 4c with control relations fulfilled as an example.

FI-V-21-4:
[(↑ OBJ) = (↑ XCOMP SUBJ)
(↑ SUBJ) = (↑ XCOMP+ {OBJ OBJ2})
(↑ XCOMP MODALITY) $=_c$ NONE
(↑ XCOMP ASPECT) $=_c$ NONE
]

zhi2de2: `worth
[FS [FORM 'zhi2de2']
FI-V-21
FI-V-21-4
FI-V-STATE
]

lun2dao4: `alternate
[FS [FORM 'lun2dao4']
FI-V-21
FI-V-21-4
]

4c-f. `Ta1 zhi2de2 wo3men she4fa3 bang1zhu4.



3.3.21.5 Bei4

1. a. Lao3shu3 bei4 yao3 le.
mouse BEI bite LE
The mouse was bitten.

b. Lao3shu3 bei4 mao1 yao3 le.
mouse BEI cat bite LE
The mouse was bitten by the cat.

c. Lao3shu3 bei4 mao1 yao3diao4 le yi3ba1.
mouse BEI cat bite off LE tail
The mouse had its tail bitten off by the cat.

d. Lao3shu3 bei4 mao1 ba3 yi3ba1 yao3diao4 le.
mouse BEI cat BA tail bite off LE
The mouse had its tail bitten off by the cat.

e. Lao3shu3 bei4 mao1 ba3 yi3ba1 yao3diao4 le hao3duo1
mouse BEI cat BA tail bite off LE many
mao2.
hair
The mouse had many hairs of its tail bitten off by the cat.

f. Mao1 bei4 lao3shu3 tao2 le.
cat BEI mouse escape LE
The cat had the mouse escaped on him.

The bei4 construction, which is commonly considered the passive construction in Chinese, is one of the most argued about syntactic topics in Chinese linguistics. We will make yet another attempt to analyze these various sentence types and to provide a unified account for the lexical item bei4. The above sentences in 1 are examples of the various types of bei4 sentences which we will discuss and try to account for.

We shall also discuss and compare previous accounts of bei4 sentences and argue for our analysis that syntactically bei4 should be treated as a verb and that it only occurs in a pivotal construction which requires an object and a non-finite VP complement. We shall conduct the discussion in relatively theory-independent terms but formulate our preferred analysis of bei4 in our vLFG formalism. Note that for certain native speakers bei4 in all the above sentences may be replaced with gei3. Thus our final analysis for the lexical entry of bei4 should apply to gei3 as well.

3.3.21.5.1 Existing Analyses for Chinese bei4

In the literature of Chinese linguistics, bei4 most commonly has been considered a function word (e.g., Chao 1968, Chang 1977, Hou 1979, Li and Thompson 1981, Chu 1984, and Her 1985-6), in the sense of the traditional distinction between "function words" versus "content word" in Chinese linguistics. However, in recent years there has been a radically different line of analysis which has been largely overlooked that treats bei4 as a content word, more specifically as a verb with its own predicate argument structure. There are two such analyses that we are aware of and both are formulated in

LFG (Ma 1985, Tan 1987). We shall first illustrate the different existing analyses of bei4 in the following figure.

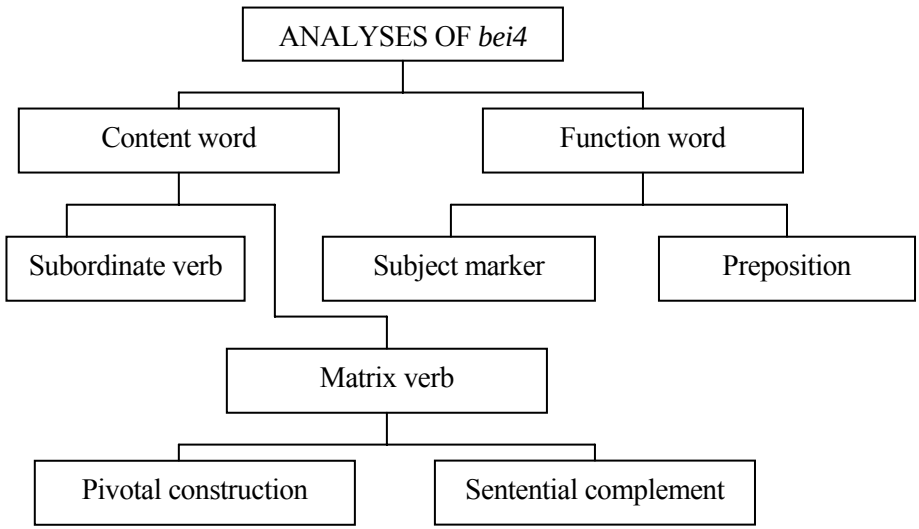


Figure 3.2 Possible Analyses for Chinese bei4

As shown in the diagram, within the analysis where bei4 is a function word, some treat it as a subject marker (e.g., Li and Thompson 1981), and yet some insist that it is a preposition (e.g., Chao 1968, Hou 1979, Chu 1984, Her 1985-6). Her's analysis is formulated in Lexicase, and Hou's in Relational Grammar. There are also some transformational accounts (e.g., Li 1972, Huang 1966, Teng 1977) while other discussions are not conducted in any particular theoretical framework. Actually Teng's analysis of bei4 is unique in that he treats it as a passive "particle," not an agent marker, subject marker, preposition, or verb, which transformationally introduced when subject and object in the deep structure are inverted. Different possible accounts also exist regarding what status bei4, when treated as a verb, has in a sentence and what predicate arguments it is required to take. Ma (1985) first proposes that bei4 has a pivotal construction, i.e., it takes a direct object and a verb phrase as arguments and its direct object functionally controls the subject of the predicated verb phrase. Tan (1987), disagreeing with Ma's analysis, argues

instead that bei4 as a verb only takes a noun phrase as its object and the [bei4 NP] phrase is subordinate to the matrix clause. Another logical alternative to their proposals of course is that bei4 is the matrix verb as Ma suggests but what follows it is a sentential complement. To put these three alternatives in LFG terms, the last alternative would have bei4 subcategorize an SCOMP, which nobody has yet argued for. Whereas, Ma proposes that bei4 subcategorizes an OBJ and an XCOMP and bei4 is the matrix verb; Tan analyzes bei4 as subcategorizing an OBJ and the bei4 phrase only functions as an adjunct of the matrix clause.

3.3.21.5.2 Dismissal of bei4 as a Subject Marker

First of all, we would like to examine and hopefully convincingly dismiss the analysis that bei4 is without lexical meaning and is merely a case marker of the subject. Within this analysis sentence 1b would have the structure [O bei4 S V] and sentence 1a would have to be considered subjectless. The dilemma is that it is quite peculiar that bei4 as a subject marker would mark nothing in 1a. Also, if sentence 1a is considered an [O V] sentence and lb an [O S V] sentence, are we to say that the corresponding English sentences, "The mouse was bitten" and "The mouse was bitten by the cat" are also of [O V] and [O V S] respectively? Such a position of course contradicts the morphology of case marking in English and makes the study of word order typology entirely meaningless. Thus, if lao3shu3 is to be recognized as the subject in all the sentences in 1, bei4 simply cannot be a subject marker. Also, as Chu (1984:140) has demonstrated, the analysis that bei4 is a pure grammatical case marker contradicts the fact that bei4 has the semantic content of "adversity" and that its presence or absence affects the grammaticality of a sentence. Thus, Chu rejects the notion that bei4 is a "subject marker," and following the same line of argument he also rejects the notion that ba3 is an "object marker." The above argument also applies to reject the transformational analysis (e.g., Li 1972, Huang 1966, Teng 1977) where subject and object in the deep structure are inverted to form the surface bei4 structure.

If one retreats from this position and claims that bei4 marks the semantic agent instead of the syntactic subject, then in effect one still leaves open the question what syntactic category bei4 belongs to. Li and Thompson (1974,

1981), again, due to their informal and subjective 'functional' approach to the Chinese grammar, along with many traditional grammarians, often fail to distinguish clearly between syntax and semantics. Consequently, it is not clear at all whether they recognize bei4 as a morphological case marker marking the nominative case of subject or as a preposition that has the semantic function of marking the agent. As we have pointed out earlier, the confusion of syntactic category and semantic function, for example, in this case always taking the semantic agent to be the syntactic subject, is unfortunately common within the study of word order typology of the Greenbergian tradition (Her 1985-6).

The study of Mandarin word order of Sun and Givon (1985) is relevant to our discussion here in two ways: first, their treatment of bei4 as an "agent marker" is another example of the confusion of semantics and syntax; second, their reformulation of Li and Thompson's finding concerning Mandarin word order and its semantic distribution of definiteness may provide another indication that bei4 should not be considered "subject marker" nor a preposition and that it should be treated as a verb. Li and Thompson (1975:171) observe that nouns preceding the verb tend to be definite, while those following the verb tend to be indefinite. Sun and Givon claim that this tendency (referred to as Tendency A in the following quote) is subject to several refinements they put forth, and one of the refinements is specifically for the element bei4: "Refinement 3: The noun following bei (the agent-marker in the passive construction), although preverbal, is immune to Tendency A" (Sun and Givon 1985:344). However, if we simply treat bei4 as a verb, this exception will no longer exist, for now the noun following bei4 is postverbal and thus tends to be definite.

3.3.21.5.3 Dismissal of bei4 as a Preposition

By far the majority of Chinese linguists consider bei4 a preposition, occasionally termed inappropriately as "coverb" (e.g., Li 1981, Chang 1977). Within this analysis, bei4 sentences would have the following constituent structures:

1a-c. [$_{S}$ Lao3shu3 [$_{VP}$ [$_{PP}$ [$_{P}$ bei4 $_{P}$] $_{PP}$] yao3 $_{VP}$] le $_{S}$]
1b-c. [$_{S}$ Lao3shu3 [$_{VP}$ [$_{PP}$ [$_{P}$ bei4 $_{P}$] mao1 $_{PP}$] yao3 $_{VP}$] le $_{S}$]

The popularity of this analysis is to some extent due to the likening of the structure of Chinese bei4 sentences to that of their English passive translations; bei4 is considered the equivalent of preposition "by" in the English passive construction. Such an analysis, unlike the previous one, is basically without flaws at the theoretical level of linguistics; however, it misses some important generalizations in Mandarin Chinese and, as we will demonstrate, makes wrong predictions. First, as both Ma (1986) and Tan (1987) have objected, when bei4 is used in "agentless" sentences such as 1a, it constitutes a violation to the universal characteristics that a preposition always takes an immediately following NP-like element as its object. Furthermore, this analysis poses a problem for analyzing the verb in a bei4 sentence and its bei4-less counterpart. The solution proposed in Her (1985-6) that treats bei4-V as a morphologically derived lexical item although getting around the first problem still faces the second problem.

2. a. Lao3shu3 bei4 yao3 le.
mouse BEI bite LE
The mouse was bitten.

a' Lao3shu3 yao3 le.
mouse bite LE
The mouse has bitten. OR,
The mouse was bitten.

b. Ji1 bei4 ta1men chi1 le.
chicken BEI they eat LE
The chicken was eaten by them.

b' Ji1 chi1 le.
chicken eat LE
The chicken was eaten. OR,
The chicken ate.

c. Lao3shu3 bei4 mao1 ba3 yi3ba1 yao3diao4 le.
mouse BEI cat BA tail bite off LE
The mouse had its tail bitten off by the cat.

c' Lao3shu3 ba3 yi3ba1 yao3diao4 le.
mouse BA tail bite off LE
The mouse bit off the tail.

d. Mao1 bei4 lao3shu3 tao2 le.
cat BEI mouse escape LE
The cat had the mouse escape him.

d' Mao1 tao2 le.
cat escape LE
The cat escaped.

e. Wo3 bei4 ta1men ba3 wo3 pian4 le.
I BEI they BA I cheat LE
I was cheated by them.

e'*Wo3 ba3 wo3 pian4 le.
I BA I cheat le
I cheated myself.

e" Wo3 ba3 zi4ji3 pian4 le.
I BA self cheat le
I cheated myself.

Sentences 2a and 2b are unambiguous as we can see from their translations in English. However, sentences 2a' and 2b', without the bei4 phrase, each has two readings, again as the translations clearly indicate. There are two possible accounts for this ambiguity. One is to say that transitive verbs in Chinese are lexically ambiguous, i.e., they can either be active or passive unless specified by certain elements such as bei4. Thus, chi1 'eat' and yao3 'bite' can be both passive and active and therefore the ambiguity in 2a' and 2b', but 2a and 2b with the specification of passive voice by the bei4 phrase have only the passive reading. Within this analysis there are serious difficulties in accounting for sentences 2c' and 2d'. First of all, 2c has only the passive reading due to the presence of the bei4 phrase, but why is 2c' not ambiguous with two readings as this analysis predicts? One might argue that 2c' only has the active reading because of the presence of a ba3 phrase but such a

statement is self-contradictory since in 2c both bei4 and ba3 are present at the same level.

Still, sentences 2d and 2d' present a different problem. Tao2 'escape' is an intransitive verb in Chinese. The universal characteristic that only transitive verbs may be passivized and examples like 2d have led some linguists to the conclusion that bei4 sentences in Chinese are not of genuine passive construction (e.g., Ma 1985, Keenan 1985 and Tan 1987). Therefore, regarding 2d, this present analysis cannot stand without posing a violation of this universal characteristic of the passive construction. The fact that tao2 'escape' in both 2d and 2d' has only one reading, the active reading, also suggests that it is the same verb in both sentences.

We now examine the second possible account for the ambiguous 2a' and 2b'. Some linguists have considered sentences like 2a' of structural ambiguity (e.g., Chu 1984, Sun and Givon 1985), one of SV structure, the other OV. In other words, the active meaning comes from the structure where lao3shu3 'mouse' and ji1 'chicken' are subjects of the verbs, and the other reading comes from the structure where they are actually objects and the sentences are subjectless. However, this account still leaves open the question of the status of the verb in a bei4 sentence. Are we to say that in 2a lao3shu3 'mouse' is also the object? How about ji1 'chicken' in 2b? If we consider them objects, then we have to say that either all bei4 sentences are subjectless, although syntactically there is nothing incomplete, or that the bei4 phrase is always the subject; such positions are of course unsubstantiated. The dilemma is that even if we take another position and claim that the presence of a bei4 phrase indicates the passive voice of the verb and that lao3shu3 'mouse' and ji1 'chicken' in 2a and 2b are the subjects, we still have the problems of accounting for 2c' and 2d'. First, we have to say that in 2d tao2 'escape' is a passivized intransitive verb and thus constitutes an exception to the universal passive construction. Secondly, we have to say that in 2c, as well as 1c, the verb yao3diao4 'bitten off', although passivized, is still transitive with yi3ba1 'tail' as its overt object. Again such behavior is extremely uncharacteristic of verbs already passivized, which universally no longer have their transitivity.

Finally, regardless of how we analyze the main verb in a bei4 sentence, if bei4 is considered a preposition and thus the bei4 phrase a prepositional phrase, there is no accounting for the fact that while 2e' is unacceptable, 2e is perfectly good. Since 2e, just like 2e' and 2e", is mono-clausal, the second

pronoun wo3 'I' in 2e, which has the first wo3, the subject, as its antecedent, must be a reflexive pronoun for the sentence to be acceptable, such is the case in the acceptable 2e". This bei4-preposition analysis therefore wrongfully predicts 2e to be ungrammatical.

The more perceptive reader could probably have detected that this last argument against the bei4-preposition analysis most convincingly leads to the hypothesis that the difference in acceptability between 2e and 2e' is that while 2e' is mono-clausal and therefore a pronoun with its antecedent in the same clause has to be reflexive, 2e is not mono-clausal and actually contains two clauses. The second pronoun wo3 in 2e thus does not have to be reflexive for the sentence to be acceptable because the second pronoun wo3 and its antecedent are not in the same minimal clause. Such a hypothesis necessarily entails the interpretation that bei4 is a verb with its own predicate structure.

A more specific piece of evidence for bei4 as a verb is available under LFG's premise that oblique functions, marked by prepositions, are semantically restricted. Thus, if bei4 is a preposition, then indeed semantically its following NP should be restricted to the agent role. Recall that Sun and Givon (1985), along with others, consider bei4 an "agent marker." However, this is certainly not always the case because 1) syntactically it is not even required for there to be an NP following bei4, and 2) the NP following bei4 is not semantically restricted to be an agent role. Rather, the semantic role associated with the NP following bei4 coincides with the role associated with the subject of the corresponding active counterpart, whether it is an agent or otherwise. This second point is quite adequately illustrated by Teng (1977:33-34).

Before we proceed to discuss and compare the possible analyses where bei4 is treated as a verb, it is important to point out that our discussion so far has also strongly supported the position that bei4 sentences should not be considered passive sentences and bei4 cannot be considered a passive marker. To be more specific, bei4 sentences are not passive in the sense that they cannot be accommodated by any of the universal generalizations proposed by Chomsky (1957), Chomsky (1981, 1982), or Bresnan (1982). Two reasons are crucial here: 1) bei4 sentences may contain intransitive verbs which cannot be passivized (e.g., 1f, 2) bei4 sentences may contain verbs still transitive with overt objects (e.g., 1c-e).

Thus, the denotation of Chinese bei4 sentences in fact is not limited to just the passive voice; it is rather similar to the Japanese indirect passive (Siewierska 1984). Also, similar to the standard -are- passive in Japanese (Keenan 1985), the Mandarin bei4 is often interpreted as negatively affecting the subject. We shall therefore suggest the term "adversative affective voice" to distinguish bei4 from the commonly-known "passive voice" in languages like English.

3.3.21.5.4 Bei4 as a Verb

Keenan (1985:260), in a typological survey of the passive construction in the world's languages, classifies Mandarin in a language group where "The passive auxiliary is a verb of experiencing," e.g., suffer. He also cautions that "their analysis as passives is in fact not obvious," confirming our position that bei4 sentences are not passives. However, unlike passive verbs in many other Southeast Asian languages in this group, bei4 can no longer occur as the main verb in a simple sentence. We now will examine the possible analyses where bei4 is treated as a verb with its own subcategorization requirements.

3.3.21.5.4.1. Problems with the subordinate status of bei4

We will now first examine the proposal that bei4(-NP) is a modifying subordinate clause (Tan 1987). In more formal LFG terms she proposes that, first of all, bei4 has the following complement structure. (Note that since both Tan (1987) and Ma (1985) follow the conventional LFG formalism, we maintain the notation they use.)

bei4 V
(↑ PRED) = 'bei4 <(SUBJ), (OBJ)/(SCOMP)>'

Note that Tan uses / to indicate a disjunction. First of all, she analyzes bei4 as subcategorizing for either <SUBJ , OBJ> or <SUBJ , SCOMP>. Secondly, Tan asserts that the bei4 clause is always an adjunct of the matrix clause and it never functions as the matrix verb. Both assertions seem rather ad hoc. A typical bei4 sentence such as 1b will therefore have the following c-structure and f-structure:

1b-c1.
[$_S$ Lao3shu3 [$_{VP}$ [$_{VP}$ [$_V$ bei4 $_V$] mao1 $_{VP}$] yao3 $_{VP}$] le $_S$]

1b-f1.

