

TOPIC and FOCUS in Cantonese: An OT-LFG Account

by

Fung Suet Man

馮雪雯

B.A. H.K.

A thesis submitted in partial fulfilment of the requirements for
the degree of Master of Philosophy
at the University of Hong Kong
July 2007

Abstract of thesis entitled

“TOPIC and FOCUS in Cantonese: An OT-LFG Account”

Submitted by

Fung Suet Man

for the degree of Master of Philosophy
at the University of Hong Kong
in July 2007

This thesis analyzes topicalization in Cantonese, investigating TOPIC and FOCUS, the two functions which a topicalized phrase may bear, within the frameworks of Lexical-Functional Grammar (LFG) and Optimality-Theoretic Lexical-Functional Grammar (OT-LFG). The study proposes functional uncertainty equations and constraint hierarchies for the phenomenon based on the results of a questionnaire survey.

The study starts by looking at the terms ‘TOPIC’ and ‘FOCUS’, which refer respectively to topicalized discourse topic and topicalized discourse focus conforming to the definition of topicalization in LFG. While the notion of ‘topic’ has been widely discussed in the literature on Chinese and the definition also applies to Cantonese TOPIC, no such definition exists which satisfactorily describes Cantonese FOCUS. The information features of this function are investigated in this study and it is proposed that the [+New] feature is what licenses FOCUS topicalization in the language and FOCUS is defined as a topicalized phrase which contains some [+New] information.

TOPIC and FOCUS in Cantonese are examined within the LFG framework. A functional approach is given, establishing two functional uncertainty equations for the phenomenon which specify the legitimate grammatical functions to which TOPIC

and FOCUS can be identified. It is shown that the two grammaticized discourse functions display an asymmetry in that they have different bottom functions. In order to complement the functional approach, an informational account is provided, adopting an independent i-structure. It is argued that TOPIC and FOCUS, though bearing discourse information, should still be included in the f-structure and it is proposed how a construction with only the PRED being topicalized, excluding its arguments or attributes, should be represented in the f-structure.

An OT-LFG account is developed for topicalization in Cantonese. New constraints including [+New]-FOCUS \wedge FOCUS-L and *TOPICALIZE-FC are introduced in order to account for the properties of Cantonese topicalization. The constraints are then ranked against each other, resulting in one subhierarchy for TOPIC topicalization and two subhierarchies for FOCUS topicalization.

(317 words)

Declaration

I declare that this thesis represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma, or other qualifications.

Signed _____
Fung Suet Man

Acknowledgements

I would first like to express my gratitude to my supervisor, Dr. Adams Bodomo, for granting me the opportunity for pursuing the M.Phil. degree under his supervision. Whenever I encounter problems in my research, he was willing to discuss them with me and gave me many insightful comments. His support and advice have indeed helped me a lot throughout my studies.

I am obliged to the University of Hong Kong for offering me the postgraduate studentship, without which I would not have been able to continue my studies. I am also grateful to the University for offering me the CRCG Conference Grant which enabled me to give a presentation at the Fifth Cambridge Postgraduate Conference in Language Research in 2007.

I would like to thank Olivia Lam, Carmen Lee, Yanhong Pan, Dr. Charles Marfo, Dr. Sam Mchombo and Dr. Stephen Matthews for their precious opinions and comments on my work. My special thanks go to Olivia who, being my tutor when I was an undergraduate student, had aroused my interest in syntax and encouraged me to do further research.

I am most grateful to my parents for their love over the years. They let me pursue another degree as I desired and are always by my side to give me their support and care whenever I am confused and frustrated.

My gratitude goes to my sister, Jollus, and my friends, especially Hard, Jason and Yvonne, who offered me their patience and encouragements when I needed them most.

I have to thank my informants for spending their time on my questionnaires. Their precious opinions were indeed a great contribution to my thesis.

I am very much indebted to the brothers and sisters from the church, especially Zoe, Lily and Gloria, who are always around to fellowship and pray with me.

My greatest thanks go to the Lord who always loves me and has been guiding me through every circumstance.

I am solely responsible for any errors in this thesis.

Contents

Declaration	i
Acknowledgements	ii
Table of Contents	iii
Abbreviations	vi

Chapter 1 Introduction

1.1 Literature Review and Objectives of Study	1
1.2 The Cantonese Language	2
1.3 The Frameworks: LFG and OT-LFG	3
1.3.1 Lexical-Functional Grammar (LFG)	3
1.3.1.1 C-Structure	4
1.3.1.2 F-structure	5
1.3.1.3 A-Structure	8
1.3.1.4 I-Structure	9
1.3.1.5 Summary	9
1.3.2 Optimality Theory (OT)	10
1.3.2.1 The Input	10
1.3.2.2 The Generator (GEN)	10
1.3.2.3 The Evaluator (EVAL)	10
1.3.2.4 Summary	12
1.3.3 OT-LFG	12
1.4 Scope of Research	13
1.5 Research Questions	14
1.6 Outline of Thesis	14

Chapter 2 TOPIC and FOCUS in Cantonese

2.1 Topicalization	15
2.2 Discourse Topic/Focus vs. Syntactic Topic/Focus	17
2.3 Topic in Cantonese	18
2.3.1 Definition of TOPIC in Cantonese	18
2.3.2 TOPIC vs. E-TOP	19

2.4 FOCUS in Cantonese	20
2.4.1 Methodology	21
2.4.2 Information Features of FOCUS: Towards a Definition	22
2.5 Summary	27

Chapter 3 Topicalization in Cantonese: An LFG Perspective

3.1 C-structure Configuration	28
3.2 A Functional Approach to Topicalization	33
3.2.1 Functional Uncertainty	34
3.2.2 Functional Uncertainty Equations for Cantonese Topicalization	39
3.2.2.1 Methodology	39
3.2.2.2 Bottom Functions of TOPIC	40
3.2.2.2.1 The SUBJ Function	40
3.2.2.2.2 The OBJ Function	41
3.2.2.2.3 The OBJ _θ Function	41
3.2.2.2.4 The XCOMP Function	42
3.2.2.2.5 The OBL Function	42
3.2.2.2.6 The ADJ Function	45
3.2.2.2.7 The PRED Function	46
3.2.2.2.8 The Illegitimate Bottom Function: COMP	47
3.2.2.3 Bottom Functions of FOCUS	48
3.2.2.3.1 The SUBJ Function	49
3.2.2.3.2 The OBJ Function	49
3.2.2.3.3 The OBJ _θ Function	49
3.2.2.3.4 The COMP Function	50
3.2.2.3.5 The XCOMP Function	50
3.2.2.3.6 The OBL Function	51
3.2.2.3.7 The ADJ Function	51
3.2.2.3.8 The PRED Function	51
3.3 An Independent I-Structure for Discourse Information	53
3.3.1 Limitations of the Functional Account and Some Solutions	53
3.3.2 Characterizing Topicalization in Cantonese	54
3.4 Summary	63

Abbreviations

1	First person	N	Noun
2	Second person	NEW/New	Newness
3	Third person	NP	Noun phrase
A-structure	Argument structure	NUM	Number
ADJ	Adjunct	O/OBJ	Primary object
AdvP	Adverb phrase	OBJ _θ	Secondary object
AP	Adjective phrase	OBL	Oblique
C	Category/Complementizer/ Constraint	OT OT-LFG	Optimality Theory Optimality-Theoretic Lexical-Functional Grammar
C-structure	Constituent structure		
CL/CL	Classifier	PART	Particle
CLP	Classifier phrase	PERF	Perfective aspect
COMP	Closed Complement	PERS	Person
CON	A set of universal constraints	PP	Prepositional phrase
CP	Complementizer phrase	PRED	Predicate
D	Determiner	Pro	Pronoun
DEF	Definiteness	PROG	Progressive aspect
DEF	Definite Determiner	PROM/Prom	Prominence
DF	Discourse function	Q	Quantifier
DOC	Double-object construction	QP	Quantifier phrase
DP	Determiner phrase	QUAN	Quantity
E-TOP	External topic	RD	Right dislocation
EVAL	Evaluator	S/SUBJ	Subject
EXP	Experiential aspect	SG	Singular
F-structure	Functional structure	TG	Transformational Grammar
GEN	Generator		
GF	Grammatical function	V	Verb
I	Inflection	VP	Verb phrase
I-structure	Information structure	XCOMP	Open Complement
IP	Inflectional phrase		
LFG	Lexical-Functional Grammar		
LMT	Lexical Mapping Theory		

Chapter 1 Introduction

Topicalization is a common syntactic phenomenon in Cantonese. TOPIC and FOCUS are the two functions which a topicalized phrase may bear. This study investigates the two functions in Cantonese within the frameworks of Lexical-Functional Grammar (LFG) and Optimality-Theoretic Lexical-Functional Grammar (OT-LFG), establishing functional uncertainty equations and constraint rankings.

This chapter serves as an introduction to the study. Section 1.1 presents a literature review and the objectives of the study. Section 1.2 states some preliminary facts about the Cantonese language. The basic assumptions of the LFG and OT-LFG frameworks are briefly described in section 1.3. Section 1.4 defines the scope of the study. Section 1.5 addresses the research questions. The outline of the thesis is given in section 1.6.

1.1 Literature Review and Objectives of Study

Topic structures in Chinese, particularly Mandarin, have been widely researched (e.g. Li 2001, Li and Thompson 1981, Lu 2000, Pan and Hu 2003, Xu 1994, Xu and Liu 1998, Zhang 1996). Most of the studies are descriptive (e.g. Li and Thompson 1981, Zhang 1996) or have been done within a derivational approach (e.g. Li 2001, Pan and Hu 2003, Xu 1994). The same syntactic structure in Cantonese has however received comparatively less attention. FOCUS topicalization in Cantonese is also an area which requires more attention. Cheung (1997) and Law (2003) have studied the phenomenon of Gap Right Dislocation (Gap RD) in Cantonese which has the same

sentence structure with certain types of FOCUS topicalization. Their studies are again done within a derivational approach. The weaknesses of the movement approach have been outlined (e.g. Lu 2000, Xu and Liu 1998). LFG is a framework which is nonderivational in nature. Research on topicalization using an LFG approach has been done in a variety of languages including Mandarin (e.g. Her 1991, Huang 1992), Chicheŵa (e.g. Bresnan and Mchombo 1987), German (e.g. Ackerman and Webelhuth 1999), Russian (e.g. King 1995) and Vietnamese (e.g. Rosén 1998). Similar research on Cantonese has not yet been documented.

OT-LFG is a fairly new theoretical framework in formal syntax. A few studies on topicalization have been done within this framework. Choi's (1997, 1999, 2001) OT-LFG account of the phenomenon in Catalan, German and Korean is among the most comprehensive. Yu's (2005) OT analysis of topicalization of NP/DP after 'de', which is one of the few OT studies on topicalization in Chinese, is done within a transformational approach. The number of OT-LFG analyses in Cantonese is relatively limited and an OT-LFG account of topicalization in the language is yet to be done.