SUBJ	[PRED	'lao3shu3']
XADJ	SUBJ	[----]
	OBJ	[PRED 'mao1']
	PRED	'bei4 <(SUBJ) (OBJ)>'
ASPECT	PERFECT	
PRED	'yao3 <(SUBJ)>'	

Tan's proposal is to a certain degree in reaction to Ma (1985) where bei4 is treated as the main verb that appears in a pivotal construction. We shall answer Tan's objections to Ma (1985) in the next section. The only piece of direct evidence that Tan provides to support her position is the dispensibility of bei4 phrases as adjuncts (Tan 1987:7):

> The meaning of the sentence with the adjunct entails its counterpart without the adjunct. This is true of Chinese bei4 sentences: the bei4 phrase disambiguates the corresponding sentences without this adjunct, which usually have both the active and non-active readings.

Sentences 1d-1e, 1f, and 2e-2e" however provide three types of counter-examples to this argument of hers. First, in the bei4-less corresponding sentences of 1d and 1e, there is no ambiguity, be it structural or lexical, and the meaning of 2d or 2e is certainly not "the intersection of the matrix predicate meaning and the adjunct meaning" (Tan 1987:7). The bei4-less corresponding sentence of 1f provides similar evidence against her position. Furthermore, if we follow Tan's proposal, then we have to also impose a lexical ambiguity on intransitive verbs such as tao2 'escape'. Thus, we have to increase the complexity of the analysis of such verbs. Therefore,

an account for sentences like 1f without such a complication, as we will demonstrate in the next section, should be preferred. Furthermore, Tan's account will also make the wrong prediction about the acceptability of sentence 2e. According to her analysis, the second wo3 'I' still has its antecedent in the same minimal clause since the bei4 phrase is merely an adjunct; therefore, the second wo3 'I' will have to be reflexive for the sentence to be acceptable. In 2e the second wo3 'I' is not reflexive; thus Tan's account wrongfully predicts that 2e is unacceptable while it is perfectly good.

The most ad hoc aspect of Tan's analysis is that the bei4 clause functions exclusively as an adjunct, never as the matrix clause. We are not aware of any other verb, in Chinese or other languages, that exhibits such peculiar behavior. Tan does not provide us with examples of other verbs in Chinese or other languages. Granted, any account for bei4 will most definitely contain peculiarities since bei4 is certainly a peculiar element which has generated much disagreement. However, a more generalized analysis will have to be preferred over one that resorts to such ad hoc features. Note also that Tan does not give an account of bei4 sentences that are "agentless" such as 1a. It is therefore unclear how this type of sentence is accounted for within her analysis.

3.3.21.5.4.2. Advantages of bei4 as the matrix verb

Ma (1985) proposes that bei4 be the matrix verb and that its object control the subject of its open VP complement, and therefore the lexical form for bei4 is the following:

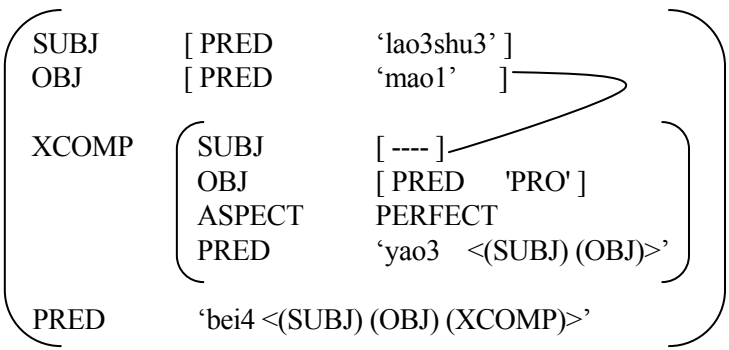
bei4 V
(↑ PRED) = 'bei4 <(SUBJ) (OBJ) (XCOMP)>'
(↑ OBJ) = (↑ XCOMP SUBJ)

Within such an analysis, a typical bei4 sentence such as 1b will have the following c-structure and f-structure:

1b-c2.

[$_S$ Lao3shu3 [$_{VP}$ [$_V$ bei4 $_V$] [$_{NP}$ mao1 $_{NP}$] [$_{VP}$ yao3 $_{VP}$] $_{VP}$] le $_S$]

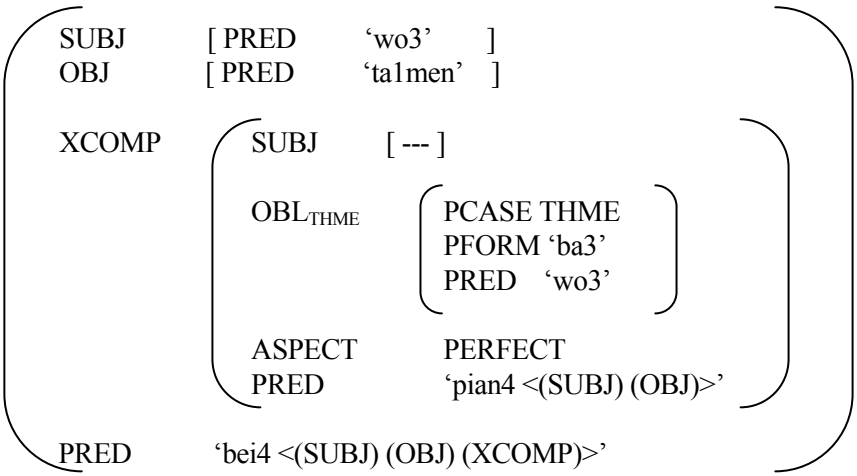
1a-f2.



We will now show that such an account of bei4 sentences avoids all the above-mentioned problems associated with other analyses. The first advantage of this analysis is that it posits a unified account for the embedded verbs in bei4 sentences, namely that all verbs in bei4 sentences are the same as those in their bei4-less corresponding sentences. This analysis therefore does not resort to the solution that in Chinese transitive verbs are ambiguous in their active and non-active voices. Accordingly, in 2b yao3 'bite' is still active with mao1 'cat' as its subject; similarly in 1f tao2 'escape' is still active with lao3shu3 'mouse' as its subject. The affective voice of the entire sentence is due to the presence of the matrix verb bei4; however, within the clause of the XCOMP the voice is active. This is most evident if we compare 2d-2e and their bei4-less counterparts which do not allow any non-active interpretation.

Secondly, this analysis correctly predicts the acceptability of 2e which has the following f-structure:

2e-f.



Clearly, the wo3 'I' in the ba3 phrase and its antecedent are in two separate clauses and therefore the second wo3 'I' does not have to be reflexive. We thus correctly predict the acceptability of this sentence. Furthermore, one may refer to Ma (1985) for more evidence from adverbial modification and other types of reflexive binding supporting this analysis.

3.3.21.5.4.3. A more complete account for bei4

Although our analysis basically follows Ma's, her account for bei4 sentences does not seem to be complete: first, it does not account for "agentless" bei4 sentences such as 1a, and second, it does not account for the fact that the XCOMP in a typical bei4 sentence, such as 1a-b, aside from the missing SUBJ, also has a missing OBJ or OBJ2, which should be identified with the matrix SUBJ, and third, it does not account for the fact that the missing OBJ or OBJ2 may be of a long distance control relation with the matrix SUBJ. Such a complement structure and control relations are not unique to bei4; they apply also to verbs such as zhi2de2 'worth', shou4 and ai1 'receive', gui1, gai1, and lun2dao4 'alternate' of the previous subcategory 21C.

3. a. Pan2zi gui1/gai1/lun2dao4 ni3 xi3.
dish alternate you wash
It's your turn to wash the dishes.

b. Ta1 zhi2de2 wo3 ai4.
he worth I love
He is worth me loving.

However, note also that in bei4 sentences 1c-f, besides the missing SUBJ, the XCOMP is otherwise complete. Thus, bei4 does not necessarily have an unbounded gap to fill in its XCOMP. In this respect bei4 behaves similarly as the verb rong2yi4 'easy' in Mandarin Chinese. Let's compare the following sentences.

1.d. Ta1 hen3 rong2yi4 gan3mao4. `(in 3.3.12.1)
he very easy catch cold
It is very easy for him to catch cold.

1.f. Mao1 bei4 lao3shu3 tao2 le. `(in 3.3.21.5)
cat BEI mouse escape LE
The cat had the mouse escaped on him.

2.a. Gan3mao4 hen3 rong2yi4 yi1zhi4. `(in 3.3.12.2)
cold very easy cure
Colds are easy to cure.

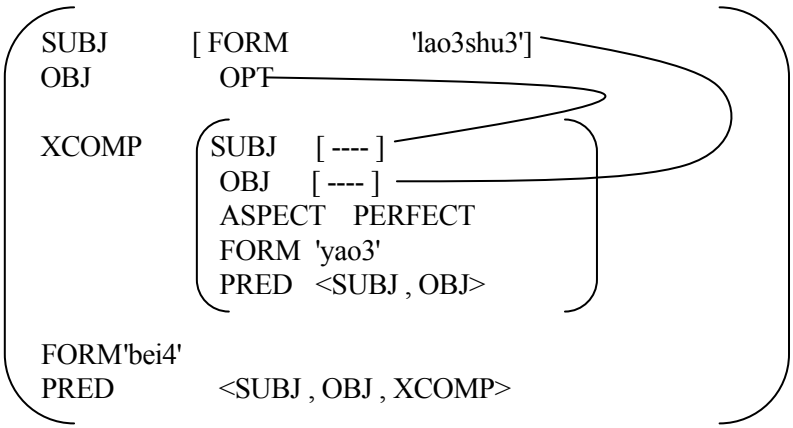
1.a. Lao3shu3 bei4 yao3 le. `(in 3.3.21.5)
mouse BEI bite LE
The mouse was bitten.

Following the same treatment of rong2yi4 'easy', we will have two separate lexical forms for bei4 to account for the two constructions, and further we will supplement Ma's analysis with extended functional expressions to account for the control relations and the fact that bei4 does not take any aspect markers or modal verbs. The entire lexical entry of bei4 is given below in our vLFG format.

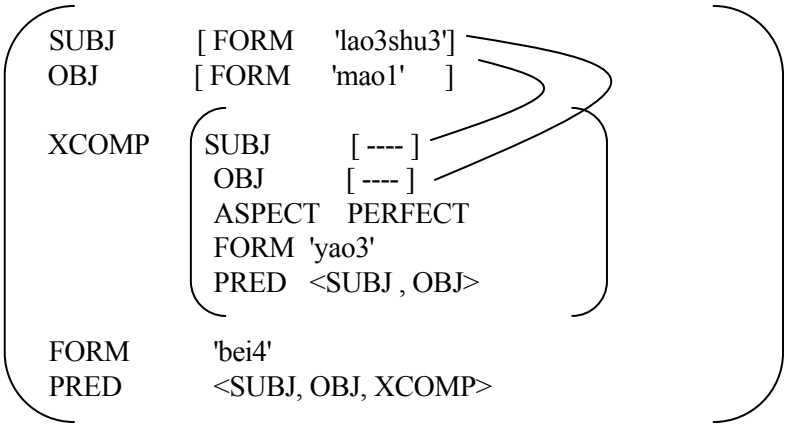
```
bei4:
{       [ CAT   V                                       `for 1a-b
          FS    [ FORM 'bei4'
                  OBJ OPT
                ]
          FI-V-21
          FI-V-21-2
          (↑ SUBJ) = (↑ XCOMP+ {OBJ OBJ2})

        ]
        [ CAT   V                                       `for 1c-f
          FS    [ FORM 'bei4'
                  OBJ OPT
                ]
          FI-V-21
          FI-V-21-2
        ]
}
```

Like causative verbs in (21B), bei4 also has the "clause union" characteristic in that it does not take aspect itself but allows its embedded non-finite verb to take aspect. Note also that since the OBJ of bei4 might not be overt, we impose a default value of OPT to its OBJ. If the OBJ of bei4 is overt, then its value will overwrite the placeholder value OPT; when its OBJ is not overt, then the value OPT will be assigned to be OBJ's default value, in which case the XCOMP's SUBJ will have the value OPT as well from unification with its controller, the matrix OBJ. Moreover, the matrix SUBJ controls the potentially unbounded OBJ or OBJ2. According to these specifications, sentences 1a, 1b, and 1c thus will have the following respective f-structures under our vLFG analysis:

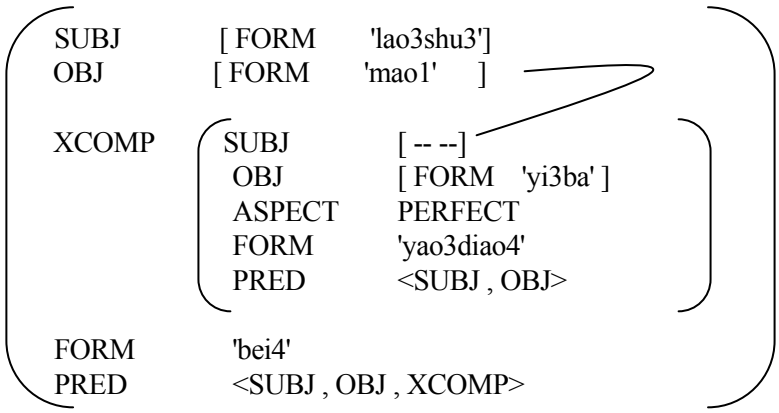
1a-f. `Lao3shu3 bei4 yao3 le.



1b-f. `Lao3shu3 bei4 mao1 yao3 le.



1c-f. `Lao3shu3 bei4 mao1 yao3diao4 le yi3ba.



3.3.21.5.4.4. Answers to Tan's objections to bei4's matrix status

Within this improved unified account for all bei4 sentences, we now answer each of the objections that Tan (1987) raised against Ma (1985). Tan's line of argument is that bei4 as a verb does not parallel other verbs with the same complement structure. What we will demonstrate below is exactly the opposite: bei4 does parallel other verbs with the same complement structure in a pivotal construction. We will show that Tan's conclusion is reached due to inappropriate analyses of the example sentences she uses or the over-generalizations she makes. The first objection she makes is that a bei4 sentence does not have the characteristic of preposing its OBJ as other similar type of verbs do, such as rang4 'let'. The examples she gives are the following:

4. a. Ta1 rang4 xiao3hair2 jin4 wu1. (Tan (1987) (16)a.)
he let child enter house
He let children enter his house.

a' Xiao3hair2, ta1 rang4 jin4 wu1. ((16)b.)
Child he let enter house
Children, he let them enter his house.

a" Ta1 ba3 xiao3hair2 rang4 jin4 wu1. ((16)c.)
he BA children let enter house
He let children enter his house.

b. Ta1 rang4 mei3 ge ren3 jin4 wu1. ((17)a.)
he let every CLS person enter house
He let everybody enter his house.

c. Ta1 shui2 dou1 rang4 jin4 wu1. ((17)b.)
he WHO all let enter house
He let everybody enter his house.

c' Ta1 dou1 rang4 shui2 jin4 wu1? ((17) b'.)
he all let who enter house
Who all did he let enter his house?

c" Shui2, ta1 dou1 rang4 jin4 wu1. ((17)c.)
WHO he all let enter house
He let everybody enter his house.

Tan is certainly correct in her observation that the OBJ of bei4 can not be preposed in the same manner as that of rang4 as the following examples show:

5. a. Lao3shu3 bei4 mao1 yao3 le.
mouse BEI cat bite LE
The mouse was bitten by the cat.

a'*Mao1, lao3shu3 bei4 yao3 le.
cat mouse BEI bite LE
The mouse was bitten by the cat.

a"*Lao3shu3 ba3 mao1 bei4 yao3 le.
Mouse BA cat BEI bite LE
The mouse was bitten by the cat.

b. Lao3shu3 bei4 mei3 zhi1 mao1 yao3 le.
mouse BEI every CLS cat bite LE
The mouse was bitten by every cat.

c.*Lao3shu3 shei2 dou1 bei4 yao3 le.
mouse WHO all BEI bite LE
The mouse was bitten by everyone.

c' Lao3shu3 dou1 bei4 shei2 yao3 le?
mouse all BEI who bite LE
By whom was the mouse bitten?

c"*Shei2, lao3shu3 dou1 bei4 yao3 le.
WHO mouse all BEI bite LE
The mouse was bitten by everybody.

However, Tan rather hastily made the conclusion that therefore bei4 cannot have the same complement structure as rang4, without further examining other similar verbs, such as yao4 'want', that are typically considered as having this <SUBJ , OBJ , XCOMP> complement structure and whose OBJ controls the SUBJ of its XCOMP.

6. a. Ba4ba4 yao4 xiao3hair2 kan4shu1.
papa want children read
Papa wants the children to read.

a'*Xiao3hair2, ba4ba4 yao4 kan4shu1.
children papa want read
Papa wants the children to read.

a"*Ba4ba4 ba3 xiao3hair2 yao4 kan4shu1.
papa BA children want read
Papa wants the children to red.

b. Ba4ba4 yao4 mei3 ge ren2 kan4shu1.
papa want every CLS person read
Papa wants everybody to read.

c.*Ba4ba4 shei2 dou1 yao4 kan4shu1.
papa WHO all want read
Papa wants everybody to read.

c' Ba4ba4 dou1 yao4 shei2 kan4shu1?
papa all want who read
Who does papa want to read?

c"*Shei2, ba4ba4 dou1 yao4 kan4shu1.
WHO papa all want read
Papa wants everybody to read.

We find exactly the same pattern between bei4 and yao4 'want'; this convincingly establishes the fact that the properties of preposing the object are lexically determined and are not general to this pivotal construction. To reject bei4 as having the pivotal construction based on its inability for its OBJ to be rearranged is also to reject yao4 'want'. We therefore dismiss Tan's first objection. Tan's second objection comes from the observation of how verbs of the pivotal construction interact with a resultative clause. We cite her examples first and will then dispute her analysis.

7. a. Wo3 ting1 ta1 ku1 de shang1xin1. (Ma 1987 (19)a.)
I hear he cry DE sad
I heard that he was crying sadly.

a' Wo3 bei4 ta1 ku1 de shang1xin1. (Ma (19)b.)
I BEI he cry DE sad
I, affected by his crying, am sad.

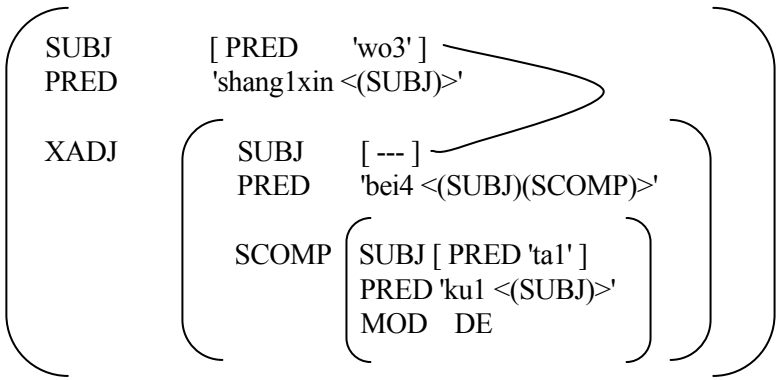
b. Xue2sheng1 kan4 lao3shi1 jiang3 de ke1shui4 le. (Ma (20)a.)
student see teacher lecture DE sleepy LE
The students saw that the teacher was lecturing sleepily.

b' Xue2sheng1 bei4 lao3shi1 jiang3 de ke1shui4 le. (Ma (20)b.)
student BEI teacher lecture DE sleepy LE
The students, affected by the teacher's lecturing, were sleepy.

Note that in sentences like 7a' and 7b' bei4 is considered as subcategorizing <SUBJ , SCOMP>, not <SUBJ , OBJ>. Tan thus assigns the following c-structure and f-structure to 7-a'.

7a'-c. [$_S$ Wo3 [$_{VP}$ [$_{VP}$ bei4 [$_S$ ta1 ku1 de $_S$] $_{VP}$] shang1xin1 $_{VP}$] $_S$]

7a'-f.



Tan first argues that in 7a and 7b, the objects, ta1 'he' and lao3shi1 'teacher' control the subjects of 'sad' and 'sleepy', but in the bei4 sentences the subjects, wo3 'I' and xue2sheng1 'students' are the controllers instead. Therefore, she concludes that bei4 cannot have ting1's and kan4's complement structure which she assumes to be <SUBJ , OBJ , XCOMP>. We challenge her assumption that ting1 'hear' and kan4 'see' have such a complement structure; instead we propose that they have the complement structure of verbs like shuo1 'say' and fa1xian4 'find': <SUBJ , SCOMP>. The typology of the complementation of verbs of perception is an interesting topic in itself. While

in English such verbs may have both complement structures: <SUBJ , SCOMP>, as in "I can hear that she is singing," and <SUBJ , OBJ , XCOMP>, as in "I can hear her singing," in Chinese, as we are trying to argue here, they can only take the former complement structure. Several pieces of evidence support our position. First, in a pivotal construction, the second verb, i.e., the verb in the infinitival XCOMP, can not be modified by the modal adverb zai4, which can only be attached to the matrix verb. However, such a restriction does not apply to verbs of perception like ting1 'hear' or kan4 'see'.

8. a.*Wo3 yao4 ta1 zai4 lai2.
I want he PROG come
I wanted him to be coming.

b.*Wo3 rang4 ta1 zai4 lai2.
I let he PROG come
I let him be coming.

c.*Wo3 qiang2po4 ta1 zai4 lai2.
I force he PROG come
I forced him to be coming.

d.*Wo3 qing3 ta1 zai4 lai2.
I invite he PROG come
I invited him to be coming.

e. Wo3 ting1 ta1 zai4 ku1.
I hear he PROG cry
I heard him crying.

e' Wo3 ting1 ta1 ku1-zhe.
I hear he cry ZHE
I heard him crying.

f. Wo3 kan4 ta1 zai4 ku1.
I see he PROG cry
I saw him crying.

f' Wo3 kan4 ta1 ku1-zhe.
I see he cry-ZHE
I saw him crying.

Another piece of evidence is that verbs of the pivotal construction typically cannot take a resultative complement such as dao4, but ting1 'hear' and kan4 'see' most often do.

9. a.*Wo3 yao4-dao4 ta1 lai2.
I want-RC he come
I wanted him to come.

b.*Wo3 rang4-dao4 ta1 lai2.
I let-RC he come
I let him come.

c.*Wo3 qian2gpo4-dao4 ta1 lai2.
I force-RC he come
I forced him to come.

d.*Wo3 qing3-dao4 ta1 lai2.
I invite-RC he come
I invited him to come.

e. Wo3 ting1-dao4 ta1 ku1.
I hear-RC he cry
I heard him cry.

f. Wo3 kan4-dao4 ta1 ku1.
I see-RC he cry
I saw him cry.

Finally, as Tan has observed herself, the sentential complement in a <SUBJ , SCOMP> structure can be topicalized. This applies to verbs like ting1 'hear' and kan4 'see', but not verbs like yao4 'want', qing3 'invite', and

qiang2po4 'force'. Note that it is precisely due to this observation that we totally agree with Tan that bei4 as a verb does not have the complement structure of <SUBJ , SCOMP>. The following examples clearly illustrate the points made here.

10. a.*Ta1 lai2, wo3 yao4.
he come I want
I wanted him to come.

b.*ta1 lai2, wo3 rang4
he come I let
I let him come.

c.*Ta1 lai2, wo3 qian2gpo4.
he come I force
I forced him to come.

d.*Ta1 lai2, wo3 qing3
he come I invite
I invited him to come.

e. Ta1 ku1, wo3 ting1-dao4.
hecry I hear-RC
I heard him cry.

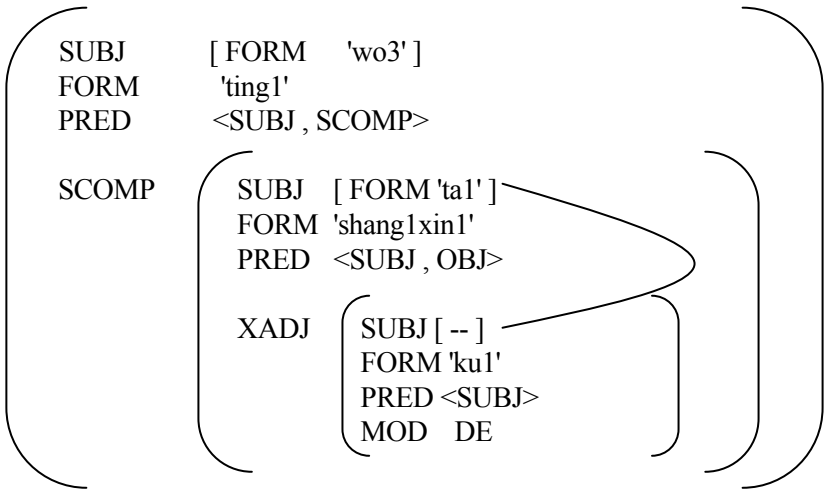
f. Ta1 ku1, wo3 kan4-dao4.
hecry I see-RC
I saw him cry.

g.*Mao1 yao3, lao3shu3 bei4.
cat bite mouse BEI
The mouse was bitten by the cat.

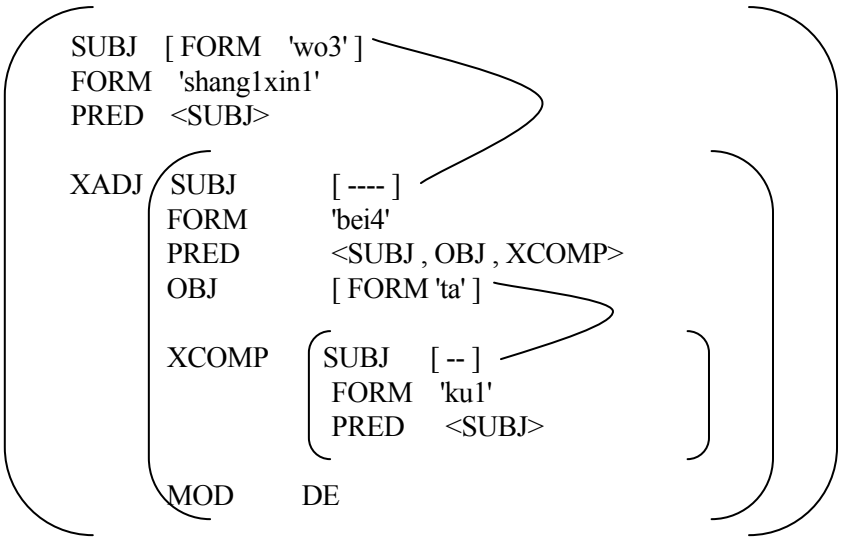
Based upon all the evidence we conclude that verbs like ting1 'hear' and kan4 'see' do not have the complement structure of the pivotal construction; rather they have the structure of <SUBJ , SCOMP>. In addition, we postulate

the analysis that de is a complementizer which indicates that the constituent preceding it is a modifying element and the constituent following it is the head. Therefore, in the sentence Ta1 ku1 de hen3 shang1xin1 'He is crying sadly', shang1xin1 'sad' is the matrix verb and ku1 'cry' is an adjunctive subordinate clause. Therefore, a more faithful translation of the sentence should be like "He is very sad in his crying." We therefore reject Tan's analysis of bei4 with the subcategorization of <SUBJ , SCOMP> and propose that 7a and 7a', still within our unified analysis, have the following f-structures respectively:

7a-f. `Wo3 ting1 ta1 ku1 de shang1xin1.
`I heard that he was crying sadly.



7a'-f2. `Wo3 bei4 ta1 ku1 de shang1xin1.
`I, affected by his crying, am sad.



We should acknowledge the fact that the analysis of sentences like "Ta1 ku1 de hen3 shang1xin1" that have the construction of [V1 de V2] is rather controversial in the field of Chinese linguistics. Our account that the V2 is the head verb and that the [V1 de] constituent is adjunctive has been quite convincingly supported by Huang and Mangione (1985). Huang (1990) summarizes their arguments: 1) V2 may form A-not-A question, not V1, 2) aspect particles may appear with V2, not V1, 3) negation scope involves V2 only, 4) the analysis of V2 as matrix better accounts for co-referential relations, 5) V1 cannot be morphologically complex, 6) the attachment of the question particle ma turns the sentence into a yes-no question at V2. J. Huang (1988), on the other hand, argues for an analysis within the GB theory where V1 is considered the matrix verb and V2 a subordinate verb. Although J. Huang has refuted some of the arguments of Huang and Mangione (1985), C. Huang (to appear) in turn refutes J. Huang's argumentation, and most importantly he points out that even if J. Huang's argumentation is valid, it only indicates that the analysis of V2 as matrix is not absolute, but it is not direct evidence for the analysis of V1 as matrix.

We have successfully dismissed all the objections that Tan has raised against our account for bei4, and our analysis does not have any of the difficulties we pointed out that her account for bei4 has. Lastly, one may object to the verb status of bei4 because of its inability to appear with aspect particles such as le. Our answer to this is that this is not unique to bei4: the same restriction applies to some other verbs of the same complement structure such as yao4 'want' and qian2gpo4 'force'. As we have pointed out earlier, this is a characteristic of the so-called "clause union verbs" of Chinese. In the case of aspect, they fuse together with the embedded verb and the aspect particle has to appear with the embedded verb. This fact is accounted for in bei4's lexical entry.

11. a.*Lao3shu3 bei4 le mao1 yao3.
mouse BEI LE cat bite
The mouse has been bitten by the cat.

b.*Wo3 yao4 le ta1 lai3.
I want LE he come
I have wanted him to com.

c.*Wo3 qiang2po4 le ta1 lai3.
I force LE he come
I have forced him to come.

3.3.21.5.5 Conclusion

To summarize, we first dismissed the analyses where bei4 is a function word, either as a subject marker or a preposition, and established the verb status of bei4. Further, we rejected Tan's account for bei4 as having a subordinate status due to several difficulties and the ad hoc nature of her analysis. Finally, we argued for the advantages for Ma's account that bei4 has the complement structure of <SUBJ , OBJ , XCOMP> and supplemented her account to make it complete. Within this complete and unified analysis for bei4, we finally successfully rebutted all the objections Tan raised to Ma's account.

3.3.22 <SUBJ , OBJ , SCOMP>

3.3.22.1 Regular Type

Examples: gao4su4 'tell', tong1zhi1 'notify', jing3gao4 'warn', da1ying4 'promise', quan4gao4 'advise', ti2xing3 'remind', bao4gao4 'report', bao3zheng4 'guarantee', da3du3 'bet'

1. a. Wo3 gao4su4 le ta1 ni3 ming2tian1 bu4 hui4 lai2.
I tell LE him you tomorrow not will come
I told him that you would not come tomorrow.

b. Wo3 gao4su4 le ta1 shei2 ming2tian1 bu4 hui4 lai2.
I tell LE him who tomorrow not will come
I told him who would not come tomorrow.

The embedded clause is evidently finite, and the fact that it may be interpreted as an indirect question when it has a wh-element in it leads to the conclusion that it is a subcategorized SCOMP. However, notice that verbs here allow but do not require their SCOMP to be interrogative.

FI-V-22:
[CAT V
FS [PRED <SUBJ , OBJ , SCOMP>]
]

bao3zheng4:
[FS [FORM 'bao3zheng4']
FI-V-22
]

3.3.22.2 Verbs Requiring an interrogative SCOMP

Examples: wen4, qing3wen4 'ask', zhi2wen4, pan2wen4 'question', kao3wen4 'torture and question'

2. a. Ta1 wen4 wo3 ni3 he2shi2 lai2.
he ask I you when come
He asks me when you will come.

b.*Ta1 wen4 wo3 ni3 mien2tian1 lai2.
heask I you tomorrow come
*He asks me you will come tomorrow.

Though both 3.3.22.1 and 3.3.22.2 subcategorize the same grammatical functions, the SCOMP required here must have an interrogative element and therefore must have the feature-value pair [Q +]. This embedded clause may then be interpreted as an indirect question, such as 2a. Verbs of the previous subcategory in 3.3.22.1 have no such constraint. Recall that the SCOMP in the subcategory 3.3.13.2 has exactly the same constraint.

FI-V-13-2:
[(↑ SCOMP Q) $=_c$ +]

zhi2wen4:
[FS [FORM 'zhi2wen4']
FI-V-22
FI-V-13-2
]

3.3.23 <SUBJ , OBL_{GOAL} , XCOMP>

3.3.23.1 Regular Type

Examples: jian4yi4 'suggest', an4shi4 'imply'

1. a. Wo3 xiang4 ta4 jian4yi4 ming2tian1 chu1fa1.
I to he suggest tomorrow start off
I suggest to him to start off tomorrow.

b. Wo3 xiang4 ta1$_i$ jian4yi4 ming2tian1 chu3fa2 ta1$_{j/*i}$.
I to he suggest tomorrow punish he
I suggest to him$_i$ to punish him$_{j/*i}$ tomorrow.

Although in LFG there is a constraint that oblique elements may not be the functional controller, it seems that for verbs in this subcategory the SUBJ of the XCOMP is identical with the noun in OBL$_{GOAL}$. The fact that in 1b above the second pronoun ta1 cannot be co-referential with the noun in OBL$_{GOAL}$ indicates that it is the noun in the OBL$_{GOAL}$ that unifies with SUBJ of the XCOMP. There are some English verbs, such as "signal," "wave," and "write" etc., that also seem to have the same property.

1. c. The general signaled to the samurai to jump.
d. The general$_j$ signaled to the samurai$_i$ to kill him$_{j/k/*i}$.
e. The general$_j$ signaled to the samurai$_i$ to kill himself$_{*j/i}$.

1a-f.

```
[ FORM          'jian4yi4'
  PRED          <SUBJ , OBL_GOAL , XCOMP>
  SUBJ          [ FORM   'wo3' ]
  OBL_GOAL      [ FORM    'ta1'
                  PFORM   'xiang4'
                  PCASE GOAL
                ]
  XCOMP    [ FORM   'chu1fa1'
             PRED   <SUBJ>
             SUBJ   [ --- ]
             ADJ    { [ FORM 'ming2tian1' ] }
           ]
]
```

FI-V-23:

[CAT V
FS [PRED <SUBJ , OBL_{GOAL} , XCOMP>
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
(↑ OBL_{GOAL}) = (↑ XCOMP SUBJ)
]

jian4yi4:

[FS [FORM 'jian4yi4']
FI-V-23
]

3.3.23.2 Verbs Requiring an interrogative SCOMP

Examples: jie3shi4 'explain', shuo1ming2 'illustrate', xun2wen4 'question'

2. a. Ta1 gen1 wo3 jie3shi4 ru2he2 zuo4ren2.
he with I explain how behave
He explains to me how to behave myself.

The XCOMP subcategorized by verbs in this subcategory must have an interrogative element, thus the feature-value pair [Q +]. This embedded clause may then be interpreted as an indirect question. FI-V-13-2 consists of such a constraint on SCOMP. We need a similar constraint on the XCOMP here.

FI-V-23-2:

[(↑ XCOMP Q) $=_c$ +]

jie3shi4:

[FS [FORM 'jie3shi4']
FI-V-23
FI-V-23-2
]

3.3.24 <SUBJ , OBL_{GOAL} , SCOMP>

3.3.24.1 Regular Type

Examples: shuo1 'say', biao3shi4, biao3da2 'express', bao3zheng4 'guarantee', bao4gao4 'report', shuo1ming2 'illustrate', du3 'bet', da3du3 'bet', biao3ming2 'clarify'

1. a. Wo3 gen1 ni3 shuo1 le ta1 bu2 hui4 lai2.
I with you say LE he not will come
I said to you that he won't come.

b.*Wo3 shuo1 ni3 ta1 bu2 hui4 lai2.
I say you he not will come
I tell you that he won't come.

c. Wo3 gen1 ni3 du3 ta1 bu2 hui4 lai2.
I with you bet he not will come
I bet you that he won't come.

b. Wo3 du3 ni3 ta1 bu2 hui4 lai2.
I bet you he not will come
I bet you that he won't come.

As we can see from the examples above, a few verbs, such as du3 'bet' and bao3zheng4 'guarantee', in this subcategory have counterparts in 3.3.22.1 that subcategorizes <SUBJ , OBJ , SCOMP>, while most of them, e.g., shuo1 'say' and biao3shi4 'express' do not.

FI-V-24:

[CAT V
FS [PRED <SUBJ , OBL_{GOAL} , SCOMP>]
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
]

jie3shi4:
[FS [FORM 'jie3shi4']
FI-V-24
]

3.3.24.2 Verbs Requiring an Interrogative SCOMP

Examples: wen4dao4 'ask', xun2wen4 'inquire'

2. a. Wo3	xiang4	ta1	wen4dao4	shei2	bu2	hui4	lai2.
I	to	he	ask	who	not	will	come
I ask him who will not come.							
b.*Wo3	xiang4	ta1	wen4dao4	Ma3li4	bu2	hui4	lai2.
I	to	he	ask	Mary	not	will	come
*I ask him Mary will not come.							

Again, verbs here require [Q +] in their SCOMP, which may be interpreted as an indirect question.

wen4dao4:
[FS [FORM 'wen4dao4']
FI-V-24
FI-V-13-2
($\uparrow$ OBL_{GOAL} PFORM) $=_c$ 'xiang4'
]

3.3.25 <SUBJ , OBJ , OBJ2 , SCOMP>

Examples: du3 'bet', da3du3 'bet'

1. a. Wo3	da3du3	ni3	wu3	quai4	qian2	ta1	bu2	hui4	lai2.
I	bet	you	5	CLS	money	he	not	will	come
I bet you 5 dollars that he won't come.									

b. Wo3 da3du3 ni3 wu3 quai4 qian2 shei2 bu2 hui4 lai2.
I bet you 5 CLS money who not will come
I bet you 5 dollars who won't come.

It seems that du3 and da3du3 'bet' are the only two verbs in Chinese that are ditransitive verbs that may also require a finite embedded clause. One piece of evidence for this embedded clause to be taken as SCOMP is that it cannot be topicalized. It thus has to be postverbal. Furthermore, this embedded clause with a wh-element may be interpreted as an indirect question, e.g., 1b; this fact indicates definitely that it is subcategorized for by the head verb.

FI-V-25:

```
[ CAT V
  FS   [ PRED <SUBJ , OBJ , OBJ2 , SCOMP> ]
]
```

3.3.26 <SUBJ , OBL_{GOAL} , OBJ , SCOMP>

Examples: du3 'bet', da3du3 'bet

1. a. Wo3 gen1 ni3 du3 wu3 quai4 qian2 ta1 bu2 hui4 lai2.
I with you bet five CLS money he not will come
I bet you five dollars that he won't come.