Given that topicalization in Cantonese has not yet been studied within the LFG and OT-LFG frameworks, the present study aims at providing an LFG and OT-LFG account of Cantonese topicalization through developing the functional uncertainty equations and the constraint rankings for the phenomenon with a view to demonstrating how the theoretical frameworks can be applied to the language.

1.2 The Cantonese Language

Cantonese is a Chinese language belonging to the Sino-Tibetan language family, spoken by approximately 71 million people. It is an official language of the Hong Kong Special Administrative Region and the Macau Special Administrative Region.

Cantonese has the canonical word order of SVO. It is a topic-prominent language (Li and Thompson 1976) and the word order can be altered depending on the position of the discourse topic in a clause. For instance, the OSV order is possible if the discourse topic is designated by the OBJ function. Topic constructions are commonly found in the language and they constitute one of the core components of the present study.

1.3 The Frameworks: LFG and OT-LFG

This section introduces the two frameworks, LFG and OT-LFG. In the following, the main features of Lexical-Functional Grammar (LFG) will be outlined, followed by a description of Optimality Theory (OT). It will be shown how OT and LFG work together in Optimality-Theoretic Lexical-Functional Grammar (OT-LFG) as one framework.

1.3.1 Lexical-Functional Grammar (LFG)

Lexical-Functional Grammar (LFG), developed by Bresnan and Kaplan in the 1970s, is a constraint-based grammar with parallel correspondence architecture (Bresnan 2001; Falk 2001). One of the main differences between LFG and Transformational Grammar (TG) is that no movement is assumed in LFG. LFG is constraint-based in the sense that grammaticality is determined by the satisfaction of constraints (Falk 2001). The three major parallel structures assumed in this theory include c(onstituent)-structure, f(unctional)-structure and a(rgument)-structure. Some other structures such as i(nformation)-structure and semantic structure have also been introduced. These structures are parallel in that none of them is regarded as more superior or more central than the other, which is contrary to the TG assumption that most of the syntactic concepts are defined in the constituent structure. A connection

is established between the structures by mapping principles. A brief description of c-structure, f-structure, a-structure (Bresnan 2001; Falk 2001) and i-structure (Choi 1999; King 1997) is provided below.

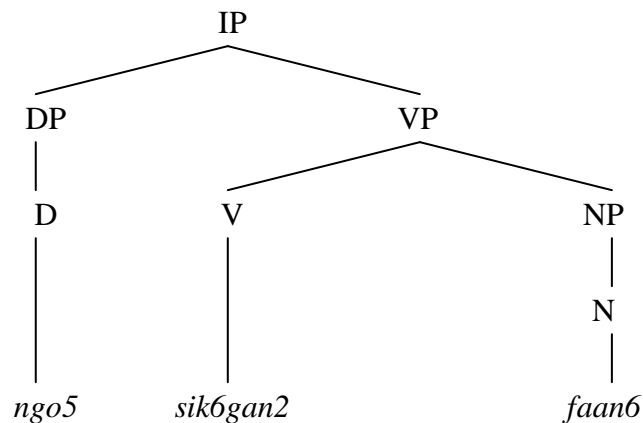
1.3.1.1 C-structure

C(onstituent)-structure is the representation of the overt structure of an expression in the form of a tree diagram. It shows the combination of words into a phrase and phrases into a clause. It can have an endocentric organization, which involves a hierarchical structure, or a lexocentric organization, which involves a flat structure. The X-bar notation (see Bresnan (2001) for a brief description) is employed for endocentrically organized c-structures. The c-structure of sentence (1) is shown in (2) as an example.

- | | | | |
|-----|----------------------|-----------------|----------------|
| (1) | 我 | 食緊 | 飯 ¹ |
| | <i>ngo5</i> | <i>sik6gan2</i> | <i>faan6</i> |
| | 1.SG | eat.PROG | meal |
| | 'I'm having a meal.' | | |

¹ All well-formed Cantonese examples consist of four lines. The first line shows the original Cantonese characters. The second line shows the romanization of the characters, employing the Jyutping system (Linguistic Society of Hong Kong 2003). The third line provides the word-for-word translation. The fourth line is the free English translation of the construction. No free translation is given for ill-formed constructions.

(2) C-structure of (1)



Empty categories in the c-structure are largely avoided in LFG in accordance with the principle of Economy of Expression (Bresnan 2001: 91):

- (3) Economy of Expression: All syntactic phrase structure nodes are optional and are not used unless required by independent principles (completeness, coherence, semantic expressivity).

With this principle, c-structure is more faithful to the actual structure of an expression.

1.3.1.2 F-structure

Grammatical functions in LFG are elements independent of the c-structure configuration, contrasting with TG's assumption that they are derived from the constituent structure. F(unctional)-structure is the level of expression which accommodates grammatical functions. It is a matrix consisting of attribute-value pairs, where the attribute, appearing on the left-hand side of the matrix, can be a symbol denoting a function (e.g. SUBJ and OBJ) or a symbol denoting a feature (e.g.

PRED, TENSE and NUM), and the value, appearing on the right-hand side of the matrix, can be represented by a symbol (e.g. PROG, SG and +), a semantic form (e.g. *book* and *tree*) or a subsidiary f-structure. The following shows the f-structure of sentence (1).

(4) F-structure of (1)

$$\left[\begin{array}{ll} \text{PRED} & \text{'sik6 } \langle (\uparrow\text{SUBJ}) (\uparrow\text{OBJ}) \rangle \text{' } \\ \text{ASPECT} & \text{PROG} \\ \text{SUBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'Pro'} \\ \text{PERS} & \text{1} \\ \text{NUM} & \text{SG} \end{array} \right] \\ \text{OBJ} & \left[\begin{array}{ll} \text{PRED} & \text{'faan6'} \end{array} \right] \end{array} \right]$$

'PRED' indicates meaningfulness. The ' $\langle (\uparrow\text{SUBJ}) (\uparrow\text{OBJ}) \rangle$ ' beside '*sik6*' is the subcategorization frame of the PRED, showing that the PRED subcategorizes for a SUBJ and an OBJ. 'PROG' for the ASPECT feature indicates progressive aspect. The value of the SUBJ and OBJ functions is a subsidiary f-structure. 'Pro' is the PRED value for pronouns. '1' for the PERS feature refers to first person and 'SG' for the NUM feature refers to singular number (see page vi for a list of abbreviations used in this thesis).

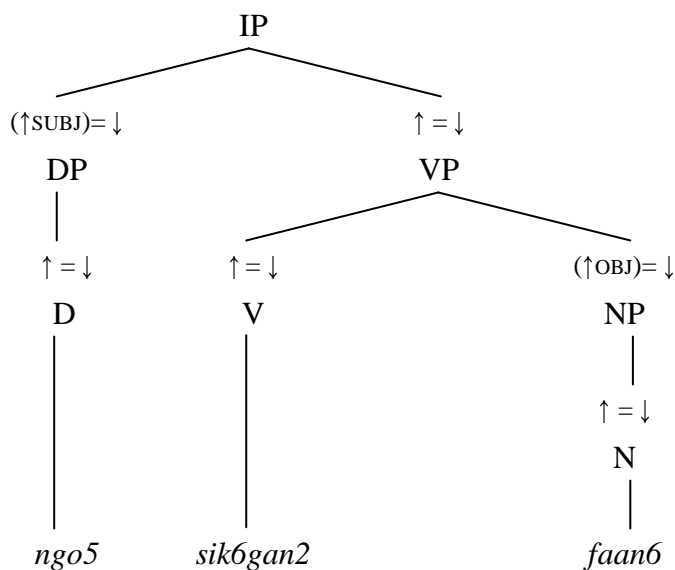
F-structures are subject to the following three well-formedness conditions (Bresnan 2001: 47, 63):

- (5) Uniqueness Condition: Every attribute has a unique value.
- (6) Completeness Condition: Every function designated by a PRED [should] be present in the f-structure of that PRED ... [I]f a designator ($\uparrow\text{GF}$) is associated with a semantic role by the PRED, the f-structure element satisfying the designator must itself contain a semantic feature [PRED *v*].

- (7) Extended Coherence Condition: All syntactic functions ... [should] be integrated appropriately into the f-structure ... Argument functions [i.e., SUBJ, OBJ, OBJ_θ, OBL and COMP] are integrated into the f-structure when they are designated by a PRED ... Nonargument functions [i.e., ADJ, TOPIC and FOCUS] are integrated if they bear an appropriate relation to the PRED. An ADJ is integrated if its immediate f-structure contains a PRED. A TOP[IC] or FOC[US] function is integrated whenever it is identified with, or anaphorically linked to, an integrated function.

A link can be established between c-structure and f-structure through the structure-function mapping principles resulting in an annotated c-structure. An annotated c-structure is a c-structure with functional equations attached to each node but the top node. The annotated c-structure of example (1) is shown below:

- (8) Annotated C-structure of (1)



The up-arrow '↑' refers to the mother of the present node and the down-arrow '↓' the

present node. The equation ‘ $\uparrow = \downarrow$ ’ means the f-structure of the mother node is the same as that of the present node. Therefore, the IP, VP and V share the same f-structure. The equation ‘ $(\uparrow\text{SUBJ}) = \downarrow$ ’ above the DP specifies that the SUBJ of the IP is the DP. By the same token, ‘ $(\uparrow\text{OBJ}) = \downarrow$ ’ above the NP indicates that the OBJ of the VP, and in turn of the IP, is the NP.

1.3.1.3 A-structure

A-structure is the representation of the participants in the event expressed by the PRED. It consists of a PRED and its arguments ordered according to the thematic hierarchy (Bresnan and Kanerva 1989 and Kiparsky 1987 cited in Bresnan 2001: 307):

- (9) Thematic Hierarchy: agent > beneficiary > experiencer/goal > instrument > patient/theme > locative

The a-structure of sentence (1) is given below:

- (10) A-structure of (1)
sik6 <Agent, Patient>

This structure shows that there are two participants in the event of *sik6*, which include an agent and a patient. Agent precedes patient according to the thematic hierarchy. This thematic hierarchy has influence on the mapping between a-structure and f-structure governed by the Lexical Mapping Theory (LMT) (see Bresnan 2001 and Falk 2001 for details).

1.3.1.4 I-structure

King (1997) proposes an independent i(nformation)-structure for accommodating discourse information. Diagram (11) below shows one of the possible i-structures of (1):

(11) I-structure of (1)



This structure specifies that *ngo5* is the discourse topic whereas *sik6gan2* and *faan6* are [+New] information. The [+New] feature was first introduced by Choi (1999), as opposed to [-New], to refer to ‘new’ information (see chapter 2 for a detailed description of information newness). Diagram (11) is only one of the possible i-structures of construction (1). In order to decide the information distribution of a construction, the context has to be taken into consideration. The i-structure in (11) is resulted from, for instance, the question *nei5 zou6gan2 me1 aa3?* (*what are you doing?*). With a different question, for example, *bin1 go3 sik6gan2 faan6 aa3?* (*who is having a meal?*), the i-structure will be different (with *ngo5* being the [+New] information). The motivation for an independent i-structure will be discussed in detail in chapter 3.