The above sentence illustrates the maximum grammatical functions that du3 and da3du3 'bet' may require. Similarly as the SCOMP in the previous subcategory, the embedded clause here cannot be topicalized and when it is interrogative it may be interpreted as an indirect question. It seems that in this and the previous subcategory they are the only verbs. They require their OBL_{GOAL} to have preposition gen1, he2, or tong2. The following FI entry specifies this constraint.

FI-V-26:

[CAT V
FS [PRED <SUBJ , OBL_{GOAL} , OBJ , SCOMP>]
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
(↑ OBL_{GOAL} PFORM) $=_c$ { 'gen1' , 'he2' , 'tong2' }
]

3.4 Summary of Syntactic Subcategorization of Verbs

Not considering the semantic classification of stativity and process, we have observed twenty-six possible subcategories of Chinese verbs thus far according to the various patterns of grammatical functions they subcategorize. A few of the subcategories are divided into subgroups due to differences in syntactic requirements of subcategorized elements. We will summarize our syntactic classification of Mandarin verbs in the following table.

Table 3.3
Syntactic Classification of Mandarin Verbs

1	SUBJ A) Intransitive Verbs: san4bu4 'walk' B) Weather Verbs: di4zhen4 'earthquake'
2	SUBJ , OBL_{LOCT}: zhu4 'live'
3	SUBJ , OBL_{GOAL}: zuo4ai4 'make love'
4	SUBJ , OBJ A) Typical Transitive Verbs: da3 'hit' B) Verbs Requiring a Locative OBJ: di3da2 'arrive' C) Locative Inversion Verbs: zuo4 'sit' D) you3 'exist, have' E) Verbs Requiring TOPIC: zuo4zhu3 'take charge'
5	SUBJ , OBL_{THME}: da3duan4 'break'
6	SUBJ , OBL_{THME} , OBJ: zhuang1man3 'fill'
7	SUBJ , OBL_{GOAL} , OBJ: jie3shi4 'explain'
8	SUBJ , OBJ , OBL_{LOCT}: xie3 'write'
9	SUBJ , OBL_{THME} , OBL_{LOCT}: gua4 'hang'

Table 3.3 (Continued) Syntactic Classification of Mandarin Verbs

10	SUBJ , OBL_{THME} , OBJ , OBL_{LOCT}: diu1 'throw'
11	SUBJ , NCOMP A) Equational Verbs: shi4 'be' B) Verbs Denoting Inherent Quality: zhi2 'worth'
12	SUBJ , XCOMP A) Regular Type: ji4hua4 'plan', kai1shi3 'start' B) Tough Construction: nan2 'difficult' C) Modal Verbs: hui4 'will', neng2 'can' D) Clause Union Verbs: she4fa3 'manage'
13	SUBJ , SCOMP A) Regular Type: shuo1 'say', que4ding4 'sure' B) Verbs Requiring a [Q +] SCOMP: yan2jiu4 'study'
14	SUBJ , OBJ2 , OBJ A) Regular Type: song4 'give', ti2gong1 'provide' B) Idiomatic Expressions: chi1dou4fu3 'tease'
15	SUBJ , OBJ , OBL_{BNFC}: gong1ying4 'provide'
16	SUBJ , OBL_{THME} , OBJ , OBJ2: jie4 'loan'
17	SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}: ji4 'mail'
18	SUBJ , OBL_{THME} , OBL_{BNFC}: ti2gong1 'provide'
19	SUBJ , OBL_{GOAL} , OBL_{THME}: shuo1 'say'
20	SUBJ , OBL_{GOAL} , OBL_{THME} , OBJ: jie3shi4 'explain'
21	SUBJ , OBJ , XCOMP A) Regular Type: jian4yi4 'suggest' B) Clause Union Verbs: shi3 'make', bi1 'force' C) "Promise" Verbs: da1ying4 'promise', dai4biao3 'represent' D) Tough Construction: zhi2de2 'worth' E) bei4 'suffer or incur (adversely)'
22	SUBJ , OBJ , SCOMP A) Regular Type: gao4su4 'tell', ti2xing3 'remind' B) Verbs Requiring an Interrogative SCOMP: wen4 'ask'
23	SUBJ , OBL_{GOAL} , XCOMP A) Regular Type: jian4yi4 'suggest', ti2xing3 'remind' B) Verbs Requiring a [Q +] SCOMP: jie3shi4 'explain'

Table 3.3 (Continued) Syntactic Classification of Mandarin Verbs

24	SUBJ , OBL_{GOAL} , SCOMP A) Regular Type: biao3shi4 'express', shuo1 'say' B) Verbs Requiring a [Q +] SCOMP: wen4dao4 'ask'
25	SUBJ , OBJ , OBJ2 , SCOMP: du3, da3du3 'bet'
26	SUBJ , OBL_{GOAL} , OBJ , SCOMP: du3, da3du3 'bet'

In all the subcategorization patterns, there seems to be an upper limit of four grammatical functions that a verb can require in Mandarin Chinese. We expect the maximum number of subcategorized functions in a lexical form is not universally uniform in all languages.

3.5 Comparison with Previous Classifications

There have been several previous classifications of Mandarin verbs. We will compare our classification based upon the subcategorization of grammatical functions with three of the existing classifications, Chao (1968:663-748), Li (1971:10-13) and CKIP (1989:62-72). Note that the focus of this comparison is not to view the previous works critically in terms of their comprehensiveness and coherence within their individual framework; rather, by contrasting with other classifications, strengths and contributions of our current work may be revealed. Likewise, strengths and contributions of previous works may suggest areas or provide a foundation for further improvements on our work.

Chao's work in Chinese linguistics has been the source of inspiration for many Chinese linguists. Both Li's and CKIP's classifications are extensions of Chao's classification which is based on two dichotomies: (a) the semantic criterion of activity versus stativity and (b) the syntactic criterion of transitivity versus intransitivity (Chao 1968:665).

Table 3.4
Chao's Classification of Mandarin Verbs (Chao 1968)

Intransitive:	
1. Action (Vi)	lai3 'come'
2. Quality (A)	da4 'big'
3. Status (Vst)	bing4 'sick'
Transitive:	
4. Action (Vt)	chi1 'eat'
5. Quality (Va)	ai4 'love'
6. Classificatory (Vc)	xing4 'be surnamed'
7. shi4	shi4 'be'
8. you3	you3 'have, exist'
9. Auxiliary (Vx)	hui4 'will, can'

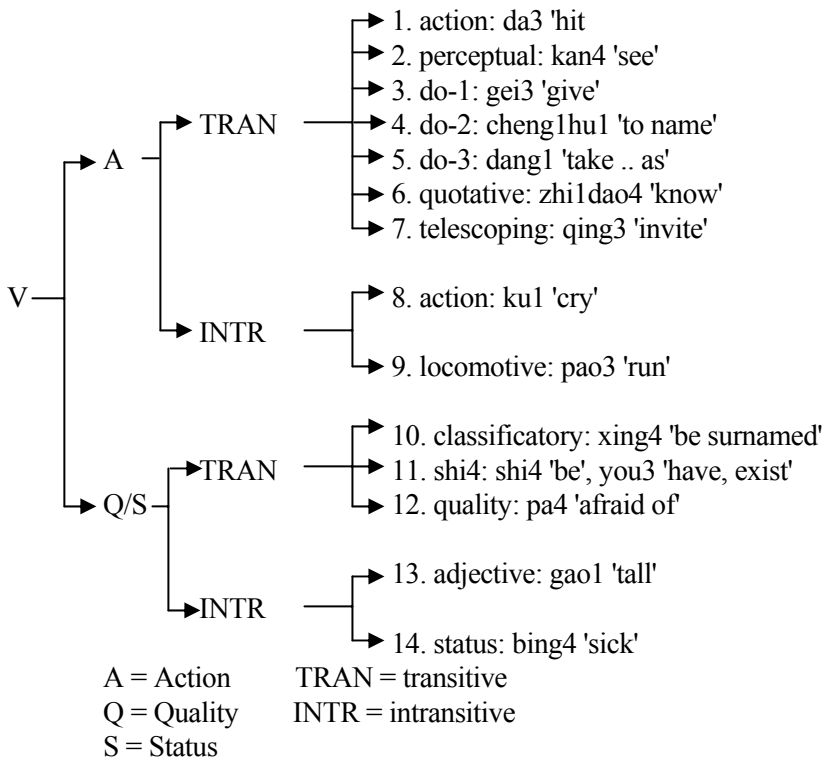
Chao also lists twelve syntactic constructions where these nine verb classes' possibility of occurrence is specified: (a) bu4 'not', (b) mei2 'not', (c) degree adverbs, (d) bie2 'do not', (e) cognate objects of extent, e.g., san1 nian2 'three years', (f) cognate objects of frequency, e.g., yi1 hui2 'one time', (g) reduplication, e.g., zuo3zuo3 'walk a little', (h) progressive aspect zhe, (i) experiential aspect guo4, (j) perfective aspect le, (k) imperative and (l) A-not-A question.

Chao's classification therefore encompasses three basic perspectives: syntactic criteria, semantic criteria, and co-occurrence constraints. Despite our focus on subcategorization for grammatical functions, our classification does have systematic schemes of both semantic criteria, [ACTIVE $\pm$] and [PROCESS $\pm$], and syntactic criteria, subcategorization for grammatical functions such as SUBJ, OBJ, etc. Nonetheless, although we have sporadically discussed the specific co-occurrence constraints that each verb subcategory impose on syntactic constructions, e.g., degree modification, modal verbs, and aspect markers, to achieve a more comprehensive classification of Mandarin verbs, Chao provides a good basic model for a more detailed and systematic study of the co-occurrence restrictions and requirements between our verb subcategories and functional elements of minor categories such as phrase markers, aspect markers, negation markers,

and sentential particles, and others like modal adverbs, degree modifiers, and duration and range phrases.

Li's classification (1971:10-13), incorporating subcategorization schemes in Chao (1968) and Wang (1963), is still basically based on the criteria of activity and transitivity, although finer distinctions are made according to other semantic characteristics.

Table 3.5
Li's Classification of Mandarin Verbs (Li 1971)



Li's study, following a transformational model of Fillmore's case grammar (Fillmore 1968), however, is primarily concerned with case in Chinese and the co-occurrence requirements between verb classes and different case roles

manifested by noun phrases. Therefore, a subcategory of verb is justified if members of the subcategory share the same requirements of case roles, or in structural terms, they share a common deep structure and thus may undergo the same transformational rules.

What Li (1971) has accomplished in the Fillmorean case grammar in terms of case roles and transformations is to be handled, in LFG, in the newly developed Lexical Mapping Theory in terms of thematic structures and morpholexical rules. Given the LFG position that linguistic information is co-described at three planes: thematic structure, functional structure and constituent structure, further study of the thematic structures of Mandarin verbs is needed to be incorporated into our current study of the subcategorization of grammatical functions and the syntactic encoding of grammatical functions. Despite the recognition of the excessive power of transformations, previous studies of case roles in relation to verbs, e.g., Li (1971) and Teng (1975), should still provide practical information regarding the thematic structure of Mandarin verbs in the application of Lexical Mapping Theory.

The CKIP's (1989) classification is primarily based on the requirements of the types and number of constituent categories, although the semantic distinction of action versus state is also employed. Despite the insignificance of transitivity in this scheme, verbs that require either a single NP or S constituent, i.e., subcategories Va and Vh, are considered intransitive and all others transitive. Notice that although the primary consideration seems to be subcategorized categorial constituents, this classification does not contain the category of prepositional phrase (PP), and it is said that subcategorized PPs are resolved by phrase structure rules. We will show CKIP's classification of Mandarin verbs in the following table.

Table 3.6
CKIP's Classification of Mandarin Verbs (CKIP 1989)

Argument Type	Action v. State	Number of Argument	Verb Subcategory
S	Action	NP S	Ve: cheng2ren4 'admit'
	State	NP S	Vk: zhi1dao4 'know' Vg: shi4 'be'
VP	Action	NP VP	Vf: she4fa3 'try'
	State	NP VP	Vl: you3yi4 'intend'
NP	Action	NP	Va: pao3 'run'
		NP NP	Vb: da3 'hit' Vc: dui4..ju2gong1 'bow'
		NP NP NP	Vd: song4 'give'
	State	NP	Vh: da4 'big'
		NP NP	Vi: dui4..ke4qi4 'polite Vj: teng2 'love'
		NP NP NP	Vd: qian4 'owe'

Although not incorporated in the classification scheme, case roles (Fillmore 1968 and Teng 1975) are assigned to arguments. Sixteen different kinds of case assignment are identified. Semantic consideration in this classification however is also hinted by classifying ju2gong1 'bow' (Vc) and ke4qi4 'polite' (Vi) as subcategorizing for <NP NP>, although they both require a prepositional phrase [xiang4 NP] or [dui4 NP]. CKIP also lists fifty-seven sentence patterns in which subcategorized constituents may appear. However, no indication is given regarding whether or how these sentence patterns correlate to the verb subcategories. Along with other types of syntactic and semantic information assigned to verbs, the CKIP's verb lexicon no doubt contains detailed and useful information; nonetheless, from what we have seen so far, it also demonstrates how in a loosely defined eclectic framework analyses can expatiate without concisely stating generalizations.

Our classification schemes employ both semantic criteria as well as syntactic ones. We have further divided the traditional distinction of action verbs versus state verbs and employed two binary features to distinguish four

classes of verbs. While the previous classifications are based on case roles and constituent categories, the contribution our syntactic classification makes to the study of Mandarin verbs is that it is based on the notion of subcategorization for grammatical functions. We have justified a set of subcategorizable functions in Mandarin: SUBJ, OBJ, OBJ2, XCOMP, SCOMP, OBL_{LOCT}, OBL_{GOAL}, OBL_{THME}, OBL_{BNFC}, and also presented the syntactic encoding of these functions. We distinguish twenty-six verb subcategories in Mandarin, and further subgroupings are made when differences in the syntactic requirements of the subcategorized elements within a certain subcategory are observed. Since notions of grammatical relations and subcategorization are not unique to LFG, our study fills the gap in the study of Mandrin verbs due to the lack of a systematic and extensive account of Chinese verb subcategorization for grammatical functions.

However, our classification of Chinese verbs, though fairly extensive, cannot be exhaustive. Finer distinctions can still be made within the same subcategory of verbs according to different thematic roles of or semantic selectional restrictions on the arguments. Ultimately it has to be recognized that every word behaves in some way unique to itself and we believe that the vLFG formalism we have demonstrated provides adequate and expressive mechanisms for the individual idiosyncrasies of individual verbs to be accounted for, and at the same time it also allows maximum generalizations among word classes.

CHAPTER 4
GRAMMATICAL THEORIES AND NATURAL LANGUAGE PROCESSING

A systematic analysis of verb subcategorization in Chinese is essential for the applications of a Chinese grammar in natural language processing (NLP). In any given sentence, there must be a predicative element that serves as the nucleus of that sentence. In Mandarin Chinese, only one particular type of sentence allows the predicative elements to be noun phrases. These noun phrases are however highly constrained: they must denote a certain inherent quality, such as worth, weight, length, age, and height, of the subject. They have to be considered predicative for they may appear in the predicate of a sentence without verbs, after the subject. Thus, this small class of nouns has to be identified as subcategorizing for <SUBJ>.

1. a. Ta1 yi1 bai3 sui4 le.
he one hundred year LE
He is one hundred years old.

b. Yi1 ge xi1gua1 liang3 kuai4 qian2.
one CLS watermelon two dollar money
One watermelon sells for two dollars.

Other than this class of predicative nouns, verbs are the necessary predicative elements that serve as the nucleus of a sentence (e.g., Li and Thompson 1981:85, Zhang 1983:96). A simple sentence may then be defined as a grammatically complete word string with only one predicative element and grammatical word strings with two or more predicative elements are complex sentences.

Since the nuclear element in a Chinese sentence is most commonly a verb, which holds together other dependent elements in the sentence, the information provided by verbs is most essential for the efficient and appropriate computational processing of Chinese sentences. As we have discussed in sections 2.5 and 3.1, elements that are dependents of the verb

may be dichotomized into two classes: the subcategorized arguments and the non-subcategorized adjuncts. In the theory of LFG, all dependent elements must be assigned certain grammatical functions in relation to the head verb, and subcategorized elements must have subcategorizable functions and non-subcategorized elements must be assigned non-subcategorizable functions. What we have attempted to accomplish in the previous two chapters is to first distinguish a set of subcategorizable grammatical functions in Chinese and second determine what each Chinese verb may require in terms of these subcategorizable functions. Therefore, information of verb subcategorization, together with the constraints of phrase structure rules, enables the computer processor to distinguish which dependent elements are required arguments and which are optional adjuncts in relation to the head verb, and thus efficiently reach a syntactic analysis of the sentence.

However, for the purpose of processing natural languages by computer, the first, and probably the most crucial, task is to choose an appropriate linguistic theory and formalism which must allow and require precise, concise, formal, and explicit statements of linguistic generalizations. They must provide a sound framework for the analysis as well as the generation of natural languages. While they should avoid the problem of over-generation, they should be expressive and not over-constrained. The formalism should be declarative, that is, the rules within the framework should be separate from the formal procedure of processing. In other words, changes made on rules should not affect the procedure of processing. Moreover, ideally, the processing process should reflect the psychological reality of human processing of natural languages. We believe that LFG theory and the variant formalism we have adopted in this study provide such a framework.

In the rest of the chapter we will examine in what ways Lexical Functional Grammar and two other current grammatical theories, Lexicase and Government Binding Theory, are related to computational applications in NLP. Within relevant contexts we will also touch upon actual NLP projects of Chinese. Most importantly we will attempt to demonstrate the suitability of LFG, especially in the vLFG formalism we have adopted in this study, for NLP applications.

4.1 Lexicase, Government and Binding, and LFG

The scene of linguistics, for nearly three decades, has been dominated by Chomsky and his theories of language. His current Government and Binding Theory, a direct descendant of his earlier theory of Transformational Grammar (TG), is still frequently referred to as the "main stream" or "conventional" theory of linguistics. However, in recent years, this perception may only be impressionistic rather than realistic. First of all, though accurate statistics are not available, in the linguistic literature GB does not seem to be the dominant theoretical framework that the majority of working linguists choose to work with when analyzing various natural languages. This is also true in the field of Chinese linguistics; GB practioners by no means constitute the majority of Chinese linguists. More importantly, some of the alternative linguistic theories, most of which have been developed out of and in reaction to Chomsky's earlier work, are increasingly gaining popularity (e.g., Sells 1985, Horrocks 1987, and Shieber 1987). In particular, with the advances in computer science, natural language processing for various applications has been vigorously pursued in the last decade and it is precisely in the area of actual, computational application of grammatical theories where GB is gradually losing ground, due to serious problems of earlier applications of transformational grammars and the indifference to formal rigor of most GB practitioners in stating linguistic generalizations and thus their lack of interest and commitment to NLP applications. We will compare GB with two alternative frameworks, Lexicase and LFG, in terms of how linguistic theories are relevant to NLP operations: parsing, sentence generation, and machine translation, in particular.