1.3.1.5 Summary

Four parallel structures in LFG, c-structure, f-structure, a-structure and i-structure, have been described above, with examples illustrating how these structures look like and what kind of information they represent. The next section

gives a brief outline of the OT theory.

1.3.2 Optimality Theory (OT)

Optimality Theory (OT) was introduced by Prince and Smolensky in 1993. This theory was originally applied to the area of phonology, and later extended to syntax. OT grammar is an input-output device (Kager 1999). Given an input, an infinite set of candidates are generated by the generator (GEN), which are then evaluated by the evaluator (EVAL), resulting in an optimal output. In the following, the nature of the input, GEN and EVAL will be discussed.

1.3.2.1 The Input

For a syntactic analysis, the input traditionally involves a lexical head with its argument structure, assignment of lexical heads to the arguments, and the tense (Grimshaw 1991 cited in Kager 1999). The input within the framework of OT-LFG has a different structure (to be described in section 1.3.3).

1.3.2.2 The Generator (GEN)

Given an input, the generator (GEN) generates output candidates. An infinite number of candidates can be generated provided that they meet the ‘universal vocabularies of linguistic representation’ (Kager 1999: 20). In the area of syntax, for instance, the candidates should conform to the X-bar structure. The form of the candidates within the OT-LFG framework will be discussed in section 1.3.3.

1.3.2.3 The Evaluator (EVAL)

The evaluator (EVAL) is responsible for selecting the optimal output from the set of candidates. It consists of a constraint ranking, marking of violations and

optimality evaluation.

A constraint ranking or hierarchy is a set of ranked constraints. The set of constraints (CON) is assumed to be universal, while the ranking of the constraints differ from language to language. It is this language-specific constraint ranking which leads to the language-specific grammar.

One of the very unique features of OT grammar is that constraints are violable. Violation of constraints does not directly lead to ungrammaticality. Constraint violation must however be minimal in that a constraint can be violated only to avoid the violation of the higher-ranked constraints.

Domination is strict in the constraint hierarchy, which implies that ‘violation of higher-ranked constraints cannot be compensated for by satisfaction of lower-ranked constraints’ (Kager 1999: 22). In other words, a candidate is ruled out if it violates a higher-ranked constraint, even if it satisfies the lower-ranked constraints better than other candidates. The optimal output is the output which incurs the least serious violations to the constraints with the ranking being taken into consideration. The evaluation is represented by a tableau. The diagram below shows a sample tableau:

(12) Sample Tableau

	C ₁	C ₂	C ₃	C ₄
a. <i>Candidate a</i>		*	**	**
b. <i>Candidate b</i>	*!		*	*

Candidate a and *candidate b* are the output candidates generated by the GEN. C₁, C₂, C₃ and C₄ are the constraints. The solid line between C₁, C₂ and C₃ means that C₁ dominates C₂ and C₃, and C₂ dominates C₃. The dotted line between C₃ and C₄ indicates that the two constraints are on the same level of the hierarchy (i.e., violating

C_3 is as serious as violating C_4). One asterisk ‘*’ signifies one violation. The exclamation mark ‘!’ indicates that the violation is fatal and the candidate is ruled out. The boxes after the fatal violation are grayed suggesting that they no longer have to be considered. The optimal output is denoted by the symbol ‘☞’. *Candidate a* is the optimal output since the highest-ranked constraint which it violates (C_2) is ranked lower than the one violated by *candidate b* (C_1). Incurring fewer violations to those lower-ranked constraints (C_2 , C_3 and C_4) does not save *candidate b* from being ruled out.

1.3.2.4 Summary

OT is an input-output device involving a GEN and an EVAL. Inputs are processed by the GEN to generate candidates. The candidates are then evaluated by the EVAL, represented by a tableau, according to the constraint hierarchy and the violations of constraints so that an optimal output is selected. The next section explains how OT and LFG work together in the framework of OT-LFG.

1.3.3 OT-LFG

OT-LFG does not only imply that assumptions from OT and LFG are adopted. In fact, as mentioned above, the forms of inputs and outputs within the OT-LFG framework are different. The input, instead of a lexical head with its arguments, is an underspecified f-structure (Bresnan 2000). Based on the i-structure proposed by King (1997), the present study includes an extra i-structure in the input. The underspecified f-structure of sentence (1) is shown below (see diagram (11) for the i-structure).

(13) Underspecified f-structure of (1)

$$\left[\begin{array}{l}
 \text{PRED} \quad \text{'sik6} \langle x, y \rangle \\
 \text{ASPECT} \quad \text{PROG} \\
 \text{GF}_1 \quad \left[\begin{array}{l}
 \text{PRED} \quad \text{'Pro'} \\
 \text{PERS} \quad 1 \\
 \text{NUM} \quad \text{SG}
 \end{array} \right] x \\
 \\
 \text{GF}_2 \quad \left[\text{PRED} \quad \text{'faan6'} \right] y
 \end{array} \right]$$

In an underspecified f-structure, grammatical functions are represented by GF_i , with different subscript digits indicating different functions. The functions subcategorized for by the PRED are expressed by the variables x and y (z if there exists a third subcategorized function) and the same variables appear beside the subsidiary f-structures of GFs, indicating a correspondence.

The output candidates generated within the OT-LFG framework consist of a fully-specified f-structure and an annotated c-structure which shows the mapping between the c- and f-structure (see diagrams (4) and (8)).

1.4 Scope of Research

In this thesis, the phenomenon of topicalization in Cantonese is studied, investigating the two functions which a topicalized phrase may bear, TOPIC and FOCUS. This study examines the acceptable and unacceptable topicalization constructions with the aim of establishing the functional uncertainty equations for the syntactic phenomenon and specifically explores OBJ topicalization constructions in an attempt to develop the constraint rankings for the language.

1.5 Research Questions

The following research questions will be answered in this thesis:

- (i) How should TOPIC and FOCUS in Cantonese be defined?
- (ii) How should TOPIC and FOCUS topicalization in Cantonese be characterized with Functional Uncertainty in LFG? Does the functional approach suffice?
- (iii) What are the OT constraints relevant to Cantonese OBJ topicalization and how should they be ranked against each other?

The definition of Chinese-style topic (Chafe 1976) will be adopted and the information features of FOCUS in Cantonese will be explored in order to answer question (i). Question (ii) will be addressed by establishing the functional uncertainty equations for the phenomenon and investigating the f-structure and i-structure of some topicalization constructions. Question (iii) requires an examination of the reasons for the grammaticality and ungrammaticality of OBJ topicalization constructions.

1.6 Outline of Thesis

This thesis is organized as follows. After introducing the study in this chapter, chapter 2 investigates the TOPIC and FOCUS functions in Cantonese, defining the two notions. Chapter 3 provides an LFG perspective on Cantonese topicalization, looking specifically at the functional uncertainty equations and the c-structure, f-structure and i-structure of the constructions. Chapter 4 presents an OT-LFG account, proposing the constraint rankings for Cantonese based on the observations on topicalization. The study is summarized and concluded in chapter 5.

Chapter 2 TOPIC and FOCUS in Cantonese

TOPIC and FOCUS are the two functions which a topicalized phrase may bear. This chapter explains what topicalization refers to and defines TOPIC and FOCUS in Cantonese.

This chapter has the following structure. Section 2.1 defines topicalization. Section 2.2 differentiates between discourse topic/focus and syntactic topic/focus. In section 2.3, TOPIC in Cantonese is defined and another type of topic, E-TOP, is identified. In section 2.4, the information features of Cantonese FOCUS are examined, with a view to establishing the definition of FOCUS. Section 2.5 summarizes the chapter.

2.1 Topicalization

Topicalization is a syntactic phenomenon to which a variety of definitions have been given (e.g. Gregory and Michaelis 2001, Krapova 2004, Lambrecht 1994, LaPolla 1988, Matthews and Yip 1994, Rosén 1998). This study employs Rosén's (1998) definition of topicalization as the basis of the analysis, which conforms to the assumptions in LFG.

According to Rosén (1998: 184), '[t]opicalization is a construction in which a leftmost constituent is understood as filling a missing constituent in the sentence'. Sentence (1) is an example of topicalization in Cantonese according to this definition.

- (1) 呢 本 書 我 鍾意
nei1 bun2 syu1 ngo5 zung1ji3
 DEFCL book 1.SG like
 ‘As for this book, I like it.’ or
 ‘It is this book that I like.’

The leftmost constituent *nei1 bun2 syu1* fills the missing OBJ function of the verb *zung1ji3*. A topicalized phrase can bear either the TOPIC or the FOCUS function. In Cantonese, a topicalized phrase followed by no particle can be interpreted as either of the functions depending on the context and the intonational pattern of the utterance. Therefore, with *nei1 bun2 syu1* followed by no particle, the construction has two possible interpretations. Particles in Cantonese specify different speech-acts, evidentiality and emotions of the speaker (Matthews and Yip 1994). Some particles tend to follow a TOPIC and some a FOCUS. In this thesis, *aa4* and *ne1* will be used for the TOPIC function and *aa3*, *laa1* and *gaa3* for the FOCUS function².

Though a topicalized phrase normally occurs in the leftmost position of a sentence, there are instances where it does not, for example, when there are multiple instances of topicalization in one sentence. Consider (2).

- (2) 佢 啊 本 書 我 俾咗
keoi5 aa3 bun2 syu1 ngo5 bei2zo2
 3.SG PART CL book 1.SG give.PERF
 ‘It is him/her who I have given the book to.’

In this construction, the OBJ_θ *keoi5* and the OBJ *bun2 syu1* are topicalized to become the FOCUS (note the particle *aa3* for the FOCUS function) and the TOPIC (a TOPIC is usually followed by no particle with a FOCUS preceding it) respectively. The TOPIC is

² A pilot questionnaire survey was done in an attempt to categorize Cantonese particles into TOPIC and FOCUS particles. A conclusion, however, can hardly be drawn as there are too many exceptions. Further research has to be done in this area in order to classify the particles in Cantonese.

preceded by the FOCUS so that it is no longer sentence-initial. *Bun2 syu1* is still regarded as a topicalized phrase even though it does not occur in the left periphery since it is extracted from its canonical post-verbal OBJ position to the pre-SUBJ position.

2.2 Discourse Topic/Focus vs. Syntactic Topic/Focus

‘Topic’ and ‘focus’ are sometimes confusing terms in the sense that they are used to refer to both the discourse notion and the syntactic notion. To resolve this kind of confusion, different terms have been invented for the discourse and the syntactic concept. For instance, Her (1989) uses ‘topic’ to denote the syntactic notion and ‘frame’ the discourse notion.

According to Bresnan and Mchombo (1987: 746), ‘grammaticized topics — constituents that bear the TOP[IC] function — designate discourse topics; but not all discourse topics are grammatically marked’. In other words, topicalized topics, TOPICS, designate discourse topics while discourse topics are not necessarily topicalized. The same idea is also assumed for focus, i.e., topicalized focuses, FOCUSES, designate discourse focuses while discourse focuses are not necessarily topicalized. Following their suggestions and the general practice in LFG which represents grammatical functions in small capitals, ‘TOPIC’ and ‘FOCUS’³ are used in the present study to refer to topicalized discourse topics and topicalized discourse focuses respectively. Apart from TOPIC, another type of syntactic topic, external topic (Aissen 1992), will be introduced in the following discussion and will be represented as ‘E-TOP’ (King 1995). ‘Topic’ will be used to generally denote syntactic topics, including both TOPIC and E-TOP. Discourse topic and discourse focus will be stated

³ It should be noted that there are elements other than topicalized phrases which can bear the TOPIC and FOCUS functions. For example, the relative pronoun of a relative clause (TOPIC) and the question phrase of a question (FOCUS).

explicitly when being referred to.

2.3 Topic in Cantonese

Topic in Cantonese will be examined in this section. In section 2.3.1, Cantonese TOPIC is defined. Another type of topic, E-TOP, will be introduced in section 2.3.2.

2.3.1 Definition of TOPIC in Cantonese

This study adopts Chafe's (1976) definition of Chinese-style topic to characterize the information features of TOPIC in Cantonese. Cantonese TOPIC is defined as a sentence-initial constituent filling a missing constituent in the sentence which 'limit[s] the applicability of the main predication to a certain restricted domain' and 'sets a spatial, temporal, or individual framework within which the main predication holds' (Chafe 1976: 50). The following example illustrates how this definition applies:

- (3) 呢 本 書 啊 我 睇過 好多 次
nei1 bun2 syu1 aa4 ngo5 tai2gwo3 hou2do1 ci3
 DEFCL book PART 1.SG read.EXP many times
 'As for this book, I have read it for many times.'

The sentence-initial DP *nei1 bun2 syu1* is identified with the missing OBJ function of the sentence. This DP limits the applicability of the main predication *ngo5 tai2gwo3 hou2do1 ci3* (*I have read (it) for many times*), by designating what exactly the speaker has read for many times. Therefore, according to the above definition, *nei1 bun2 syu1* is a TOPIC.

2.3.2 TOPIC vs. E-TOP

While the sentence-initial constituent in example (3) above fills a missing function in the sentence, there are constructions of which the leftmost constituent conforms to Chafe's (1976) definition of Chinese-style topic without being identified with any missing constituent in the sentence. Based on Aissen's (1992) analysis of three Mayan languages, it is proposed that there are two types of topics in Cantonese, internal topic (TOPIC) and external topic (E-TOP). Consider the following example:

- (4) 呢 件 事 呢 佢 的確 衝動咗 啲
nei1 gin6 si6 ne1 keoi5 dik1kok3 cung1dung6zo2 di1
 DEFCL matter PART 3.SG indeed impetuous.PERF a bit
 'As for this matter, it is indeed quite impetuous of him/her (to act like that).'

The PRED *cung1dung6* subcategorizes only for a SUBJ and the construction without the sentence-initial DP *nei1 gin6 si6* is already well-formed, with the DP *keoi5* being the SUBJ. *Nei1 gin6 si6* cannot be identified with any missing function in the sentence and the construction is therefore not an instance of topicalization.

Aissen (1992) identifies two types of topic in the Mayan languages she investigates, which are internal topic⁴ and external topic. An internal topic is in fact a TOPIC. It has a closer syntactic connection with the following clause than an external topic (Aissen 1992). An external topic, or an E-TOP according to King (1995), is 'less integrated into basic clause structure, being essentially prefixed to what is otherwise a fully well-formed clause' (Aissen 1992: 44). It is similar to what Krapova (2004) calls a hanging topic in that it has a rather loose relation with the following clause. Unlike TOPIC, an E-TOP is not required to bind any argument,

⁴ The internal topic here should be differentiated from the internal topic in Paul (2002, 2005) which is defined as a preposed object appearing between the SUBJ and the verb.

though it can be coreferent with a pronoun. There can be a pragmatic connection but no syntactic connection between the E-TOP and the following clause. It should be noted that, since Cantonese is a pro-drop language (Lee 2003), a ‘fully well-formed clause’ does not necessarily demand that all arguments be present in the clause. A clause is still well-formed even if some arguments are missing provided that they can be recovered from the context. Consider (4) again. The sentence-initial DP *nei1 gin6 si6* does not bind any argument and has no syntactic relation with the following already well-formed clause. Its relation to the clause is only pragmatic in that it restricts the applicability of the predication *keoi5 dik1kok3 cung1dung6zo2 di1* (*it is indeed quite impetuous of him/her (to act like that)*) to the scope of *nei1 gin6 si6* (*this matter*). In other words, it is for this particular matter that the addressee is regarded as being impetuous and this utterance is not commenting on the general personality of the addressee. Given these observations, it is concluded that the sentence-initial DP *nei1 gin6 si6* is an E-TOP.

To summarize, there are two types of topic in Cantonese, TOPIC and E-TOP. They differ with respect to their syntactic relation with the following clause in that a TOPIC is identified with a missing function while an E-TOP is not.

2.4 FOCUS in Cantonese

Languages use different means to encode discourse focus. Green and Jaggar (2003) suggest the following morphosyntactic and prosodic means and observe that languages often make use of more than one of the following options:

- (5) a. Focus *in situ*
 b. Focus movement
 - Clause-initial

- Pre-verbal
- Post-verbal
- c. Focus markers
- d. Focal stress

This section investigates when and how Cantonese makes use of topicalization to encode discourse focus (clause-initial focus movement in Green and Jaggar's sense; the present study does not analyze this phenomenon as a kind of movement as will be demonstrated in chapter 3) in an attempt to give a definition to FOCUS in Cantonese. The methodology is presented in section 2.4.1. The properties of FOCUS topicalization are discussed and the definition of FOCUS in Cantonese is established in section 2.4.2.

2.4.1 Methodology

In order to find out the properties of FOCUS and the contexts under which a phrase is topicalized to become a FOCUS in Cantonese, question-answer pairs were constructed, where the questions determine which element(s) in the answers designate(s) the discourse focus and the answers are topicalization constructions with different elements being the FOCUS. A survey was conducted in order to confirm the legitimacy of the answers with respect to the questions. A questionnaire consisting of three parts (see appendix I) was distributed to 40 Cantonese native speakers. In part III (parts I and II to be discussed in chapter 3), the informants were given four questions followed by two to six answers. Each question and answer was pronounced aloud by the interviewer (the author) and the informants were asked to indicate the acceptability ('acceptable', 'strange, but still acceptable', 'very strange' and 'unacceptable') of the answers with respect to the given questions. The

questionnaires were then processed by assigning marks to each option of acceptability and calculating the mean of the marks for each construction (Lam 2004). Marks ranging from -2 to +2, instead of 1 to 4, are assigned to the options so that the acceptability of the constructions can be decided by the plus or minus sign of the mean. +2 marks and +1 mark are assigned to each ‘acceptable’ and ‘strange but still acceptable’ respectively. -1 mark and -2 marks are assigned to each ‘very strange’ and ‘unacceptable’ respectively. Constructions having a positive mean are considered as acceptable in the following discussion and those with a negative mean are considered as unacceptable (see appendix II for the calculated mean of each construction). The results of this part of the questionnaire survey are also used in the OT-LFG account in chapter 4.

2.4.2 Information Features of FOCUS: Towards a Definition⁵

Choi (1997, 1999, 2001) uses the features of [New] and [Prom] to characterize discourse topic and discourse focus. [New] refers to newness and [Prom] refers to prominence⁶. The following table shows how the interaction of these two features contributes to defining different types of information (Choi 1999: 92):

(6)

	+Prom	-Prom
-New	Topic	Tail
+New	Contrastive Focus	Completive Focus

In the present study, only discourse focus is characterized by the [New] and [Prom] features (see section 2.3 for the defining properties of a topic in Cantonese).

⁵ Part of this section has been presented by the author at the Annual Research Forum of the Linguistic Society of Hong Kong (Fung 2006) and the Fifth Cambridge Postgraduate Conference in Language Research (Fung 2007).

⁶ Note that ‘prominence’ here refers to the prominence of information, to be differentiated from, for example, phonological prominence.

As shown in the table, there are two types of discourse focus including contrastive discourse focus which represents [+New] and [+Prom] information and completive discourse focus which conveys [+New] but [-Prom] information, where [+New] refers to new, informative or unknown information and [+Prom] refers to important, urgent, unexpected, contrastive or surprising information (Choi 2001). Choi (1999) shows that [+Prom] is the motivation for topicalization in English. In order to define FOCUS in Cantonese, it will be investigated in the following what licenses FOCUS topicalization.

The phenomenon of Gap Right Dislocation (Gap RD) in Cantonese, which refers to a kind of leftward movement in which a constituent is ‘moved’ from the end to the initial position of a sentence (Cheung 1997), has the same sentence structure with certain types of FOCUS topicalization. When the FOCUS is identified with a constituent of which the canonical position is at the end of the sentence, the topicalization construction has the same structure as Gap RD. As mentioned in chapter 1, the Gap RD phenomenon in Cheung (1997) is studied within a derivational approach, contrasting with the LFG approach taken in the present study (see chapter 3). Regarding information structure, Cheung proposes that the left-dislocated phrase is the domain of focus where ‘the function of focus is to indicate the presence of alternatives to the expression in focus that are under discussion’ (Krifka 1997 cited in Cheung 1997: 95). FOCUS is a topicalized phrase designating the discourse focus (Bresnan and Mchombo 1987). The present study however does not employ the concept of ‘alternatives’ to encode discourse focus since a TOPIC may also involve alternatives as illustrated in the following example:

- (7) 呢 本 書 呢 我 就 買 咗,
nei1 bun2 syu1 ne1 ngo5 zau6 maai5zo2,
 DEFCL book PART 1.SG zau buy.PERF
- 嗰 本 書 呢 我 就 未 買
go2 bun2 syu1 ne1 ngo5 zau6 mei6 maai5
 DEFCL book PART 1.SG zau not yet buy

‘As for this book, I have bought it. As for that book, I haven’t bought it yet.’

In this example, *nei1 bun2 syu1* and *go2 bun2 syu1*, followed by the TOPIC particle *ne1*, are TOPICS. At least two alternatives are involved here, *this book* and *that book*. Therefore, ‘alternatives’ is not an exclusive property of discourse focus. This study adopts Choi’s (1999) [+New] feature to characterize discourse focus, employing Lambrecht’s (1994) concept of ‘new’ and Halliday’s (1967) idea of ‘informative’ to interpret ‘[+New]’.