Lexicase, a lexicalist word-driven dependency grammar theory, has been developed at the University of Hawaii for the last two decades. It no doubt has a most serious commitment to formal rigor and actual application and is probably the most constrained syntactic theory among the three under discussion (Starosta 1985, 1988). As a linguistic theory, LXC is well-developed and has many attractive features that capitalize upon current linguistic trends. It has been applied to the analysis of linguistic phenomena in many different languages. Two major areas of the grammar of Mandarin Chinese have been explored in LXC theory: the analysis of prepositions (Starosta 1985) and word order (Her 1985-6). While Lexicase theory is better

known in Europe and Asia, that it has been somewhat overlooked by many linguists in the U.S. is probably due to a combination of factors: the fact that most of LXC papers are semi-published, the relative isolation of Hawaii from the mainland, and certainly to some degree the "ethnocentrism" of various linguistic camps. However, the recent publication of The Case for_Lexicase (Starosta 1988), which provides the first comprehensive account of the theory, and the positive reviews of the book in Language (Blake 1989) and Computational Linguistics (Fraser 1989) should definitely help promote the theory.

One of the advantages of LFG over LXC and GB is that in the past decade it has emerged to become an established linguistic framework within which there has been a great deal of interaction between theoretical construction and computational application. The emergence of unification grammars and their popularity among NLP researchers certainly has been an important factor. However, the advantage that LFG has over other unification formalisms is that LFG has been developed as a linguistic theory and has also been vigorously applied as a grammatical framework in analyzing many different languages, for example, English (e.g., Bresnan 1982a and Simpson 1983), Italian (e.g., Baker 1983), Malayalam (e.g., Mohanan 1983), Chichewa (e.g., Bresnan and Kanerva 1989), Japanese (e.g., Ishikawa 1985), Dutch (e.g., Zaenen and Maling 1983) and others. Its application in Chinese is also fairly extensive (see section 1.3.5 for references). The 1989 annual R.O.C. conference of computational linguistics, with a pre-conference workshop solely devoted to LFG, had a focus on the computational application of LFG in Chinese NLP.

Many existing NLP projects employ LFG or an LFG-like framework for linguistic analysis, such as the German-Japanese SEMSYN project at the University of Stuttgart, English-Japanese project at UMIST (University of Manchester Institute of Science and Technology), and, perhaps most prominently, the KBMT (knowledge-based machine translation) system under development at CMU (Carnegie-Mellon University). Even within the EUROTRA community some researchers are suggesting that the design may be better off if LFG innovations are integrated into the system (Gebruers 1989). The ECS MT systems that cover five language pairs also employ LFG for linguistic analysis (Her 1989). The functional formalism of LFG also means its greater portability to other functional formalisms such as Functional

Unification Grammar (FUG) and PATR II, and portability certainly is also a serious, practical consideration when it comes to choosing a framework for linguistic analysis in an NLP system. LFG has clearly been established as a practical working framework for NLP applications within which linguists also contribute collectively to the study of language.

4.2 Tasks of Natural Language Processing

NLP can be roughly perceived as having several modules where linguistic analysis plays an essential role. We illustrate the subdivisions in the following figure:

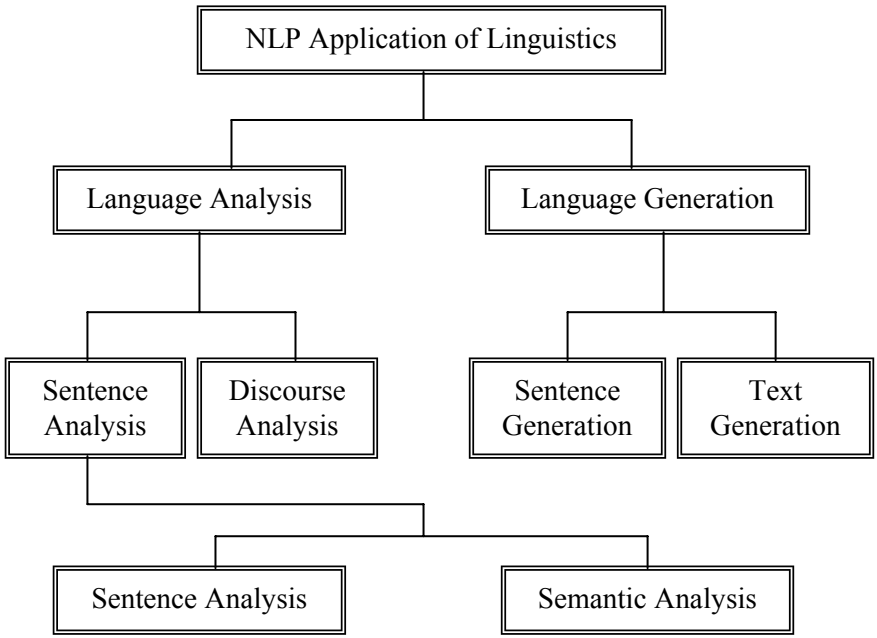


Figure 4.1 NLP Application of Linguistics

Syntactic analysis of sentences is commonly referred to as parsing. Here we will not address issues of semantic, discourse or dialogue analysis. In terms of

generation, we will discuss primarily the generation of sentences, for again the generation of multi-sentence text involves determination of both the content and the structure of a coherent discourse and is beyond the scope of our discussion here. More recent systems of machine translation of the indirect approach involve both parsing and sentence generation, while there is generally very little analysis of discourse structure integrated.

There are three fundamental functions that may be performed by formal grammars: parsing (the analytic function), generation (the synthetic function), and checking (the acceptor function). Different theories may have different claims about what functions are performed by a grammar and whether the same grammar is used for different functions or just one of the functions. For instance, a grammar in Kay's FUG formalism is seen as a competence grammar for generation but not directly involved in parsing. That of course can be considered as a weaker or more conservative position (Kay 1986).

LFG, whose formalism is otherwise very similar to that of FUG with the major difference being that in LFG constituent structure and informational feature structure are clearly distinguished into c-structure and f-structure, makes a stronger claim in that the same grammar in LFG may be reversible for parsing and generation. In other words, LFG claims to have all three functions of a grammar. A Lexicase grammar is a dependency grammar purely based upon the lexicon. Constituent relations are dictated by the valencies of head words in a syntactic construction. And, given also its sisterhood constraint, a Lexicase grammar should be equivalent to a context-free grammar and the same grammar may be used for parsing as well as generation, both of which have to be lexicon-driven. The analytical and synthetic functions also logically subsume the acceptor function.

GB, as a transformational grammar, involves two distinct strata of representation of syntactic structure, the d-structure and the s-structure, linked by a transformation. Parsing in a GB grammar must first construct the s-structure out of the input string and then transform, actually de-transform, it into the d-structure. The operation of parsing in GB is therefore quite different from that of generation which is much more straightforward for a GB grammar. Reversing and undoing the various effects of transformations from the s-structure to the d-structure in parsing has caused some serious concerns (King 1983). Furthermore, since the base X-bar component is responsible for only the generation or acceptance of d-structures and an input string to be

parsed has to be given an s-structure not directly generated or admitted by the base X-bar component, e.g., s-structures with adjunction of moved elements such as wh-questions and extraposition, it is obvious that the grammar for generation is not sufficient for parsing.

Whether psychologically human processing of language in terms of comprehension and production involves separate grammars is of course an empirical question and sound linguistic theories should reflect that psychological reality. However, in terms of efficiency for computational application, a grammar that may be applied to both parsing and generation is generally preferred. Therefore, the gap between human language processing and computational models is largely a separate issue. A parsing grammar may on the one hand produce syntactic structures that match native intuition and mimic certain aspects of human parsing, but on the other hand it does not necessarily reflect the mental process. We will look at one example.

Crain and Fodor (1986) conducted their own experimental study and also reexamined several earlier studies of the relationship between grammar and human sentence parsing. They observe that all evidence indicates that for the human mental parser information of both constituent structure as well as constraints on filler-gap dependencies is available simultaneously, and thus they argue against a modular transformational grammar model, such as GB, where constituent structure and constraints on filler-gap dependencies reside in different modules. They claim that GPSG, where there is a single, uniform system of rules, reflects human sentence parsing more faithfully in this respect. GB, as well as LXC and LFG, all fail to reflect that psycholinguistic insight, if Crain and Fodor's observation is correct. GB is a theory of high modularity, in which constituent structure is assigned by phrase structure rules and constraints of filler-gap dependency is a separate module that is effective after the application of the phrase structure rules.

Similarly in LXC the grammar first assigns constituent structure according to dependency specifications to form the "simple syntactic representation," then phrasal level semantic/anaphoric rules apply to fill gaps to form the "augmented syntactic representation" (Starosta 1988). Current LXC formalism, which employs "chaining rules" to account for filler-gap information (Lee 1989) still does not seem to allow the possibility of integrating information of filler-gap dependency with that of constituent structure simultaneously. LFG, according to the formulation in Bresnan (1982)

and the parser of LFG Grammar Writer's Workbench developed by Kaplan and associates at CSLI, builds c-structure through the application of phrase structure rules first; unification procedures then follow the specifications of functional schemata to build the corresponding f-structure of the sentence. It is at the f-structure that grammatical gaps are filled. Like GB and LXC, LFG therefore does not reflect Crain and Fodor's finding of human sentence parsing either. However, LFG theory does not rule out an alternative formalism which may account for their claim. That is, unification does not have to wait until a parse is reached, i.e., the whole c-structure of the sentence is built; there is an alternative to unify partial information to construct the f-structure simultaneously with the building of any component of the entire c-structure. Such is part of the proposal made by Wedekind (1986) where he also presents a mono-level variation of LFG. This alternative approach of building f-structure simultaneously with c-structure is precisely what we have adopted in our vLFG formalism.

4.3 Parsing Natural Language

In terms of parsing by a computer, there are many existing LFG parsers developed and employed by various NLP projects that employ LFG as the linguistic model. There is also a LXC parser developed at the University of Hawaii. While there are quite a few parsers, developed mostly during the 60's and the 70's, based on Chomsky's earlier transformational models, there are far less parsers based on the GB model. Also it does not seem that any of the GB parsers has been implemented in any actual NLP application. As mentioned earlier, any parser based on a multi-stratal grammar, such as GB, has to deal with the problem of analyzing the surface string and undoing the effects of transformations such as Move-α and deletions. The most common technique has been devising an additional set of phrase structure rules, often referred to as the "covering grammar," to assign s-structure and then reverse or undo the transformations that might have applied to generate this surface string. One of the problems at this point is that there is often more than one deep structure reached; each one of them is then checked by the base phrase structure rules to make sure that it is legal, and finally for each legal base tree reached thus far (forward) transformations are then applied to generate a surface string to be checked against the s-structure of the original input string.

Due to this complicated computation of transformations, compared to parsing in LFG and LXC, GB parsers are definitely more complicated in procedure and much less straightforward in the way rules in the grammar are interpreted by the parser.

It is interesting to note that a fairly recent report on a dissertation project attempts to develop a GB parser that avoids the above-mentioned problems by proposing a GB parser based on a variant GB model mentioned by Chomsky in passing in Chomsky (1981 pp.89-92). The parser is a dissertation project in progress at Syracuse University (Correa 1987). We first outline this variant GB model:

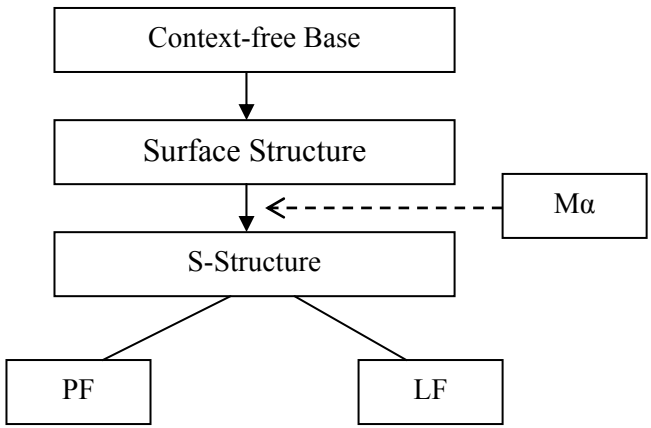


Figure 4.2 Correa's Variant GB Model (Correa 1987)

Essentially, the GB model adopted by Correa's parser is in fact transformationless and surface structure is base generated. What is accomplished in conventional GB by Move-α is now performed by an interpretive system called Mα. Chomsky has considered the assumption that syntax is subdivided into base rules and Move-α not a central one to the GB theory and that an obvious alternative to this assumption is:

> ...to suppose that base rules generate S-structure directly with new interpretive rules of the LF-component associating

> phrases and their traces (now base-generated). (Chomsky 1981 p.90)

On a different occasion, Chomsky again remarked on the insignificance of the distinction between these two alternatives:

> It is immaterial ... whether Move-α is regarded as a rule forming s-structure from d-structure, or whether it is regarded as a property of s-structures that are 'base-generated' ... It is in fact far from clear that there is a distinction apart from terminology between these two formulations. (Chomsky 1982:33)

Chomsky nonetheless rejected the argument that because it reduces the class of transformations not just to the single rule Move-α but further eliminates the entire class, the alternative view should be preferable. He argued that if Move-α is eliminated from syntax, a new category of LF-rules has to be added with exactly the same properties of Move-α. Correa's GB model is slightly different from the alternative that Chomsky was describing in that Correa's model posits a Mα interpretive component that applies to the base-generated surface structure (including trace and empty categories) to generate s-structure; therefore, surface structure and s-structure are identical in their constituent structure with the only difference being the latter has identified antecedents for its traces and empty categories; Mα is not in LF. Interestingly, there is a prototype Chinese parser based on GB that J. Huang, the most prominent GB proponent in the field of Chinese linguistics, was involved with, reported in Lin, et al (1986), which employs an approach rather similar to Correa's where empty categories are base-generated.

Correa (1987) concludes that parsing in this revised GB is much more efficient and straightforward than in conventional GB. Here we see a good example that one formalism can make parsing very cumbersome and another may be more suited for the computation that a parser has to perform, and yet the two formalisms may provide essentially the same linguistic information about the sentences to be parsed. That is also one of the important motivations behind the creation of LFG: it makes the computational process more efficient, and also better accounts for certain experimental data on human sentence

parsing, to replace certain transformations with lexical rules. Efficiency for processing is thus an important issue for LFG. Such is also the case for LXC (Starosta and Nomura 1986).

A Lexicase parsing system was outlined in Starosta and Nomura (1986). As expected, given the well-constrained dependency lexicon-driven grammar, parsing is a rather straightforward matter and the algorithm rather simple. Efficiency is enhanced by the one-bar constraint and the requirement that all terminal nodes be words; it is therefore impossible for the LXC parser to be trapped in a recursive loop. The Sisterhood constraint limits possible linking: linking (i.e., forming a higher constituent) can only be initiated by positive contextual features (which may require or allow sisters) marked on the head words. The Placeholder Substitution and Placeholder Expansion components described in the Lexicase parsing algorithm are essential for the economy of Lexicase parsing; they are similar in function to a tree-sharing device in other parsers and cut down considerably the redundancy of parse trees created by homographs, and given the Lexicase's treatment of words the number of homographs is definitely high. Lexicase parsing, being lexicon-driven, is necessarily bottom-up, but, as Starosta and Nomura have argued (1986), "Since it forms constituents simultaneously, it thus incorporates virtues of both top-down and bottom-up parsers." It would seem that due to the Sisterhood constraint the lexicase parser works in a breadth-first manner, rather than depth-first which enhances parsing speed if memory allows. We thus see that Lexicase has good potential for application of sentence analysis in NLP systems. A special advantage that LXC has is that its well-developed theory of case and thematic relations enables LXC to naturally and generally describe ergative languages as well as accusative languages. It may turn out that in analyzing ergative languages in the same consistent system as analyzing accusative languages LXC may have the leading edge. As we speculated earlier that Lexicase has been often overlooked is due to circumstantial factors not pertinent to the theory itself.

The Lexicase parser presented by Starosta and Nomura however does not seem to be able to handle control relations and filler-gap dependencies. That part of Lexicase theory is relatively recent, developed by Pagotto and Lindsey, and if it can be incorporated in the parser it will be a good test for the formulation of rules of that component.

Similar to Lexicase, lexicon plays an important role in LFG. However, unlike Lexicase, constituent structure is specified with conventional context-free phrase rules, which are further augmented with functional schemata prescribing manners of unification to construct f-structures. Parsing in LFG, in terms of c-structure analysis, should be of the same efficiency as parsing in a context-free phrase structure grammar. PSRs may apply either top-down or bottom-up. The LFG parser developed at CSLI is a top-down parser. The ECS parser proceeds bottom-up first then top-down (Her 1989). In the conventional LFG formalism, when final parses are reached, unification procedures then apply to build an f-structure out of each parse and throw out the ones that violate constraints. Control relations and filler-gap dependencies are specified when the f-structure is built. In our vLFG formalism, a partial f-structure is a "by-product" whenever a partial c-structure is build. It is clear then that the addition of f-structure description in either formalism helps ensure the accuracy of parsing but does not complicate the parsing process of a context-free phrase structure grammar. The operation of unification also enhances the efficiency of parsing. We will illustrate with a simple example:

2. a. The deer swims.
   b. The deer swim.
   c. The deer swam.

Within non-unification-based grammars, such as GB and LXC, it is necessary to produce two parses for 2c. For 2a and 2b, even given tree-sharing facilities, one parse tree is successful and accepted with the other thrown out for violating subject-verb agreement. Within a unification-based approach, all three sentences reach only one parse with no other parses violating subject-verb agreement and failing. We will illustrate this point in a relatively theory-independent unification formalism.

3.a. the: CAT DET

$$\begin{pmatrix} \text{SPEC} & \text{'the'} \\ \text{DEFINITE} & + \\ \text{NUMBER} & \text{ANY} \end{pmatrix}$$

b. deer: CAT N

$$\begin{pmatrix} \text{PRED} & \text{'deer'} \\ \text{NUMBER} & \text{ANY} \end{pmatrix}$$

c. swim: CAT V

$$\begin{pmatrix} \text{SUBJ} & [\ \text{NUMBER PL}\ ] \\ \text{PRED} & \text{'swim <SUBJ>'} \end{pmatrix}$$

d. swims: CAT V

$$\begin{pmatrix} \text{SUBJ} & [\ \text{NUMBER SG}\ ] \\ \text{PRED} & \text{'swim <SUBJ>'} \end{pmatrix}$$

e. swam: CAT V

$$\begin{pmatrix} \text{SUBJ} & [\ \text{NUMBER OPT}\ ] \\ \text{PRED} & \text{'swim <SUBJ>'} \end{pmatrix}$$

4. a. S --> NP V
<↑ SUBJ = ↓> <↑ = ↓>

b. NP --> DET N
<↑ = ↓> <↑ = ↓>

We will not go through the unification process step by step here. We will simply point out that, through the unification of the verb and its subject, subjects in 2a and 2b get their NUMBER specification SG and PL respectively, and for 2c the number of subject remains the unspecified OPT;

no parse that violates subject-verb agreement is ever pursued any further. Whether psycho-linguistically this unification approach matches human sentence processor or not is a separate, empirical issue. The operation of unification, however, does allow simplicity and better efficiency in terms of computation. More specifically in this case it allows the lexical items to be under-specified and produces a single parse, while in a non-unification approach the lexical items have to be fully specified and thus multiple parses cannot be avoided.