A phrase is regarded as carrying new or unknown information if ‘the denotatum of this phrase stands in a pragmatically construed relation to the proposition such that its addition makes the utterance of the sentence a piece of new information’ (Lambrecht 1994: 210). Whether the constituent has been previously mentioned does not matter. A piece of information is informative in the sense that the speaker intends it ‘to be interpreted as informative’ (Halliday 1967: 204). Consider the following example:

- (8) Q: 你 買 咗 咩 啊?
nei5 maai5zo2 me1 aa3?
 2.SG buy.PERF what PART
 ‘What have you bought?’

A: 三 本 字典 啊 我 買咗
saam1 bun2 zi6din2 aa3 ngo5 maai5zo2
 three CL dictionary PART 1.SG buy.PERF
 ‘It is three dictionaries that I have bought.’

Question-answer pairs can be used to identify the discourse focus of a sentence (Rochemont and Culicover 1990). By asking *nei5 maai5zo2 me1 aa3?* (*what have you bought?*), the addresser assumes that the addressee has bought something. *Saam1 bun2 zi6din2* in the answer then represents new information in the sense that its addition to the presupposition *I have bought something* makes the utterance a piece of new information. This piece of information is also presented by the speaker as the most informative part of the whole utterance. Therefore, this phrase bears [+New] information and is thus the discourse focus according to Choi’s (1999) classification. A topicalized phrase which designates the discourse focus is a FOCUS (Bresnan and Mchombo 1987).

While the grammatical function assigned to a particular element is fixed, the informational status assigned can differ from speaker to speaker (Choi 2001). For instance, while a speaker regards an element as [+Prom], another speaker under the same context can regard it as [-Prom]. In example (8), if buying three dictionaries is not regarded as surprising, the phrase *saam1 bun2 zi6din2* in the answer represents [+New] but [-Prom] information and thus designates a completive discourse focus. If buying three dictionaries is considered surprising, the phrase bears [+New] as well as [+Prom] features and therefore designates a contrastive discourse focus. Whether carrying [+Prom] or [-Prom] information, the phrase *saam1 bun2 zi6din2* can still be topicalized. It is concluded that the [Prom] feature does not affect the acceptability of FOCUS topicalization and it is the [+New] feature which licenses FOCUS topicalization in Cantonese.

In the above example, the FOCUS corresponds to the discourse focus, that is the [+New] information. A topicalized phrase can include at the same time [+New] and [-New] information as in the following example:

- (9) Q: 你 買咗 三 本 咩 啊?
nei5 maai5zo2 saam1 bun2 me1 aa3?
 2.SG buy.PERF three CL what PART
 ‘What are the three books that you have bought?’
- A: 三 本 字典 啊 我 買咗
saam1 bun2 zi6din2 aa3 ngo5 maai5zo2
 three CL dictionary PART 1.SG buy.PERF
 ‘It is three dictionaries that I have bought.’

Given the question, the presupposition here is that the addressee has bought three books. *Zi6din2* is therefore the [+New] information as its addition to the presupposition makes the utterance a piece of new information. The topicalized phrase contains both [+New] (*zi6din2*) and [-New] (*saam1 bun2*) information. This phrase should still be regarded as a FOCUS since it contains the [+New] information. Therefore, it is proposed that a FOCUS in Cantonese is a topicalized phrase which *contains* the discourse focus (i.e., [+New] information).

While the FOCUS in the above example contains the entire discourse focus, the FOCUS in the following example only contains part of the discourse focus:

- (10) Q: 你 買咗 咩 啊?
nei5 maai5zo2 me1 aa3?
 2.SG buy.PERF what PART
 ‘What have you bought?’

A: 字典 啊 我 買咗 三 本
zi6din2 aa3 ngo5 maai5zo2 saam1 bun2
 dictionary PART 1.SG buy.PERF three CL
 ‘I have bought three DICTIONARIES’

The question presupposes that the addressee has bought something. *Saam1 bun2 zi6din2* is what makes the utterance new and is thus the [+New] information. Only part of the [+New] information *zi6din2* is topicalized to be the FOCUS, with *saam1 bun2* in its canonical postverbal OBJ position. This example demonstrates that it is not obligatory for the entire phrase corresponding to the discourse focus to appear inside the FOCUS.

To conclude this section, [+New] is the feature which licenses FOCUS topicalization and FOCUS in Cantonese is defined as a topicalized phrase which contains the entire or part of the discourse focus ([+New] information), whether contrastive ([+Prom]) or completive ([-Prom]).

2.5 Summary

This chapter has outlined topicalization in Cantonese and defined TOPIC and FOCUS. Section 2.1 has defined topicalization. Section 2.2 has distinguished syntactic topic/focus from discourse topic/focus. In section 2.3, Cantonese TOPIC has been defined as a topicalized phrase which limits the applicability of the main predication and another type of topic, E-TOP, has been identified. Section 2.4 has concluded that [+New] is the feature which licenses FOCUS topicalization and defined FOCUS as a topicalized phrase which contains the entire or part of the discourse focus ([+New] information).

In the next chapter, the phenomenon of topicalization in Cantonese will be accounted for within an LFG approach.

Chapter 3 TOPIC and FOCUS: An LFG Perspective

The LFG framework offers a systematic account of the topicalization phenomenon. A phrase structural approach was first introduced for characterizing TOPIC and FOCUS (Kaplan and Bresnan 1982, Zaenan 1980, 1983 cited in Kaplan and Zaenan 1989). This approach was then rejected and a functional approach was suggested (Kaplan and Zaenan 1989). Later, an informational approach was proposed to represent discourse information (King 1997). This chapter investigates to what extent a functional approach can be applied to topicalization in Cantonese and how the informational approach complements the functional account to attain a satisfactory account.

This chapter is organized as follows. The c-structure configuration of TOPIC, FOCUS and E-TOP constructions is presented in section 3.1. Section 3.2 provides a functional account of topicalization and illustrates its application to Cantonese. In section 3.3, the motivation and essence of the informational approach and its application to Cantonese topicalization are discussed. Section 3.4 summarizes the chapter.

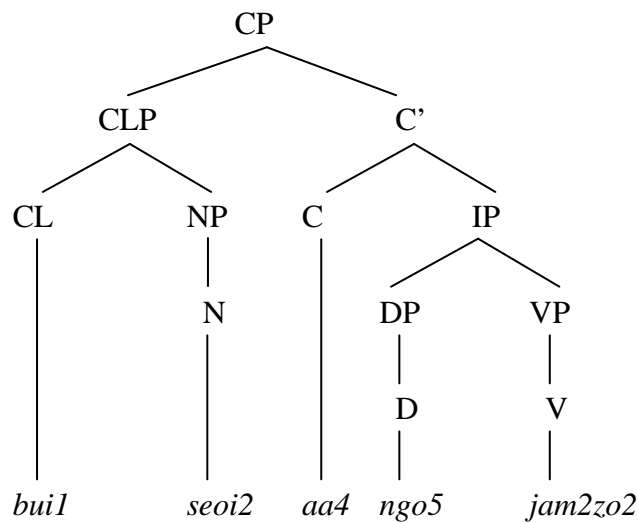
3.1 C-structure Configuration

In this section, it will be illustrated how the TOPIC, FOCUS and E-TOP constructions in Cantonese should be represented in the c-structure. Specifier positions of functional projections accommodate grammaticized discourse functions (Bresnan 2001). It is proposed that TOPIC and FOCUS in Cantonese occupy the specifier position of the functional projection CP with the TOPIC/FOCUS particle

occupying the C^0 position. Sentence (1) is an instance of TOPIC topicalization and the c-structure of this example is shown in (2).

- (1) 杯 水 啊 我 飲咗
bui1 seoi2 aa4 ngo5 jam2zo2
 CL water PART 1.SG drink.PERF
 ‘As for the cup of water, I have drunk it.’

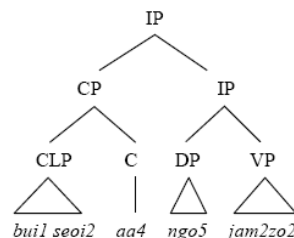
- (2) C-structure of (1)⁷



⁷ See Svolacchia, Mereu and Puglielli (1995) for a similar c-structure for FOCUS constructions in Somali and King (1995) for *li* yes-no questions in Russian.

Though there is a C' level which corresponds to *aa4 ngo5 jam2zo2*, it does not mean that this sequence of words is a constituent since C' is only an intermediate level of the maximal projection CP.

Alternative c-structures have been suggested for topicalization constructions, with adjunction structure (e.g. Bresnan 2001, King 1995, Rosén 1998) being one of those. An alternative c-structure of (1) is shown below, where the TOPIC CP is adjoined to IP:



In this c-structure, *bui1 seoi2 aa4* is considered as a constituent. As mentioned in chapter 2, some particles have a higher tendency to follow a TOPIC and some a FOCUS. The choice of the particle however depends on the speaker's attitude towards the event expressed by the entire construction rather than the TOPIC or FOCUS only. Therefore, the present analysis employs a c-structure where the particle C heads the CP corresponding to the entire construction. More research is necessary in order to confirm this analysis.

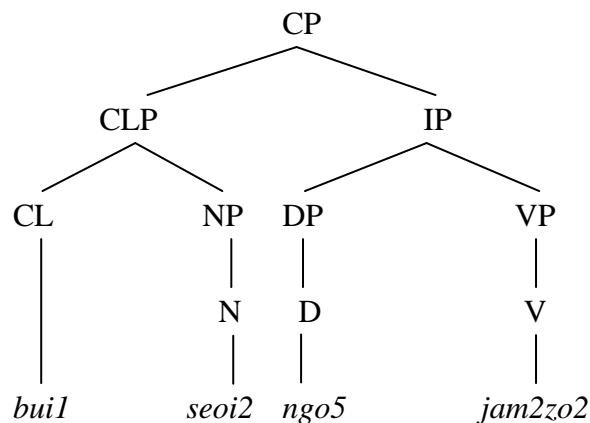
The [SPEC, CP] is occupied by the CLP *bui1 seoi2*, which is a TOPIC, and the C^0 is occupied by the TOPIC particle *aa4*, which selects an IP as its complement.

For constructions with no particle following the topicalized phrase, where the topicalized phrase can be interpreted as either a TOPIC or a FOCUS, the CP has no head.

The c-structure of sentence (3) is given in (4) as an example.

- (3) 杯 水 我 飲咗
 bui1 seoi2 ngo5 jam2zo2
 CL water 1.SG drink.PERF
 ‘As for the cup of water, I have drunk it.’ (TOPIC topicalization)
 ‘It is the cup of water that I have drunk.’ (FOCUS topicalization)

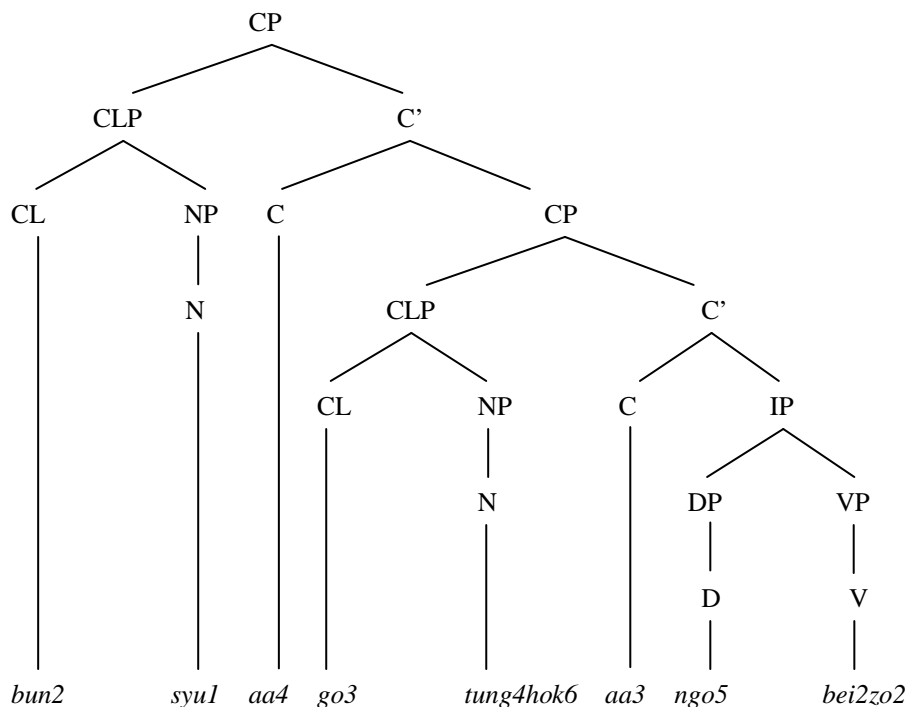
- (4) C-structure of (3)



There can be more than one topicalized phrase co-existing in a construction. In this case, CP is instantiated recursively (Svolacchia, Mereu and Puglielli 1995). Sentence (5) is a construction with both TOPIC and FOCUS, where the TOPIC *bun2 syu1* is identified with the OBJ and the FOCUS *go3 tung4hok6* is identified with the OBJ_θ (note the different particles). The c-structure of this construction is shown in (6).

- (5) 本 書 啊 個 同 學 啊 我 俾 咗
bun2 syu1 aa4 go3 tung4hok6 aa3 ngo5 bei2zo2
 CL book PART CL classmate PART 1.SG give.PERF
 ‘As for the book, I have given it to THE CLASSMATE.’

- (6) C-structure of (5)



Two CPs are postulated in order to create two [SPEC, CP] positions to accommodate the TOPIC and the FOCUS and two C⁰ positions for the particles.

It has been proposed in chapter 2 that there is a second type of topic in Cantonese, which is E-TOP. Example (4) in chapter 2 is repeated below as (7):

- (7) 呢 件 事 呢 佢 的 確 衝 動 咗 啲
nei1 gin6 si6 nei1 keoi5 dik1kok3 cung1dung6zo2 di1
 DEFCL matter PART 3.SG indeed impetuous.PERF a bit
 ‘As for this matter, it is indeed quite impetuous of him/her (to act like that).’

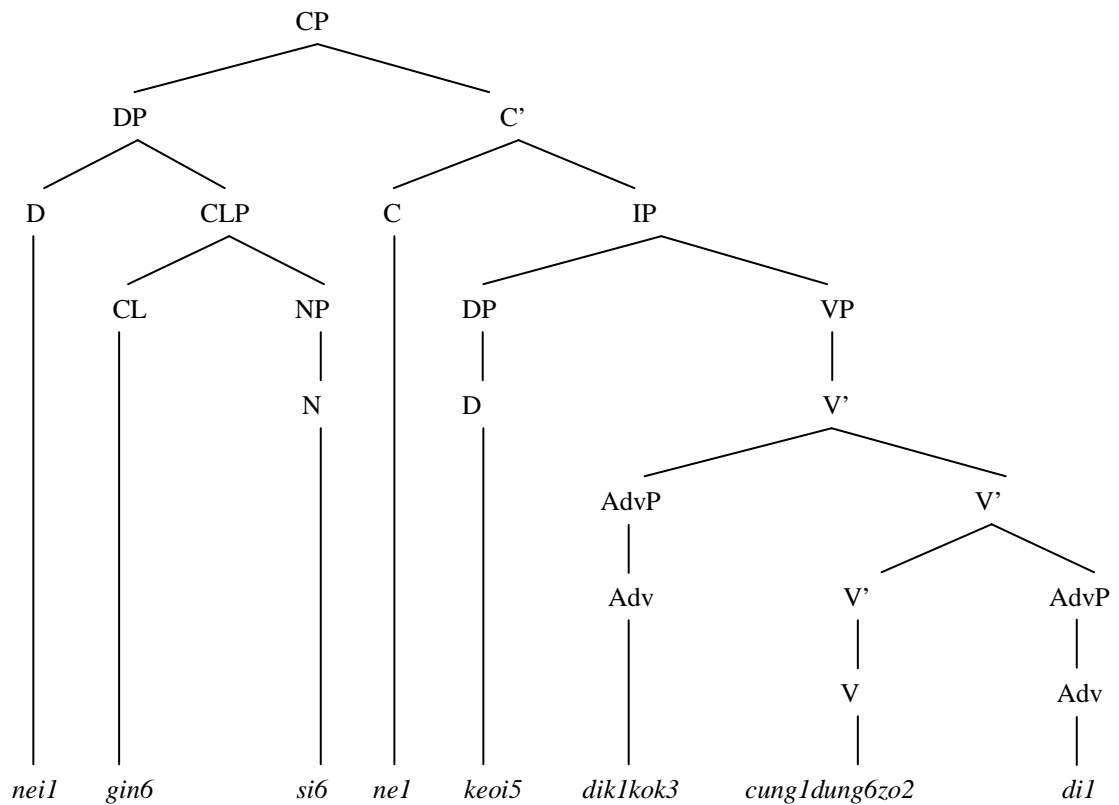
The ‘E(xpression)’ notation which only appears in the matrix clause is used as the top node for E-TOP constructions in Aissen (1992). An E-TOP in Cantonese can however occur inside a subordinate clause as shown in (8) below:

- (8) 我 覺得 [呢 件 事 呢 佢 的確
ngo5 gok3dak1 [neil gin6 si6 nel keoi5 dik1kok3
 1.SG think DEF CL matter PART 3.SG indeed
 衝動咗 啲]
cung1dung6zo2 di1]
 impetuous.PERF a bit

‘I think, as for this matter, it is indeed quite impetuous of him/her (to act like that).’

Based on this observation, the ‘E’ node is not employed in the present study. The c-structure of (7) is proposed in (9) below. Like the case of topicalization, the E-TOP appears in the specifier position of the CP.

(9) C-structure of (7)



In the c-structures of TOPIC, FOCUS and E-TOP constructions presented above, it is not indicated what types of discourse function the phrase in the [SPEC, CP] position bears. This information is shown in the f-structure and the functional annotations in an annotated c-structure, which will be illustrated in the functional analysis in the next section.

3.2 A Functional Approach to Topicalization

Topicalization is a kind of long distance dependencies. In early LFG, long distance dependencies were handled in the c-structure level (Kaplan and Bresnan 1982, Zaenan 1980, 1983 cited in Kaplan and Zaenan 1989). There are however extraction constraints which cannot be accounted for in purely phrase structural

terms. Kaplan and Zaenan (1989) illustrate with Icelandic data that it is more appropriate to characterize extraction constraints in functional terms and propose Functional Uncertainty as a formal device for handling long distance dependencies. In section 3.2.1 below, a brief outline of Functional Uncertainty is given. The functional uncertainty equations for Cantonese topicalization are derived in section 3.2.2.

3.2.1 Functional Uncertainty

Functional uncertainty is the formal device in LFG for characterizing long distance dependencies. As suggested by its name, it is a device defined in functional terms, rather than phrase structural terms. It ‘permits a functional statement of constraints on unbounded dependencies’ (Kaplan and Maxwell 1988: 297). The following provides a brief outline of how Functional Uncertainty is derived.

Kaplan and Bresnan (1982) first find out that the relation between a gap (i.e., the constituent which is missing) and a filler (i.e., the phrase which fills the missing constituent) can be captured by an equation. Sentence (10) is a topicalization construction. The equation in (11) indicates that the TOPIC *Mary* is at the same time the OBJ of the COMP (Kaplan and Zaenan 1989: 26):

(10) Mary John claimed that Bill telephoned yesterday.

(11) $(\uparrow\text{TOPIC}) = (\uparrow\text{COMP OBJ})$

Kaplan and Zaenan (1989) point out that sentence (10) can be extended by adding more COMPS as in their example sentence *Mary John claimed that Bill said that ... that Henry telephoned yesterday*, where the number of COMPS is uncertain. They introduce the following equation for this phenomenon (1989: 26):

$$(12) \quad (\uparrow\text{TOPIC}) = (\uparrow\text{COMP COMP} \dots \text{OBJ})$$

They immediately notice the problem of this equation being that ‘this is an infinite family of equations, and hence impossible to enumerate in a finite disjunction appearing on a particular rule of grammar’ (1989: 26). In order to represent this family of equations in a finite way, they first extend the function-application expression (13) to (14) to allow it to represent strings of symbols.

$$(13) \quad (f s) = v \text{ holds if and only if } f \text{ is an f-structure, } s \text{ is a symbol, and the pair } \langle s, v \rangle \in f.$$

$$(14) \quad (f x) \equiv ((f s) y), \text{ where } x = s y \text{ with } s \text{ being an initial symbol and } y \text{ being a possibly empty suffix string}$$

$$(f \epsilon) \equiv f, \text{ where } \epsilon \text{ is the empty string.}$$

The extended function-application expression in (14) allows only for individual strings and therefore is not yet able to characterize the uncertainty of long distance dependencies. Since the OBJ can be embedded in *any* number of COMPS, from zero to infinity, an equation has to be derived which includes strings from ($\uparrow\text{OBJ}$) to ($\uparrow\text{COMP COMP} \dots \text{OBJ}$) and any other possible strings. Kaplan and Zaenan (1989: 27) devise Functional Uncertainty which is able to denote a set of strings:

$$(15) \quad (f \alpha) = v \text{ holds if and only if}$$

$$((f s) \text{Suff}(s, \alpha)) = v \text{ for some symbol } s,$$

where α is a (possibly infinite) set of strings and $\text{Suff}(s, \alpha)$ is the set of suffix strings y such that $s y \in \alpha$.