As we have discussed earlier, unification in conventional LFG comes after the construction of c-structure and that an alternative is to build the c-structure and f-structure simultaneously. If unification is executed simultaneously, as the vLFG formalism we describe in this book does, parse trees with illegal f-structures are detected earlier and thrown out and the same path will not be continued; however, for every possible constituent unification has to be performed to decide whether the constituent is associated with a legal partial f-structure. Therefore, in terms of efficiency for parsing, there is a tradeoff and which approach is preferable is a mathematical and statistical matter. However, as mentioned earlier the latter approach of the vLFG probably better reflects the psychological reality of human sentence parsing.

4.4 Natural Language Generation

In our discussion of parsing, we have assumed that what is to be accomplished by a parser is to provide syntactic analyses in a formal representation of the sentence structure. However, in actual NLP systems very often this formal representation of sentence structure needs to be further translated into a logical, semantic, or knowledge representation. Likewise, sentence generation generally involves first accessing information from a knowledge base and then translating the information accessed into a knowledge representation which then needs to be mapped into a form from which the grammar can generate sentences; discourse principles then decide the organization of the final output text. The generation of sentences is therefore only one part of the whole system and how a grammar generates sentences often relies on how much linguistic information is available as its input. (Sentence generation in MT systems is usually simpler, which will be discussed momentarily.)

We are not aware of any NLP system that employs GB or LXC frameworks for the linguistic component for sentence generation. There are several systems that employ LFG or other unification approaches such as FUG for generation and there also exist earlier systems that employ Transformational Grammar. For a description of a sentence generation mechanism that maps a logical representation into the deep structure and then from the deep structure to sentences, refer to Grishman (1986 pp.160-168). The reason LFG is preferred over a theory like GB by NLP system designers may be that the functional or informational structure of a sentence in LFG is independent from the constituent structure. The functional structure provides an intermediate stage for generation: semantic or knowledge representation is mapped into the functional structure and then the sentence with constituent structure is generated based on the functional structure. To map a semantic or knowledge representation directly into an entirely language-specific constituent structure of a sentence is much more involved due to the complication of language-specific word order constraints. Another advantage of LFG is that there has been some substantial work done within the framework on the representation of the formal semantic structure of a sentence (Holversen 1983). (The LFG Grammar Writers' Workbench parses a sentences and can provide c-structure, f-structure, and Semantic Structure.) In this respect, LFG better facilitates the interface among different modules of a natural language generation system than GB or LXC.

Another advantage of a separate f-structure and the operation of unification is that often the input information for the sentence to be generated does not have to be complete: the missing information will be supplied later automatically by unification. We will illustrate this point with a very simple example. Suppose a query system is providing the answer "three warships" in Chinese to the question "What objects are approaching?" and suppose the system only provides san1 'three' as the quantifier and jun1jian4 'warships' and thus the semantic structure and then the following f-structure:

5.
```
[ ADJ   [ FORM 'san1' ]
  FORM 'jun1jian4'
]
```

Note that the f-structure given does not have any information on the classifier that the noun is required to have. The generator first looks up the words from the dictionary and unifies the given f-structure with the functional structure of the words found in the dictionary according to the relevant phrase structure rule and thus expands this given functional structure.

6.a. san1:

```
[ CAT QTFR
  FS   [ FORM 'san1' ]
]
```

b. jun1jian4:

```
[ CAT N
  FS   [  FORM 'jun1jian4'
          CLASS 'sao1'
       ]
]
```

c. sao1:

```
[ CAT CLS
  FS   [ CLASS 'sao1' ]
]
```

7. NP --> QTFR CLS N
(↓ ε ↑ ADJ) (↑ = ↓) (↑ = ↓)

After unification as specified by this rule we have:

8.

```
[ FORM   'jun1jian4'
  ADJ    { [ FORM 'san1'] }
  CLASS 'sao1'
]
```

We thus successfully generated the answer san1 sao1 jun1jian4 with the appropriate classifier. Among all the classifiers only sau1 has the compatible value of CLASS that can unify with the value of CLASS specified in the noun jun1jian4.

This example may also be used to demonstrate the better efficiency of our vLFG formalism in sentence generation compared to the conventional LFG. In conventional LFG, since unification proceeds only after a final c-structure is generated, it would over-generate many noun phrases of san1_CLS jun1jian4 where every classifier in the lexicon can be inserted under the category CLS to form a legal c-structure, and only when unification applies to all these c-structures, will the wrong classifiers be filtered out. In our vLFG however, since an f-structure is always being build simultaneously, none of the wrong classifiers can be successfully inserted as part of the noun phrase san1 CLS jun1jian4. The vLFG thus totally avoids the problem of over-generation here.

4.5 Machine Translation

Machine translation is an NLP task that involves both parsing and sentence generation. Refer to Her (1987) or Slocum (1985) for an overview of current MT approaches. In this last section we will discuss these three frameworks in relation to the transfer approach and interlingua approach of MT, and points made earlier on parsing and generation will not be repeated here unless it is necessary to do so.

It is likely, though there is no actual statistics, that LFG, along with its variations, is the most popular choice among established, non-system-specific grammar frameworks among MT projects that are currently under development. Once again it seems that it is the division of c-structure and f-structure that makes LFG more suitable for both the interlingua as well as the transfer approach. We will discuss the interlingua approach first. Interlingua is a supposedly universal representation of meaning. If either GB or LXC is used as the linguistic model, then it is the sentence with constituent structure that serves as the input to be mapped into this universal meaning representation. For generation of the sentence in the target language this meaning representation needs to be mapped into a language-specific constituent structure of the target sentence again. As we have pointed out before, f-structure in LFG contains functional information, with information of control relations and gaps filled, abstracted away from the constituent information. Since the functional structure is more language-independent, it should be less involved to transform the f-structure into a universal meaning

representation. The KBMT project at CMU, one of the best-known MT systems under development that employ an interlingua approach, implements an LFG-like formalism as the linguistic model in the system.

If a transfer approach is adopted in an MT system, LFG still maintains the advantage. In fact, most of the current MT systems or projects adopt the transfer approach. Again, since the functional structure contains information already abstracted away from the highly language-particular constituent structure, less manipulation needs to be performed on the source functional structure for it to become the functional structure of the target sentence. On the other hand, to transfer the highly language-specific tree or constituent structure for it to become the usually very different shape of the target sentence, more manipulation is needed. Furthermore, since the f-structure does not contain constituent information, often little transfer is needed. This economy of transfer becomes most evident when translating two languages with drastically different word order and constituent structure, for example, Japanese and English.

9. a. Taroo ga eigo o hanasu.
Taroo GA English O speak

b. Taroo speaks English.

Japanese is a verb final, relatively order-free language with a complex case marking system and English is basically SVO with little case marking. 9a and 9b therefore have quite different constituent structures. However, in terms of their f-structures, within which ordering of attribute-value pairs is entirely arbitrary, 9a and 9b are very similar.

9a-f.

```
[ SUBJ [ FORM   'Taroo'
         PCASE  'ga'
       ]
  OBJ  [ FORM   'eigo'
         PCASE  'o'
       ]
```

```
  PRED <SUBJ , OBJ>
  FORM 'hanasu'
  TENSE PRES
]
```

9b-f.

```
[ SUBJ [ FORM          'Taroo'
         NUMBER        SG
       ]
  OBJ  [ FORM          'English'
         NUMBER        SG
       ]
  PRED <SUBJ , OBJ>
  FORM 'speak'
  NUMBER SG
  TENSE PRES
]
```

For translation between the two sentences, studies on verb subcategorization for grammatical functions become very useful. Given appropriate phrase structure rules, information of subcategorization of grammatical functions, and well-formedness constraints on f-structures, little transfer needs to be performed if we specify for the English grammar to ignore the irrelevant Japanese PCASE and for the Japanese grammar to ignore the English NUMBER. Another important point is that, as we mentioned earlier in the discussion of sentence generation, the f-structure needs not be entirely complete to generate the correct sentence. Our example with the Chinese classifier can be repeated here. To translate the English phrase "three warships" into Chinese "san1 sao1 jun1jian4" there is no need to add the information about classifier in the transfer f-structure. That information will be supplied later through unification during generation. This again reduces the amount of manipulation during the transfer stage. But if we have to perform transfer on the tree representation of GB or LXC, more manipulation is needed, including the notoriously hectic task of reordering of constituents, and every detail of relevant information has to be provided. Another advantage of transferring the f-structure is that it has control relations

identified and gaps filled so that if a gap has to be realized overtly or lexically in the target language there is no need to supply that information.

However, between GB and LXC, as Starosta and Nomura (1986) have argued very briefly, because the one-bar constraint and Sisterhood constraint of Lexicase seriously limit the possible kinds of constituent structure and it follows that between any two languages the possible variance of constituent structure is drastically reduced, therefore the amount of transformation during the transfer stage for producing the structure of the target sentence should also be considerably reduced. LXC thus seems to be a better choice than GB. But as it is presently formulated, LXC does not have a formal representation of the sentence structure that has control relations and filler-gap dependencies integrated.

4.6 Summary

To summarize, in terms of theoretical linguistics, GB may still appear to be the dominant theory, but in terms of NLP application, LFG, along with other unification approaches, is ahead of GB. Judging from the very fundamental spirit of modern sciences, it is reasonable to believe that grammatical theories with strong commitment to NLP application in academia as well as in industry are to benefit greatly from this experience and eventually override theories without that commitment. Based upon this belief and also the intention to contribute to its development we have chosen LFG as the basic linguistic model for this study of Mandarin verbs as part of the grammar for Mandarin Chinese and made certain improvements to its formalism.

CHAPTER 5
CONCLUSION

The primary endeavor of this book is to develop a classification of Mandarin verbs, within the theoretical framework of Lexical Functional Grammar, in terms of the grammatical functions that verbs subcategorize for in Mandarin Chinese. We hope that this study constitutes a useful addition to Chinese linguistics and that it contains new insights into Mandarin verbs. Aside from the contribution to Chinese syntax, we also hope that the modifications incorporated in the vLFG formalism within which Mandarin verb subcategorization is presented do make theoretical as well as practical sense and that this study can bring attention to practitioners and theoreticians of the conventional LFG and thus generate constructive discussions and empirical studies concerning the possible advantages of the alternatives we have adopted. In the following pages, we will have some concluding remarks regarding the vLFG formalism and our scheme of Mandarin verb subcategorization. We will conclude the book after we suggest several areas for further research.

5.1 The vLFG Formalism

Although vLFG follows most closely the theoretical constructs of LFG, there are two differences in vLFG formalism: the notation of phrase structure rules and the implementation of feature inheritance entries in the lexicon. A difference between two grammar formalisms can be either notational or notional (Shieber 1987:2-3). Generally, if the difference is purely notational, then linguistic analyses expressed in one grammar formalism can be easily reduced, in the mathematical sense of the word, in the other formalism. However, if reduction is impossible or cannot be achieved in a straightforward and perspicuous manner, then such difference is likely to be notional. Both of the two differences between LFG and vLFG, rule notation and inheritance entries, are basically notational, for it is clearly the case that expressions in one of the two formalisms can be easily and adequately reduced in the other. The vLFG formalism is therefore precisely what its

name is intended to convey: a variant of the conventional LFG formalism (e.g., Wescoat 1987).

Given the fact that in vLFG unification takes place whenever constituents are found and a higher category is to be built, it is logical for the phrase structure rules to state the constituents before the higher category, e.g., NP VP -> S, which can be built only when unification among its constituents succeeds. No apparent advantages, nor disadvantages, can be observed regarding this variant rule notation. In contrast, the conventional phrase structure rule format, e.g., S -> NP VP, suffices for LFG, for unification does not take place until a final tree, or c-structure, is constructed and thus the construction of c-structure is independent of that of the f-structure.

The implementation of feature inheritance entries in the organization of lexicon in vLFG, though also merely notationally variant from LFG formalism, does have the clear advantage of minimizing the size of lexical entries and enables a more modular lexicon. Furthermore, in the conventional LFG lexicon, where every lexical entry is fully specified, conceptually there is no device available in its formalism to express information shared by members of a word class. In our current study, it is precisely by means of feature inheritance entries that subcategorization groupings of Mandarin verbs are stated. Like the Redundancy Rules in Lexicase, vLFG feature inheritance entries provide a conceptually explicit device for word class groupings. Moreover, the fact that they facilitate underspecified lexical entries leads to a lexicon of drastically reduced size with increased modularity. It is a reasonable assumption that, in terms of natural language processing, a more modular lexicon is easier to maintain and that a lexicon of reduced size enhances processing efficiency by requiring less computer memory and storage space.

Finally, recall that regarding the processing of syntactic rules, there are two places where vLFG differs from LFG: first, in vLFG whenever a constituent is constructed, it must have a well-formed f-structure as well, while in LFG f-structures are constructed only when final c-structures are reached; second, at any point of the processing in vLFG, a c-structure with an incoherent f-structure will not be constructed, while in LFG the Coherent Condition is checked only when final f-structures are constructed. As we have stated in the following sections: 1.4.3, 2.6, 4.2, and 4.3, the two vLFG alternatives may have certain desirability over LFG with regard to computational efficiency

and psycholinguistic reality of human sentence processing due to vLFG's earlier detection of f-structure violations. Yet, in order to substantiate or repudiate this rather intuitive inclination of ours further research is necessary.

5.2 Mandarin Verb Subcategorization in LFG

Although it has only been a few years since LFG was first introduced to the analysis of Chinese (e.g., Huang 1985 and Ma 1985), a fairly wide range of syntactic constructions in Mandarin Chinese has been treated, as we have indicated with references in section 1.1. Huang's 1987 dissertation, where LFG is compared with three other linguistic frameworks GB, GPSG, and RG in the analysis of Mandarin de, was no doubt the first major work, and he continues to be the prominent voice in bringing LFG's theoretical developments into the study of Chinese syntax (e.g., Huang 1989) and promoting the application of LFG in Chinese computational linguistics (e.g., Huang and Chen 1989). It is hoped that our current study of Mandarin verb subcategorization of grammatical functions, being the first such attempt in Chinese linguistics according to our knowledge, further broadens the range of LFG application in Chinese and contributes to this collective effort of a comprehensive description of Mandarin grammar.

Regarding the subcategorization of Mandarin verbs, there have been several previous works (e.g., Chao 1968, Li 1971, and CKIP 1989). As we have indicated at the end of CHAPTER 3, while previous classifications are often based on case roles and/or constituent categories, our current classification of Mandarin verb is unique in that it is based on the notion of subcategorization for grammatical functions. It is necessary for us to first establish a set of grammatical functions in Mandarin Chinese and justify a subset of them as subcategorizable: SUBJ, OBJ, OBJ2, XCOMP, SCOMP, NCOMP, OBL_{LOCT}, OBL_{GOAL}, OBL_{THME}, OBL_{BNFC}. Additionally, we discuss the syntactic encoding of these subcategorizable functions in Mandarin. Based on the different requirements verbs have in terms of these subcategorizable functions, we distinguish twenty-six verb subcategories in Mandarin. Within several of the subcategories, further subgroupings are made according to differences in the syntactic requirements of the subcategorized elements. Given the recognition of notions of grammatical relations and subcategorization in general linguistics, our study fills the gap in Chinese

syntax due to the lack of a systematic and extensive account of Mandarin Chinese verb subcategorization for grammatical functions.

Moreover, we have not neglected the traditional classification of Mandarin verbs along the dichotomy of stative verbs versus active verbs. In fact, following Her (1985-6), we have a more refined scheme of semantic classification in employing two binary features, [STATE ±] and [PROCESS ±], to obtain the classification of four subcategories: (a) state verbs, e.g., gao1 'tall', (b) process verbs, e.g., chen2 'sink', (c) action verbs, e.g., ti1 'kick', and (d) action-process verbs, e.g., liang4gan1 'line-dry'. Also, we furnish several examples of how this semantic classification may be relevant and applicable to certain syntactic generalizations, for instance the derivational process of resultative compounding and co-occurrence restrictions of aspect markers, phrase markers, and degree adverbs.

To better assist the determination whether a dependent constituent should be deemed subcategorized-for or not, we have surveyed and organized ten potential rule-of-thumb tests: (1) obligatory co-occurrence, (2) ontological necessity, (3) constancy of semantic content, (4) distributional restrictions, (5) one per sentence, (6) proximity to the head, (7) semantic selectional restrictions, (8) possible internal gaps, (9) ability of being an indirect question, and (10) identification of a subcategory. Linguists who are interested in the study of subcategorization should find this information useful.

In addition to the classification schemes, a number of our analyses of particular syntactic constructions also differ from previous accounts. Among them, we have called for the recognition of topic as a syntactic notion and that the term "topic" be reserved for this use only, to avoid the previous confusion. Having suggested the term "frame" to denote the semantic/discourse function encoded by topic, we recounted the encoding of semantic functions of frame, focus, and contrast by syntactic relations of subject, topic, and predicate. We further argue that topic, as a syntactic notion, should not be subcategorizable in Mandarin, contrary to Huang's (1989a) proposal, and provide an alternative analysis, one which does not compromise linguistic generalizations, for a small number of verbs in Mandarin Chinese that require a sentential or discoursal topic.

Contrary to conventional wisdom, we treate OBJ2, or the indirect object, as encoded by the second postverbal NP, rather than the first. This position is consistent with and directly tied to our insistence that [ba3 NP] should not be

taken as object. We demonstrate with evidence that, among other mis-generalizations, to treat [ba3 NP] as direct object necessarily leads to the recognition of a double direct object construction, which violates the universal that only one direct object is possible in a clause. Following Her (1985-6), we analyze [ba3 NP] as a prepositional constituent encoding the subcategorizable function of OBL_{THME} and account for several types of ba3 constructions.

Another controversial lexical item is of course bei4. We argue and provide evidence against the conventional analyses treating bei4 as either a subject marker or a preposition. Aside from violating universal characteristics of the passive construction, the prepositional analysis of bei4 also leads to unnecessary complication in analyzing the status of the verb following bei4. Expounding and extending the account in Ma (1985), we formulate bei4's lexical entry as a verb in the pivotal construction subcategorizing for <SUBJ , OBJ , XCOMP> and give a more comprehensive analysis of bei4 sentences.

Our treatment of you3 'have, exist' as a main verb to certain extent also differs from previous analyses (e.g., Cheng 1979, Starosta 1985). Although we do recognize that semantically you3 has several different usages, we contend that given a general concept of existence expressed by you3, i.e., [(A) you3 B] is interpreted as "B exists in the domain of A," all these different usages and different syntactic requirements can be unified and accounted for by a single lexical entry. Furthermore, we demonstrate with various types of evidence that expressions like you3 qian2 'rich' and you3 yi4si 'interesting' should be treated as idioms and provided explicit formulation in vLFG.