With this functional uncertainty, the equation in (12) can then be modified to the following form (Kaplan and Zaenan 1989: 27):

$$(16) \quad (\uparrow\text{TOPIC}) = (\uparrow\text{COMP}^* \text{OBJ})$$

The Kleene star operator ‘*’ indicates that the OBJ can be embedded in any number of COMPS, including zero. This equation is specifically for topicalization of the OBJ function. A more general equation for long distance dependencies is provided in Kaplan and Zaenan (1989: 32):

$$(17) \quad (\uparrow\text{DF}) = (\uparrow\text{body bottom})$$

where DF is the discourse function, body is the uncertainty path, and bottom is the end of the path.

The relevant DFs for topicalization are TOPIC and FOCUS.

As an illustration, the functional uncertainty equation for the Cantonese example in (18) is shown in (19).

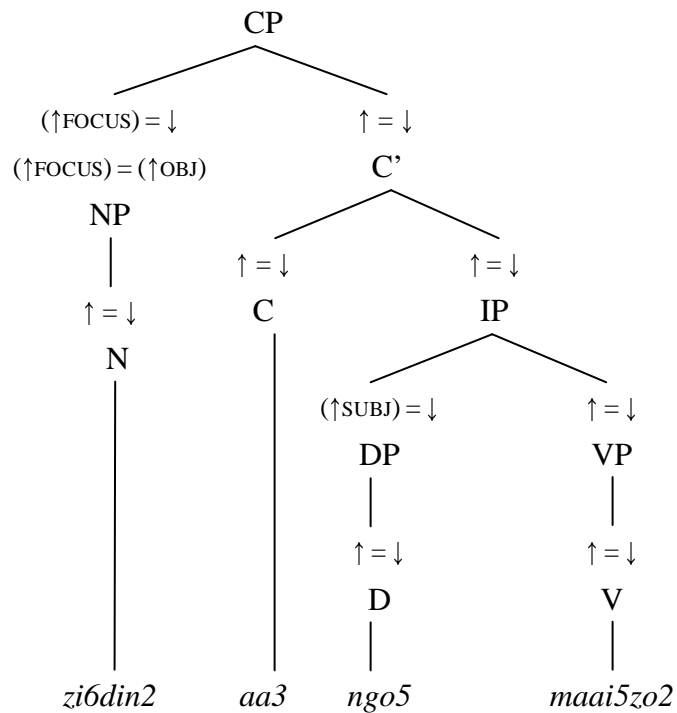
$$(18) \quad \begin{array}{l} \text{字典} \quad \text{啊} \quad \text{我} \quad \text{買咗} \\ \text{zi6din2} \quad \text{aa3} \quad \text{ngo5} \quad \text{maai5zo2} \\ \text{dictionary PART 1.SG} \quad \text{buy.PERF} \\ \text{'It is dictionaries that I have bought.'} \end{array}$$

$$(19) \quad (\uparrow\text{FOCUS}) = (\uparrow\text{OBJ})$$

Equation (19) captures the relation between the leftmost FOCUS and the missing OBJ in (18). It indicates that the FOCUS of the clause *zi6din2* is at the same time the OBJ of the clause. This equation is annotated to the c-structure node of the FOCUS in order to

specify the unification of the topicalized function and the missing function as in (20).

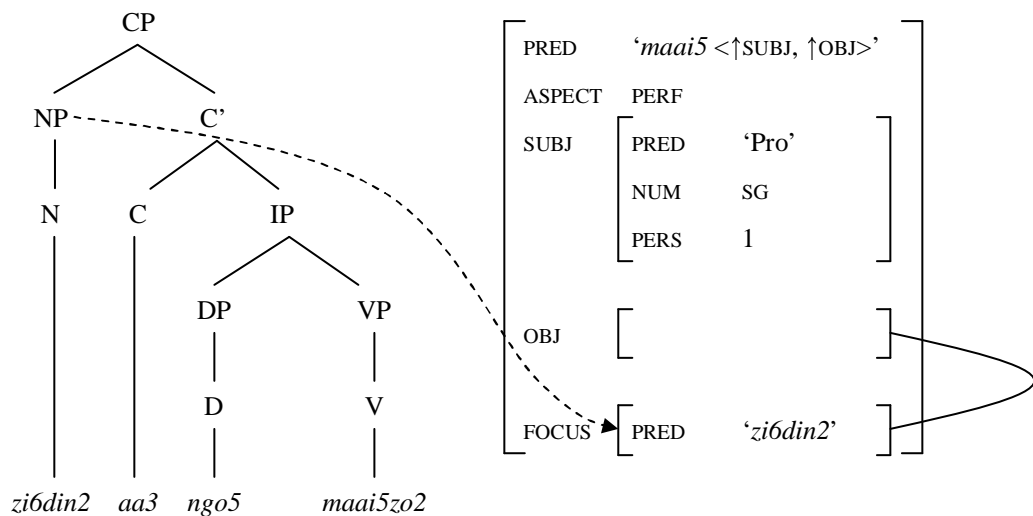
(20) Annotated c-structure of (18)



‘(↑FOCUS) = ↓’ above the NP *zi6din2* specifies that the FOCUS of the sentence is the present NP. ‘(↑FOCUS) = (↑OBJ)’ indicates that the FOCUS of the sentence is at the same time the OBJ of the sentence.

The unification is also shown in the correspondence between the c-structure and f-structure:

(21) Correspondence between the c-structure and f-structure of (18)



The dotted curved line connecting the NP in the c-structure and the FOCUS in the f-structure indicates that the NP corresponds to the FOCUS. The solid curved line in the f-structure linking the FOCUS and the OBJ shows that the two functions unify so that the Completeness condition is still observed though the OBJ function which is subcategorized for by the PRED *maai5* does not have a PRED feature because its PRED feature is provided by the FOCUS. The Extended Coherence Condition is also satisfied with the FOCUS being identified with the OBJ (see chapter 1 for the definition of the two conditions).

The functional uncertainty equation introduced above is an outside-in functional uncertainty, where the relation is traced from the discourse function to the missing function. Also possible is an inside-out approach, in which the relation is traced the other way around, i.e., from the missing function to the discourse function. Falk (2001) discusses the properties of both types of functional uncertainty equation. Unlike the inside-out approach, the outside-in approach annotates the functional uncertainty equation to the discourse function, instead of the missing function. Therefore, no empty node in the c-structure is necessary and Economy of Expression

(see chapter 1 for the definition) is observed. Falk (2001) mentions the disadvantage of the outside-in approach being that a clause without a discourse function in the c-structure will pose problems to the annotation process. He uses an English relative clause without a relative pronoun as an example. It however creates no trouble to the present analysis since the discourse function, which is either a TOPIC or a FOCUS, always exists in topicalization constructions. In this study, the outside-in approach will be employed, maintaining the faithfulness of the c-structure.

After presenting how Functional Uncertainty is derived and how it works, the next section looks specifically at the functional uncertainty equations for topicalization in Cantonese.

3.2.2 Functional Uncertainty Equations for Cantonese Topicalization

In the topicalization constructions presented so far, the TOPIC or FOCUS is identified with the OBJ function of the sentence (except example (5) above and example (2) in chapter 2 which involve multiple topicalized phrases). This section aims at determining the legitimate bottom functions of TOPIC and FOCUS in Cantonese, with a view to establishing the functional uncertainty equations for topicalization. Section 3.2.2.1 presents the methodology. The bottom functions of TOPIC and FOCUS are investigated in sections 3.2.2.2 and 3.2.2.3 respectively.

3.2.2.1 Methodology

In order to find out the legitimate bottom functions of TOPIC and FOCUS, the acceptability of the relevant constructions have to be decided. TOPIC and FOCUS topicalization constructions with different bottom functions were constructed and included in the questionnaire which was completed by 40 Cantonese native speakers (see appendix I). In part I, the informants were given 21 topicalization constructions

and asked to indicate the acceptability of each construction after the interviewer had pronounced them. Marks were assigned to each option and the acceptability of the constructions was determined by the mean score (see chapter 2 for details of the calculations and appendix II for the results). Based on the acceptability of individual sentences, it was then decided which grammatical functions are the legitimate bottom functions of TOPIC and FOCUS and the corresponding functional uncertainty equations were derived. In the following two sections, the relevant constructions will be presented from which the legitimacy of the bottom functions will be concluded.

3.2.2.2 Bottom Functions of TOPIC

This section explores the TOPIC topicalization constructions in Cantonese in an attempt to find out the possible bottom functions of TOPIC. The grammatical functions SUBJ, OBJ, OBJ_θ, XCOMP, OBL, ADJ, PRED and COMP will be in turn investigated.

3.2.2.2.1 The SUBJ Function

Since the canonical position of a SUBJ is already sentence-initial, it seems meaningless to talk about topicalizing the SUBJ function. Topicalization of the SUBJ of the matrix clause can in fact be made explicit with a particle following the SUBJ:

- (22) 佢 啊 瞓咗 喇
 keoi5 *aa4* *fan3zo2* *laa3*
 3.SG PART sleep.PERF PART
 ‘As for him/her, s/he is already asleep.’

In this example, the sentence-initial phrase *keoi5*, being the SUBJ of the clause, is followed by the particle *aa4*, which indicates that it is at the same time the TOPIC. It

is concluded from this example that SUBJ is a legal bottom function of TOPIC.

3.2.2.2.2 The OBJ Function

In most of the examples of TOPIC topicalization presented so far, the TOPIC is linked to the OBJ function (e.g. example (1) above, example (3) in chapter 2). These examples illustrate that OBJ is another legitimate bottom function of TOPIC in Cantonese.

3.2.2.2.3 The OBJ_θ Function

OBJ_θ is the function which is subcategorized for by the PRED in a double-object construction (DOC) together with SUBJ and OBJ. A canonical DOC in Cantonese has the order of SUBJ-PRED-OBJ-OBJ_θ as in the following example, where *keoi5* is the OBJ and *neil go3 man6tai4* is the OBJ_θ:

- (23) 我 問過 佢 呢 個 問題
ngo5 man6gwo3 keoi5 neil go3 man6tai4
 1.SG ask.EXP 3.SG DEFCL question
 ‘I have asked him/her this question.’

Sentence (24) below is an example of topicalization where the TOPIC is identified with the OBJ_θ function:

- (24) 呢 個 問題 啊 我 問過 佢
neil go3 man6tai4 aa4 ngo5 man6gwo3 keoi5
 DEFCL question PART 1.SG ask.EXP 3.SG
 ‘As for this question, I have asked him/her.’

Given that this construction is acceptable, OBJ_θ is also a bottom function of TOPIC.

3.2.2.2.4 The XCOMP Function

XCOMPs in Cantonese can be realized by phrases of a variety of categories such as NP, PP, VP and AP, where VP is the most typical category (Lee 2002). An XCOMP involves either SUBJ control or OBJ control depending on whether its SUBJ is controlled by the SUBJ or the OBJ of the immediate upper clause. In a canonical construction, an XCOMP immediately follows the verb, or the OBJ when there is one. Sentences (25a) and (25b) below are topicalization constructions with the TOPIC being identified with the XCOMP function involving SUBJ control and OBJ control respectively.