Given the commonality shared by contemporary grammatical theories within the generative tradition and the considerable reducibility among their formalisms, linguistic analyses need not be regarded as entirely unseparatable from their formulations within particular linguistic theories and formalisms (Shieber 1987). Since the notions of grammatical function and subcategorization are employed by most generative theories and these notions can also be understood in a theory-independent context, the results of our study of Mandarin verb subcategorization should be not only directly applicable to an LFG grammar of Mandarin Chinese but also constructive to Chinese grammarians working with other theories. We are also hopeful that our specific analyses of the various syntactic constructions in Mandarin offer new perspectives to examining these issues and succeed in revealing fresh insights. In addition, these results should be conducive to the syntactic

processing of Chinese sentences in NLP applications due to the importance of verbs as nuclei in sentences, our explicit formulation of analyses, and our attention to computational efficiency.

5.3 Suggested Areas for Further Research

However, given the complexity of language and our incomplete understanding of how language works, it is unrealistic to believe that any existing linguistic theory in its current form provides the ultimate, correct explanation of language and also implements the most appropriate formalism. Advancements in linguistic theories unquestionably should and will continue as our understanding of language progresses. Lessons learned from psycholinguistic studies and NLP applications will improve upon existing linguistic formalisms. Furthermore, while a sound and well-constrained theory and its formalism do restrict the possible formulation of linguistic analyses, they certainly do not dictate the analyses nor ensure the appropriateness of the analyses within them. Accordingly, we welcome and eagerly await discussions, debates, and researches that would further advance our classification of Mandarin verbs to be more comprehensive and to enhance our analyses of grammatical functions and specific syntactic constructions. There are several areas where further research is needed to improve upon or extend this work.

First of all, in terms of the oblique function, OBL, encoded by prepositional phrases, we have considered only a portion of prepositions in Mandarin Chinese and recognized four subtypes OBL_{LOCT}, OBL_{GOAL}, OBL_{THME}, OBL_{BNFC}. Two specific areas need to be further explored. First, the possible prepositions we did not cover and yet may encode one of the subtype we have identified. For example, we have identified postverbal zai4 'at' and dao4 'to' as encoding the OBL_{LOCT} subtype, but whether elements such as shang4 'up', xia4 'down', and jin4 'into' should also be considered as such prepositions or they should be considered as verbs being part of a verb compound in sentences like the following:

1. a. Ta1 tiao4-xia4 zhuo1zi.
   he jump-down table
   He jumped off the table.

Secondly, other subtypes of OBL may still need to be recognized as encoded by other prepositions. For example, whether there should be an instrumental oblique function encoded by yong4 'with, use' and yi3 'with' needs to be established, and it also needs to be further studied whether there is enough evidence for wei4 'for' and ti4 'for' prepositional phrases to be regarded as subcategorized-for (Ernst 1989).

Furthermore, a more careful and detailed study of the interaction and correlation between our semantic classification and syntactic subcategorization, especially in the area of co-occurrence restrictions and requirements with functional elements of minor categories such as phrase markers, aspect markers, negation markers, and sentential particles, and others like modal adverbs, degree modifiers, and duration and range phrases, will be essential for a more comprehensive description of Mandarin verbs. Chao's (1968:665) classification provides a good basic model for this work, but one which needs to be extended.

Given the linguistic stratification of thematic structure, functional structure and constituent structure within LFG, for a more complete picture of Mandarin verbs, further study of the thematic structures of Mandarin verbs is needed to be incorporated into our current study of the subcategorization of grammatical functions and the syntactic encoding of grammatical functions. Since in LFG the specific derivational relations between verb classes are to be stated in thematic terms, researches of the application of Lexical Mapping Theory in Mandarin verbs are also required. Our study of verb subcategorization and Huang's (1989) application of Lexical Mapping Theory to Mandarin verbs as well as previous studies of case roles in Mandarin such as Teng (1975) and Li (1971) will prove to be valuable in these two areas of study: thematic structure of verbs and morpholexical rules of verb derivations.

Additionally, for practical considerations of natural language processing of real Chinese texts, semantic selectional restrictions of the subcategorized elements should also be studied and verb classes be further classified into subgroups accordingly.

Finally, since we have sporadically made arguments in reference to thematic relations in our analyses of several syntactic constructions, with the further investigation and a more systematic account of the link between thematic structures and subcategorization requirements of functions, some of

the controversial issues covered in this book such as the subcategorizability of topics, the syntactic encoding of OBJ2, status of ba3 and bei3 can then be further examined and perhaps more revealing and complete accounts may still be developed.

We conclude this book by voicing once again that any scientific pursuit cannot be an individual, isolated venture; rather, to be fruitful it has to be a collective and cooperative effort. The current linguistic scene of diversified theories and formalisms should not be a lasting phenomenon. We anticipate that the similarities among the different theories will receive more and more attention and a unified and dominant theory and formalism will eventually emerge in the linguistic science.

BIBLIOGRAPHY

Alsina, A. and A. Mchombo. 1988. Lexical Mapping in the Chichewa Applicative Construction. Paper presented at the 19th Annual African Linguistics Conference, Boston University, April 14-17, 1988.

Andrews, A. 1985. The Major Functions of the Noun Phrase. In E. Shopen (Ed.) Vol. I. 1985:242-281.

Bach, E. and R. Harms. 1968. Universals in Linguistic Theory. New York, NY: Holt, Rinehart, and Winston.

Baker, M. 1983. Objects, Themes, and Lexical Rules in Italian. In L. Levin, et al. (Eds.). 1-45.

Blake, B. 1989. Review of The Case for Lexicase (S. Starosta). Language 65.3:614-622.

Bresnan, J. 1978. A Realistic Transformational Grammar. In M. Halle, J. Bresnan, and G. Miller (Ed.), Linguistic Theory and Psychological Reality. Cambridge, Mass: MIT Press.

Bresnan, J. 1982 (Ed.). The Mental Representation of Grammatical Relations. Cambridge, Mass: MIT Press.

Bresnan, J. 1982a. The Passive in Lexical Theory. In Bresnan (Ed.). 1982. 3-86.

Bresnan, J. 1982b. Control and Complementation. In Bresnan (Ed.). 1982. 282-390.

Bresnan, J. 1987. Lectures on Lexical-Functional Grammar, 1987 Linguistic Institute, Stanford University.

Bresnan, J. 1987a. (Ed.). Lexical-Functional Grammar. Course Material for LI229, 1987 Linguistic Institute, Stanford University.

Bresnan, J. and J. Kanerva. 1989. Locative Inversion in Chichewa: A Case Study of Factorization in Grammar. Linguistic Inquiry 20.1:1-50. Also appeared as CSLI Report No. CSLI-88-131, Center for the Study of Language and Information, Stanford University, Stanford, CA.

Carbonell J. and M. Tomita. (1987). Knowledge-based Machine Translation, the CMU Approach. In R. Nirenburg, (Ed.). 1987. 68-89.

Carrier-Duncan, J. 1985. Linking of Thematic Roles in Derivational Word Formation. Linguistic Inquiry 12.1:1-34.

Chafe, W. 1970. Meaning and the Structure of Language. Chicago: University of Chicago Press.

Chafe, W. L. 1976. Givenness, Contrastiveness, Definiteness, Subjects, Topics, and Point of View. In C. Li (Ed.) 1976:25-55.

Chang, C. 1989. Compounds in Mandarin Chinese: with special emphasis on resultative verb compounds. University of Hawaii Department of Linguistics Working Papers 20:59-84.

Chang, C. 1990. Complex Verbs and Argument Structure: Interaction between Syntax and Morphology. Paper Presented at the 2nd Northeast Conference on Chinese Linguistics, University of Pennsylvania, Philadelphia, May 4-6, 1990.

Chang, C. 199o. Thematic Structure and Verb-copying in Mandarin Chinese. Paper Presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, University of Texas at Arlington, October 5-7, 1990.

Chang, D. T. 1986. Design of a Conceptual Scheme for Chinese Language Processing. Proceedings, 1986 International Conference on Chinese Computing. 379-390.

Chang, L.-L., et al. 1987. Classification and Co-occurrence Restrictions in Chinese Simple Noun Phrases. Proceedings, 1987 International Conference on Chinese and Oriental Languages Computing. 107-110.

Chang, L.-L., et al. 1988. A Classification of Chinese Verbs for Language Parsing. Proceedings, 1988 International Conference on Chinese and Oriental Languages Computing. 414-417.

Chang, R. 1977. Coverbs in Spoken Chinese. Ph.D. Dissertation. University of Florida.

Chao, Y. 1968. A Grammar of Spoken Chinese. Berkeley and Los Angeles: University of California Press.

Chen, W. 1989. Functional Uncertainty and Topicalization in Mandarin Chinese. M.A. thesis, National Tsing Hua University.

Cheng, R. 1979. Taiwanese u and Mandarin you. Proceedings of 1979 Conference of Language Teaching in the Asian and Pacific Regions. 141-180.

Cheng, R. 1983. Focus Devices in Mandarin Chinese. In Hanyu Jufa Yuyixue Lunji (Studies in Chinese Syntax and Semantics, Universe and Scope: Presupposition and Quantification in Chinese), T. Tang, R. Cheng, and Y. Li. (Eds.). 50-102.

Cheng, R. 1990. Tense in Taiwanese. (In Chinese.) Proceedings of the First International Symposium on Chinese Languages and Linguistics (1990). 59-63.

CKIP (Chinese Knowledge Information Group). 1989. Analysis of Chinese Word Classes. (Revised Edition). Taipei: Academia Sinica.

Chomsky, N. 1957. Syntactic Structures. The Hague: Mouton.

Chomsky, N. 1965. Aspects of the Theory of Syntax. Cambridge, Mass.: The MIT Press.

Chomsky, N. 1981. Lectures on Government and Binding. Dordrecht: Foris.

Chomsky, N. 1982. Some Concepts and Consequences of the Theory of Government and Binding. Cambridge, Mass: MIT Press.

Chomsky, N. 1986. Knowledge of Language. New York, NY: Preager Publishers.

Chu, C. 1979. Definiteness, Presupposition, Topic and Focus in Chinese. Paper presented at 1979 MLA-CLTA, San Francisco.

Chu, C. 1984. Chinese Word Order and Its Change. Yuyan Yanjiu, 1:115-127.

Correa, N. 1987. An Attribute-Grammar Implementation of Government-binding Theory. Proceedings of 25th Annual Meeting of the ACL.

Crain, S. and J. Fodor. 1985. How Can Grammars Help Parsers? In D. Dowty, et al (Eds.) 1985:95-128.

Ding, S., S. Lu, R. Li, D. Sun, X. Guan, Q. Fu, S. Huang, and Z. Chen. 1979. Xiandai Hanyu Yufa Jianghua (Discussions on Modern Chinese Grammar). Society of Chinese Language and Literature. Beijing: Shangwu Publishing Co.

Dowty, D. 1988. Thematic Proto-Roles. Subject Selection and Lexical Semantic Defaults. Paper presented at the 1987 LSA Annual Meeting.

Dowty, D., L. Karttunen, and M. Zwicky. 1985. Natural Language Parsing. Cambridge: Cambridge University Press.

Ernst, T. 1989. On Verb Subcategorization in Chinese. Proceedings of the Third Ohio State Conference on Chinese Linguistics. Bloomington: IULC. 116-140.

Fillmore, C. 1968. The Case for Case. In E. Bach and R. Harms (Eds.) 1968:1-88.

Ford, M., J. Bresnan and R. Kaplan. 1982. A Competence-Based Theory Syntactic Closure. In J. Bresnan (Ed.) 1982:727-796.

Fraser, N. 1989. Review of The Case for Lexicase (S. Starosta). Computational Linguistics 15.2:114-115.

Gazdar, G., E. Klein, G. Pullum, and I. Sag. 1985. Generalized Phrase Structure Grammar. Cambridge, Mass: Harvard University Press.

Gebruers, R. 1989. Book Review: From Syntax to Semantics: Insights from Machine Translation. Machine Translation 4.3:231-238.

Grimshaw, J. 1982. Subcategorization and Grammatical Relations. In A. Zaenen (Ed.), Subjects and Other Subjects: Proceedings of the Harvard Conference on the Representation of Grammatical Relations. Bloomington: IULC. 35-55.

Grishman, R. 1986. Computational Linguistics: An Introduction. Cambridge University Press.

Grosz, B., K. Jones, and B. Webber. 1986. (Ed.). Readings in Natural Language Processing. Los Angeles: Morgan Kaufmann Publishers, Inc.

Gruber, J.S. 1965. Studies in Lexical Relations. Ph.D. Dissertation. MIT.

Halvorsen, P.-K. Semantics for Lexical Functional Grammar. Linguistic Inquiry 14:567-615.

Harris, M. 1985. Introduction to Natural Processing. Reston, Virginia: Reston Publishing Company.

Hellwig, p. 1986. Dependency Unification Grammar. Proceedings of COLING 1986. 196-198.

Her, O. 1985-1986. To Dispense with OV Word Order in Mandarin Chinese: A Lexicase Analysis. Papers in East Asian Languages 3:17-47.

Her, O. 1987. The ECS Machine Translation System: an Overview.Technical report and manual, ECS, Inc., Provo, Utah.

Her, O. 1989. An LFG-Based English-Chinese Machine Translation System. Proceedings 1989 International Symposium on Chinese Text Processing, March 16-17, 1989, Boca Raton, Florida. 8.3-8.7.

Her, O. 1989a. Chinese Verb Subcategorization in a Variant Lexical Functional Grammar. Paper presented at the 22nd International Conference on Sino-Tibetan Languages and Linguistics, October 5-8, 1989, Honolulu, Hawaii.

Her, O. 1989b. An LFG Account of Chinese BEI Sentences. Journal of the Chinese Language Teachers Association XXIV.3:67-89.

Her, O. 1989c. Focus Devices in Chinese Revisited. Ms. ECS, Inc.

Her, O. 1989d. The Historical Development of BA/JIANG Sentences: More Evidence from Bianwen and Chuanqi. Paper Presented at the First

Northeast Conference on Chinese Linguistics, May 5-7, 1989, the Ohio State University, Columbus, Ohio.

Her, O. 1990. On the Verb YOU in Mandarin Chinese. Paper Presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, University of Texas at Arlington, October 5-7, 1990.

Horrocks, G. 1987. Generative Grammar. London & New York: Longman.

Hou, J. 1979. Grammatical Relations in Chinese. Ph.D. Dissertation. University of Southern California.

Hu, C. 1989. Long Distance Dependencies in Chinese Relative Clauses: An LFG Approach. M.A. thesis, National Tsing Hua University.

Huang, C. 1985. On Pseudo-Possessive NPs in Mandarin Chinese. Cornell Working Papers in Linguistics 7:129-148.

Huang, C. 1987. Mandarin NP de: A Comparative Study of Current Grammatical Theories. Ph.D. Dissertation. Cornell University

Huang, C. 1988. A Unification-based Analysis of Lexical Discontinuity.Paper presented at the 4th International Workshop on East Asian Languages and Linguistics, June 15-16, Paris.

Huang, C. 1989. Mandarin Chinese and the Lexical Mapping Theory. Paper presented at the XXII International Conference on Sino-Tibetan Languages and Linguistics, Honolulu, Hawaii, October 6-8, 1989.

Huang, C. 1989a. Subcategorized Topics in Mandarin Chinese. Paper presented at the 1989 CLTA Annual Meeting, November 17-19, Boston, Mass.

Huang, C. 1989b. A Brief Introduction to the Lexical Functional Grammar. (In Chinese). Proceedings of ROCLING II & R.O.C. Computational Linguistics Workshops II (1989). 73-94.

Huang, C. 1990. Review of Functionalism and Chinese Grammar (ed. by J. Tai and F. Hsueh) Journal of Chinese Linguistics 18.2:318-334.

Huang, C. 1990a. Mandarin Double Object Construction and Morpholexical Rules. Paper Presented at the 23rd International Conference on Sino-Tibetan Languages and Linguistics, University of Texas at Arlington, October 5-7, 1990.

Huang, C. and K. Chen (Eds.). 1989. Proceedings of ROCLING II & R.O.C. Computational Linguistics Workshops II (1989). Taipei: Academia Sinica.

Huang, C. and L. Mangione. 1985. A Reanalysis of DE: Adjuncts and Subordinate Clauses. In Proceedings of the 4th Annual West Coast

Conference on Formal Linguistics, Stanford Linguistics Association:Stanford.

Huang, J. 1982. Logical Relations in Chinese and the Theory of Grammar. Ph.D. Dissertation. MIT.

Huang, J. 1983. Move Wh in a Language without Wh Movement. Linguistic Review 1:369-416.

Huang, J. 1988. Wo pao de kuai and Chinese Phrase Structure. Language 64.274-311.

Huang, S. 1966. Subject and Object in Mandarin. POLA 13.25-103.

Huang, X. 1986. A Bidirectional Grammar for Parsing and Generating Chinese. Proceedings, 1986 International Conference on Chinese Computing. 46-52.

Isabelle, P. and E. Macklovitch. 1986. Transfer and MT Modularity. Proceedings of Coling, 1986, Bonn. 115-117.

Ishikawa, A. 1985. Complex Predicates and Lexical Operations in Japanese. Ph.D. Dissertation. Stanford University.

Jackendoff, R. 1987. The Status of Thematic Relations in Linguistic Theory. Linguistic Inquiry 18:369-411.

Kaplan, R. 1987. Three Seductions of Computational Psycholinguistics. In P. Whitelock, et al (Eds.). 149-188.

Kaplan, R. 1989. The Formal Architecture of Lexical-Functional Grammar. In Proceedings of ROCLING II. 1-18.

Kaplan, R. and J. Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In J. Bresnan (Ed.), The Mental Representation of Grammatical Relations. Cambridge, Mass.: MIT Press. 173-281.

Kaplan, R. and J. Maxwell. 1988. Constituent Coordination in Lexical Functional Grammar. In Proceedings of COLING 1988, Budapest.

Kaplan, R., K. Netter, J. Wedekine, and A. Zaenen. 1989. Translation by Structural Correspondences. In Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics, 10-12 April 1989, University of Manchester.

Kaplan, R. and A. Zaenen. 1987. Functional Uncertainty. CSLI Monthly, January 1987. CSLI, Stanford University.

Kaplan, R. and A. Zaenen. 1989. Functional Precedence and Constituent Structure. In Proceedings of ROCLING II. Taipei: Academia Sinica. 19-40.

Kaplan, R. and A. Zaenen. 1989a. Long Distance Dependencies, Constituent Structure, and Functional Uncertainty. In M. Baltin & A. Kroch (Eds.), Alternative Conceptions of Phrase Structure. Chicago: University of Chicago Press. 17-42.

Kay, M. 1986. Parsing in Functional Unification Grammar. In Grosz, J. et al 1986. (Ed.).

Keenan, E. 1985. Passive in the World's Languages. In E. Shopen (Ed.) Vol. I. 1985:242-281.

King, M. 1983. Parsing Natural Language. London: Academic Press.

Kudo, I. and H. Nomura. 1986. Lexical-Functional Transfer: A Transfer Framework in a Machine Translation System based on LFG. Proceedings of Coling 1986, Bonn. 112-114.

Levin, L. 1987. Toward a Linking Theory of Relation Changing Rules in LFG. CSLI Report No. CSLI-87-115, Center for the Study of Language and Information, Stanford University, Stanford, CA.

Levin, L. 1990. Syntactic Theory and Grammar Design for Machine Translation. Paper Presented at The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 11-13 June 1990, Linguistic Research Center, University of Texas at Austin.

Levin, L., M. Rappaport, and A. Zaenen. 1987. (Eds.) Papers in Lexical-Functional Grammar. Bloomington, IN.: Indiana University Linguistic Club.

Li, C. 1975. (Ed.) Word Order and Word Order Change. University of Texas Press.

Li, C. 1976. (Ed.) Subject and Topic. New York: Academic Press.

Li, C. and S. Thompson. 1974. An Explanation of Word Order Change SVO->SOV. Foundations of Language. 12.201-214.

Li, C. and S. Thompson. 1975. The Semantic Function of Word Order. In C. Li (Ed.) 1975.

Li, C. and S. Thompson. 1976. Subject and Topic: a New Typology of Language. In Li 1976:457-89.

Li, C. and S. Thompson. 1981. Mandarin Chinese: A Functional Grammar. Berkeley:University of California Press.

Li, M. D. 1985. Reduction and Anaphoric Relations in Chinese. Ph.D. Dissertation. University of California at San Diego.

Li, Y. 1971. An Investigation of Case in Chinese Grammar. Seton Hall University Press.

Li, Y. 1972. Problems of Subject, Object , etc., in Chinese. in C. Tang et al (Eds.) Papers in Linguistics in Honor of A. A. Hill. 109-132.

Li, Y. 1972. Sentences with Be, Exist, and Have in Chinese. Language 48.3.

Li, Y. 1974. What Does "Disposal" Mean? Journal of Chinese Linguistics 2.2.200-218.

Li, Y. 1980. The Historical Development of the Coverb and the Coverbial Phrase in Chinese. Journal of Chinese Linguistics 8:1.273-293.

Li, Y. 1988. Hanyu Zhubin Guannian de Zai Tantao (Subject and Object in Chinese Revisited). Proceedings of the 2nd International Conference of Teaching Chinese. Beijing: Beijing Language Institute Publishing Co. 298-304.

Lin, L.-J. 1985. A Syntactic Analysis System for Chinese Sentences. Master Thesis. Department of Electrical Engineering, National Taiwan University, R.O.C.

Lin, L.-J., et al. 1986. SASC: A Syntactic Analysis System for Chinese Sentences. Proceedings, 1986 International Conference on Chinese Computing. 29-32.

Lin, L., J. Huang, K. Chen and L. Lee. 1986. A Chinese Natural Language Processing System Base upon the Theory of Empty Categories. Proceedings AAAI-86 Fifth Conference on Artificial Intelligence, Volume II. 1059-1062.

Lindsey, F. 1986. Pronoun Interpretation Rules: A Guided Tour. Ms., Department of Linguistics, University of Hawaii.

Liu, H. 1986. A Categorial Grammar Analysis of Chinese Separable Compounds and Phrases. Ph.D. Dissertation. Arizona State University.

Lu, S. 1984. Xiandai Hanyu Ba Bai Ci (Eight Hundred Words in Modern Chinese). Hong Kong: Shangwu Publishing Co.

Ma, L. 1985. The Classical Notion of Passive and the Mandarin 'bei'. Ms., Department of Linguistics, Stanford University.

McKeown, K. 1985. Text Generation. Cambridge: Cambridge University Press.

Mo, R. 1990. On Chinese Subcategorized Topics. Unpublished M.A. Thesis, Fu-jen Catholic University.

Mohanan, K. 1983. Functional and Anaphoric Control. Linguistic Inquiry 14.4:641-674.

Mohanan, K. 1983a. Move NP or Lexical Rules? Evidence from Malayalam Causativization. In L. Levin, et al. (Eds.). 47-111.

Lee, N. 1989. Complementation in Japanese: A Lexicase Analysis. Ph.D. Dissertation. University of Hawaii.

Moortgat, M. 1988. Categorial Investigation: Logical and Linguistic Aspects of the Lambek Calculus. Dordrecht, Holland: Foris Publication.

Nirenburg, S. (Ed.). 1987. Machine Translation: Theoretical and Methodological Issues. Cambridge University Press.

Pagotto, L. 1985. Missing Nominals in Wh-questions and Infinitives: A Lexicase View. University of Hawaii Working Papers in Linguistics 17.1:23-76.

Pagotto, L. 1985a. On Impersonal Verbs in English. University of Hawaii Working Papers in Linguistics 17.2:1-70.

Perlmutter, D. 1978. Impersonal Passives and the Unaccusative Hypothesis. Proceedings of the 4th Berkeley Linguistic Society. 157-189.

Pollard, C. and I. Sag. 1987. Information-based Syntax and Semantics. Vol. I. Fundamentals. CSLI Lecture Notes No. 13. Center for the Study of Language and Information, Stanford University, Stanford, CA.

Radford, A. 1981. Transformational Syntax. Cambridge: Cambridge University Press.

Riemsdijk, H. and E. Williams. 1986. Introduction to the Theory of Grammar. Cambridge, Mass: MIT Press.

Ross, C. 1981. On the Functions of Mandarin de. Journal of Chinese Linguistics 11 (2).

Sadler, L., I. Crookston, and D. Arnold. 1990. LFG and Translation. Paper Presented at The Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages, 11-13 June 1990, Linguistic Research Center, University of Texas at Austin.

Schachter, P. 1985. Parts-of-speech Systems. In E. Shopen (Ed.) Vol. I. 1985:242-281.

Sells, P. 1985. Lectures on Contemporary Syntactic Theories. Stanford, CA: CSLI, Stanford University.

Sheu, Y. 1987. Chinese Morphosyntax. Paper presented at the 1987 International Conference on Sino-Tibetan Languages and Linguistics.

Shieber, S. 1986. Introduction to Unification-based Approaches to Grammar. Stanford, CA: CSLI, Stanford University.

Shieber, S. 1987. Separating Linguistic Analyses from Linguistic Theories. In P. Whitelock, et al (Eds.). 1-36.

Shiu, Y. 1989. Mandarin Sentential Particles and Particle Questions. M.A. thesis, National Tsing Hua University.

Shiu, Y. and C. Huang. 1989. A Unification-based Approach to Mandarin Questions. Proceedings of ROCLING II. Nankang: Academia Sinica. 311-332.

Shopen, T. (Ed.) 1985. Language Typology and Syntactic Description. Cambridge: Cambridge University Press.

Simpson, J. 1983. Resultatives. In L. Levin, et al. (Eds.). 143-157.

Slocum, J. 1985. A Survey of Machine Translation: Its History, Current Status, and Future Prospects. Computational Linguistics 11:1-17.

Slocum, J. 1986. Is MT Linguistics? Letter to the Editor. Computational Linguistics 12 (2), p.125.

Starosta, S. 1982. Case Relations, Perspectives, and Patient Centrality. University of Hawaii Working Papers in Linguistics 14.1:1-33.

Starosta, S. 1985. Mandarin Case Marking. Journal of Chinese Linguistics 13.2:215-266.

Starosta, S. 1985a. The End of Phrase Structure as We Know It. LAUDT Series A, Paper no. 147, Duisburg, Linguistic Association - University of Duisburg.

Starosta, S. 1985b. The Great AUX Cataclysm. University of Hawaii Working Papers in Linguistics, 17.2:95-114.

Starosta, S. 1986. GB minus VP Equals UG. MS. University of Hawaii.

Starosta, S. 1987. Generative Grammar: The Hypothetico-deductive Science of Language, Taipei: The Crane Publishing Co.

Starosta, S. 1988. The Case for Lexicase. London: Pinter Publishers.

Starosta, S and H. Nomura. 1984. Lexicase and Japanese Language Processing. Technical report No.21887, Musashino Electrical Communication Laboratory, Nippon Telegraph and Telephone Public Corporation.

Starosta, S. and H. Nomura. 1986. Lexicase Parsing: A Lexicon-driven Approach to Syntactic Analysis. in M. Nagao (Ed.). Proceedings of the Eleventh International Conference on Computational Linguistics (COLING '86), Bonn: University of Bonn. 127-132.

Su, K.-Y., et al. 1987. A Powerful Language Processing System for English-Chinese Machine Translation. Proceedings,1987 International Conference on Chinese and Oriental Languages Computing. 260-264.

Sun, C. and T. Givon. 1985. On the So-called SOV Word Order in Mandarin Chinese. Language 61.329-351.

Tai, J. 1985. Temporal Sequence and Chinese Word Order. In J. Haiman (Ed.). Iconicity in Syntax. Amsterdam: John Benjamins Publishing Co.

Tan, F. 1987. The Predicate Argument Structure of Bei. Berkeley Linguistics Society: Proceedings of the Thirteenth Annual Meeting February 14-16, 1987: General Session and Parasession on Grammar and Cognition. Berkeley: Berkeley Linguistics Society. 94-99.

Teng, S. 1975. A Semantic Study of Transitivity Relations in Chinese. Taipei: Student Book Co.

Teng, S. 1979. Remarks on Cleft Sentences in Chinese. Journal of Chinese Linguistics 7.1:101-114.

Tsai, Y.-S. 1987. Parsing Chinese Conjunction Using SASC. Proceedings, 1987 International Conference on Chinese and Oriental Languages Computing. 94-99.

Tsao, F. 1978. Subject and topic in Chinese, in Cheng et al (Eds.) Proceedings of the Symposium of Chinese Linguistics, 1977 Linguistic Institute of LSA. Taipei: Student Book Co.

Tsao, F. 1979. A Functional Study of Topic in Chinese. Taipei: Student Book Co.

Tsao, F. 1987. A Topic-Comment Approach to the BA Construction. Journal of Chinese Linguistics 15.1:1-54.

Wang, L. 1947. Zhongguo Xiandai Yufa (Modern Chinese Grammar). Shanghai:China Bookstore.

Wang, W. 1963. Some Syntactic Rules for Mandarin. The Ohio State University Project on Linguistic Analysis, No. 3. Ohio State University, Columbus, Ohio.

Wedekind, J. 1986. A Concept of Derivation for LFG. Proceedings COLING 1986. 487-489.

Wescoat, M. 1987. Practical Instructions for Working with the Formalism of Lexical Functional Grammar. In J. Bresnan (Ed.). Lexical-Functional Grammar. Course Material for LI229, 1987 Linguistic Institute, Stanford University.

Whitelock, P., M. Wood, H. Somers, R. Johnson, and P. Bennet. (Eds.) 1987. Linguistic Theory and Computer Application. London: Academic Press.

Winograd, T. 1983. Language as a Cognitive Process (Volume 1: Syntax). Addison-Wesley Publishing Company.

Xu, L. and T. Langendeon. 1985. Topic Structure in Chinese. Language 61.1:1-27.

Yang, S. 1989. An Analysis of the structures of shi...de Sentences. (In Chinese) Paper presented at the 1989 International Conference on Sino-Tibetan Languages and Linguistics, University of Hawaii.

Yang, Y.-M., et al. 1984. Use of Heuristic Knowledge in Chinese Language Analysis. Proceedings, COLING 1984. 222-225.

Yang, Y.-M. 1985. Studies on an Analysis System for Chinese Sentences. Ph.D. Dissertation. Department of Information Sciences, Kyoto University, Kyoto, Japan.

Zaenen, A. 1983. On Syntactic Binding. Linguistic Inquiry 14.3:469-504.

Zaenen, A. and J. Maling. 1983. Passive and Oblique Case. In L. Levin, et al. (Eds.). 159-191.

Zhang, S. 1989. Imperative Revisited: Its Nature and Sentential Category. Proceedings of Chicago Linguistic Society's 25th Annual Meeting. CLS Press.

Zhang, Y. 1983. Review Article. Journal of Chinese Language Teachers Association 18.2:93-107.

Zheng, Y. and Y. Chen. 1989. Structural Types of Adjective Reduplication in Fuzhou Dialect. (In Chinese.) Paper presented at the 1989 International Conference on Sino Tibetan Languages and Linguistics, University of Hawaii.

Zhou, J.-Y. and S.-K. Chang. 1986. A Methodology for Deterministic Chinese Parsing. Computer Processing of Chinese and Oriental Language 1986:139-161

APPENDIX
FEATURE INHERITANCE ENTRIES

```
FI-V-STATE:
        [ FS [ ACTIVE -
               PROCESS -
             ]
        ]

FI-V-PROCESS:
        [ FS [ ACTIVE -
               PROCESS +
             ]
        ]

FI-V-ACTION:
        [ FS [ ACTIVE +
               PROCESS -
             ]
        ]

FI-V-ACTPRO:
        [ FS [ ACTIVE +
               PROCESS +
             ]
        ]

FI-V-1:
        [ CAT V
          FS    [ PRED < SUBJ > ]
        ]
```

FI-V-1-2:

```
[ FS [ SUBJ [FORM OPT
             PLACE +
            ]
      ]
]
```

FI-V-2:

```
[ CAT V
  FS   [ PRED <SUBJ , OBL_LOCT> ]
  ( ↑ OBL_LOCT PCASE ) =c LOCT
]
```

FI-V-3:

```
[ CAT V
  FS   [ PRED <OBL_GOAL , SUBJ> ]
  ( ↑ OBL_GOAL PCASE ) =c GOAL
]
```

FI-V-4:

```
[ CAT V
  FS   [ PRED <SUBJ , OBJ> ]
]
```

FI-V-4-2:

```
[ (↑ OBJ PLACE) =c +
]
```

FI-V-4-ID:

```
[ CAT V
  FS   [PRED <SUBJ , OBJ @>
        OBJ [BACKGROUND -
             FORM ANY
            ]      ]
    FI-V-STATE
]
```

FI-V-4-5:

```
[ FS [ OBJ [BACKGROUND +
                FORM ANY
              ]
       ]
]
```

FI-V-5:

```
[ CAT V
  FS   [ PRED <SUBJ , OBL_THME> ]
  (↑ OBL_THME PCASE) =c THME
]
```

FI-V-6:

```
[ CAT V
  FS   [ PRED <SUBJ , OBL_THME , OBJ> ]
  (↑ OBL_THME PCASE) =c THME
]
```

FI-V-7:

```
[ CAT V
  FS   [ PRED <SUBJ , OBL_GOAL , OBJ> ]
    ( ↑ OBL_GOAL PCASE ) =c GOAL
]
```

FI-V-8:

```
[ CAT V
  FS   [ PRED <SUBJ , OBJ , OBL_LOCT> ]
  (↑ OBL_LOCT PCASE) =c LOCT
]
```

FI-V-9:

```
[ CAT V
  FS   [ PRED <SUBJ , OBL_THME , OBL_LOCT> ]
  (↑ OBL_THME PCASE) =c THME
  (↑ OBL_LOCT PCASE) =c LOCT
]
```

FI-V-10:
[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ , OBL_{LOCT}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{LOCT} PCASE) =$_c$ LOCT
(↑ OBJ FORM) =$_c$ NONE
]

FI-V-11:
[CAT V
FS [PRED <SUBJ , NCOMP>]
(↑ NCOMP SUBJ) = (↑ SUBJ)
]

FI-V-11-1:
[(↑ NCOMP PRED) = <SUBJ @>]

FI-V-12:
[CAT V
FS [PRED <SUBJ , XCOMP>]
]
FI-V-12-1:
[(↑ SUBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODALITY) =$_c$ NONE
(↑ XCOMP ASPECT) =$_c$ NONE
]

FI-V-12-2:
[CAT V
FS [PRED <SUBJ @ , XCOMP>
XCOMP [SUBJ OPT]
]
(↑ SUBJ) = (↑ XCOMP+ {OBJ OBJ2})
(↑ XCOMP MODALITY) =$_c$ NONE
(↑ XCOMP ASPECT) =$_c$ NONE
]

FI-V-12-3:

```
[ FS [ ASPECT NONE
       MODALITY ANY
     ]
]
```

FI-V-12-3-SPEAKER:

```
[ (↑ SUBJ) = (↑ XCOMP SUBJ)
  (↑ XCOMP ASPECT) =c NONE
]
```

FI-V-12-4:

```
[ (↑ SUBJ) = (↑ XCOMP SUBJ)
  (↑ XCOMP MODALITY) =c NONE
  (↑ ASPECT) =c NONE
 (↑ XCOMP ASPECT) = (↑ ASPECT)
]
```

FI-V-13:

```
[ CAT V
  FS    [ PRED <SUBJ , SCOMP> ]
]
```

FI-V-13-2:

```
[ (↑ SCOMP Q) =c + ]
```

FI-V-14:

```
[ CAT V
  FS    [ PRED <SUBJ , OBJ , OBJ2> ]
]
```

FI-V-14-2:

```
[ CAT V
    FS [ PRED <SUBJ , OBJ , OBJ2 @> ]
]
```

FI-V-15:

[CAT V
FS [PRED <SUBJ , OBJ , OBL_{BNFC}>]
(↑ OBL_{BNFC} PCASE) =$_c$ BNFC
]

FI-V-16:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ , OBJ2>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBJ2 FORM) =$_c$ NONE
]

FI-V-17:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBJ , OBL_{BNFC}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{BNFC} PCASE) =$_c$ BNFC
(↑ OBJ FORM) =$_c$ NONE
]

FI-V-18:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{BNFC}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{BNFC} PCASE) =$_c$ BNFC
]

FI-V-19:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{GOAL}>]
(↑ OBL_{THME} PCASE) =$_c$ THME
(↑ OBL_{GOAL} PCASE) =$_c$ GOAL
]

FI-V-20:

[CAT V
FS [PRED <SUBJ , OBL_{THME} , OBL_{GOAL} , OBJ>]
(↑ OBL_{THME} PCASE) $=_c$ THME
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
(↑ OBJ FORM) $=_c$ NONE
]

FI-V-21:

[CAT V
FS [PRED <SUBJ , OBJ , XCOMP>]
]

FI-V-21-1:

[(↑ OBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP ASPECT) $=_c$ NONE
(↑ XCOMP MODAL) $=_c$ NONE
]

FI-V-21-2:

[(↑ OBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODAL) = NONE
(↑ ASPECT) $=_c$ NONE
(↑ XCOMP ASPECT) = (↑ ASPECT)
]

FI-V-21-3:

[(↑ SUBJ) = (↑ XCOMP SUBJ)
(↑ XCOMP MODALITY) $=_c$ NONE
]

FI-V-21-4:

[(↑ OBJ) = (↑ XCOMP SUBJ)
(↑ SUBJ) = (↑ XCOMP+ {OBJ OBJ2})
(↑ XCOMP MODALITY) $=_c$ NONE
(↑ XCOMP ASPECT) $=_c$ NONE
]

FI-V-22:

[CAT V
FS [PRED <SUBJ , OBJ , SCOMP>]
]

FI-V-23:

[CAT V
FS [PRED <SUBJ , OBL_{GOAL} , XCOMP>
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
(↑ OBL_{GOAL}) = (↑ XCOMP SUBJ)
]

FI-V-23-2:

[(↑ XCOMP Q) $=_c$ +]

FI-V-24:

[CAT V
FS [PRED <SUBJ , OBL_{GOAL} , SCOMP>]
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
]

FI-V-25:

[CAT V
FS [PRED <SUBJ , OBJ , OBJ2 , SCOMP>]
]

FI-V-26:

[CAT V
FS [PRED <SUBJ , OBL_{GOAL} , OBJ , SCOMP>]
(↑ OBL_{GOAL} PCASE) $=_c$ GOAL
(↑ OBL_{GOAL} PFORM) $=_c$ { 'gen1' , 'he2' , 'tong2' }