- (25) a. 去 睇 戲 啊 佢 應承咗 我 架
heoi3 tai2 hei3 aa4 keoi5 jing1sing4zo2 ngo5 gaa3
 go see movie PART 3.SG promise.PERF 1.SG PART
 ‘As for going to the movies, s/he has promised me.’

- b. 買 嗰 本 書 啊 我 叫過 佢 啦
maai2 go2 bun2 syu1 aa4 ngo5 giu3gwo3 keoi5 laa1
 buy DEFCL book PART 1.SG ask.EXP 3.SG PART
 ‘As for buying the book, I have already asked him/her to do so.’

Due to space limitations, only examples of XCOMPs of the VP category are shown. XCOMP is treated as the same functional notion regardless of its category. Therefore, by proving that XCOMP of the VP category can be topicalized as a TOPIC, it is concluded that XCOMP should be included in the set of bottom functions of TOPIC in Cantonese.

3.2.2.2.5 The OBL Function

The OBL function does not have a fixed canonical position. For example, it can occur postverbally as in (26a) or preverbally as in (26b).

- (26) a. 佢 住過 嘍 香港 架
keoi5 zyu6gwo3 hai2 Hoeng1gong2 gaa3
 3.SG live.EXP in Hong Kong PART
 ‘S/he used to live in Hong Kong.’
- b. 我 同 佢 傾咗 五 分鐘 咋
ngo5 tung4 keoi5 king1zo2 ng5 fan1zung1 zaa3
 1.SG with 3.SG talk.PERF five minute PART
 ‘I have talked with him/her only for five minutes.’

The following shows the topicalization counterparts of the above two constructions where the TOPIC is interpreted as filling the OBL function:

- (27) a. 嘍 香港 啊 佢 住過 架
hai2 Hoeng1gong2 aa4 keoi5 zyu6gwo3 gaa3
 in Hong Kong PART 3.SG live.EXP PART
 ‘In Hong Kong, s/he used to live.’
- b. 同 佢 啊 我 傾咗 五 分鐘 咋
tung4 keoi5 aa4 ngo5 king1zo2 ng5 fan1zung1 zaa3
 with 3.SG PART 1.SG talk.PERF five minute PART
 ‘With him/her, I have talked only for five minutes.’

The TOPIC *hai2 Hoeng1gong2* in sentence (27a) is linked to the locative OBL while the TOPIC *tung4 keoi5* in sentence (27b) is linked to the comitative OBL.

Huang (1992: 262) gives a Mandarin example where the sentence-initial PP can only be interpreted as the matrix ADJ but not as an embedded locative PP and concludes that OBL should be excluded from the set of bottom functions of TOPIC. Sentence (28) is the corresponding example in Cantonese:

- (28) 係 德國 我 睇到 [電視 報導 [奧運
hai2 Dak1gwok3 ngo5 tai2dou2 [din6si6 bou3dou6 [Ngou3wan6
 in Germany 1.SG see television report the Olympics
 已經 開始咗]]
ji5ging1 hoi1ci2zo2]]
 already start.PERF

‘In Germany, I saw the TV report that the Olympics has already started.’

In part II of the questionnaire (see appendix I), two to three interpretations were proposed for three topicalization constructions respectively (sentences (28), (29) and (30)) and the informants were asked to indicate the acceptability of the interpretations with respect to the given constructions. According to the results of the survey, in (28) above, *Dak1gwok3* (*Germany*) can only be interpreted as the place where the speaker saw the TV report, but not as the place where the TV reported or where the Olympics has started. In other words, like Huang’s (1992) Mandarin example, the sentence-initial PP *hai2 Dak1gwok3* cannot be linked to any embedded function. Even if it can, however, it is the ADJ function, instead of the OBL function, to which it should be linked. Therefore, it cannot be concluded from this example that OBL is not a legal bottom function.

In fact, there are examples of a sentence-initial PP being identified with the OBL in a subordinate clause in Cantonese. According to the results of the survey, in (29) below, *nei1 gaan1 fong2* (*this room*) should be interpreted as the place where the pot of flowers has been put. Therefore, the TOPIC *hai2 nei1 gaan1 fong2* is interpreted as filling the OBL function in the embedded clause.

- (29) 嚟 呢 間 房 我 估 [佢 擺 咗 盆 花]
hai2 nei1 gaan1 fong2 ngo5 gu2 [keoi5 baai2zo2 pun4 faa1]
 in DEF CL room 1.SG guess 3.SG put.PERF CL flower
 ‘In this room, I guess s/he has put a pot of flowers.’

Huang (1992) argues that, in Mandarin, a PP occurring in the sentence-initial position should be conceived as an ADJ of the matrix clause instead of a TOPIC, unless the sentence is not interpretable with the PP being treated as the matrix ADJ, in which case the speaker will link the PP to an embedded function in order to find a reasonable interpretation. The present study, however, will not investigate under what circumstances a TOPIC is to be interpreted as filling the OBL function and when the ADJ function. Since it has been shown above that there are situations where the TOPIC and the OBL identify, OBL is concluded as a bottom function of TOPIC in Cantonese.

3.2.2.2.6 The ADJ Function

Like the OBL function, ADJ in Cantonese may occur in different positions. For instance, locative ADJs usually appear right after the SUBJ while frequency expressions may appear immediately after the verb or the OBJ. According to Li and Thompson (1981), time and locative phrases acting as ADJs can be topics. An ADJ appearing in the sentence-initial position is treated in this study as a TOPIC which fills the ADJ function, rather than simply an ADJ of the matrix clause as argued by Huang (1992). Therefore, example (28) above is an instance of ADJ topicalization. While the TOPIC in example (28) cannot be linked to an embedded ADJ, there are cases where the TOPIC and an embedded ADJ identify. Consider the following example:

- (30) 嚟 香港 我 以為 [佢 有 三 間 屋]
hai2 Hoeng1gong2 ngo5 ji5wai4 [keoi5 jau5 saam1 gaan1 nguk1]
 in Hong Kong 1.SG think 3.SG have three CL house
 ‘In Hong Kong, I thought s/he has three houses.’

In a canonical construction, the ADJ *hai2 Hoeng1gong2*⁸ occurs either right after the SUBJ of the subordinate clause (as in *ngo5 ji5wai4 keoi5 hai2 Hoeng1gong2 jau5 saam1 gaan1 nguk1*) or right after the OBJ (as in *ngo5 ji5wai4 keoi5 jau5 saam1 gaan1 nguk1 hai2 Hoeng1gong2*). In this example, the ADJ of the COMP is extracted to the sentence-initial position as the TOPIC. Similar to the case of OBL, the identification of the TOPIC and the ADJ might be due to the impossibility of interpreting the TOPIC as the ADJ of the matrix clause with the PRED *ji5wai4*. This however does not render untenable the conclusion that ADJ is one of the bottom functions of TOPIC.

3.2.2.2.7 The PRED Function

Matthews and Yip (1994) in their study of Cantonese mention that, when a verb is fronted (‘topicalized’ in their sense), it is typically repeated, or replaced by the copula verb *hai6*, in its canonical position as in (31) where the verb *maai5* is repeated. Topicalizing the verb by leaving the canonical position of the verb empty leads to ungrammaticality as in (32).

- (31) 買 呢 我 就 會 買 呢 本 書 □⁹
maai5 nei ngo5 zau6 wui3 maai5 nei1 bun2 syu1 ge2
 buy PART 1.SG zau will buy DEFCL book PART
 ‘As for buying, I’ll buy that book.’

⁸ Note that the PP *hai2 Hoeng1gong2* here is optional and is therefore an ADJUNCT while the same PP in (26a) and (27a) is subcategorized for by the PRED *zyu6* and is an OBL.

⁹ Particles having no corresponding Chinese characters are represented by the symbol ‘□’.

- (32) *買 呢 我 就 會 呢 本 書 □
 **maai5* *ne1* *ngo5* *zau6* *wui3* *ne1* *bun2* *syu1* *ge2*
 buy PART 1.SG zau will DEF CL book PART

Example (32) seems to suggest that PRED is not a possible bottom function of TOPIC. In fact, there are instances where a TOPIC and a PRED unify as in the following example.

- (33) 喊 呢 我 就 唔 會
haam3 *ne1* *ngo5* *zau6* *m4* *wui3*
 cry PART 1.SG zau not will
 ‘As for crying, I won’t.’

The PRED *haam3* is an intransitive verb. With the auxiliary *wui3* staying in its canonical position, the PRED can be topicalized to become a TOPIC. Though the PRED is not always a legitimate bottom function, it is still included here as one of the possible bottom functions of TOPIC. Unacceptable constructions like sentence (32) should be accounted for in terms of violations of some other constraints, which will be left for future research.

3.2.2.2.8 The Illegitimate Bottom Function: COMP

After identifying the legitimate bottom functions of TOPIC, the illegitimate function, COMP, will be discussed presently. The COMP function immediately follows the verb in a canonical construction as illustrated in (34). The PRED *ji5wai4* subcategorizes for a COMP.

- (34) 我 以爲 [你 買咗 本 書]
ngo5 ji5wai4 [nei5 maai5zo2 bun2 syu1]
 1.SG think 2.SG buy.PERF CL book
 'I thought you'd bought a/the book.'

The sentence is ill-formed when the COMP is topicalized to be the TOPIC:

- (35) *你 買咗 本 書 啊 我 以爲
 **nei5 maai5zo2 bun2 syu1 aa4 ngo5 ji5wai4*
 2.SG buy.PERF CL book PART 1.SG think

The unacceptability of this construction reveals that COMP is not a legitimate bottom function of TOPIC in Cantonese.

In the above sections, the legitimate bottom functions of TOPIC in Cantonese have been identified, which include SUBJ, OBJ, OBJ_θ, XCOMP, OBL, ADJ and PRED. The COMP function has been shown to be illegitimate. The following functional uncertainty equation summarizes the findings:

- (36) $(\uparrow\text{TOPIC}) = (\uparrow\{\text{SUBJ, OBJ, OBJ}_\theta, \text{XCOMP, OBL, ADJ, PRED}\})$

This equation specifies that the TOPIC of the sentence is at the same time any of the functions in the curly brackets.

3.2.2.3 Bottom Functions of FOCUS

After proposing a functional uncertainty equation for TOPIC topicalization, FOCUS topicalization constructions will be investigated in this section and the bottom functions will be identified.

3.2.2.3.1 The SUBJ Function

Sentence (37) is a topicalization construction where the FOCUS *keoi5* (note the particle *aa3*) is identified with the SUBJ.

- (37) 佢 啊 考 第一
keoi5 aa3 haau2 dai6jat1
 3.SG PART exam first
 ‘It is s/he who got the first prize.’

Like the case of TOPIC constructions, the FOCUS status of the SUBJ is indicated by the additional particle *aa3*. Given the acceptability of this example, it is concluded that a FOCUS can be identified with the SUBJ of a clause.

3.2.2.3.2 The OBJ Function

Sentence (18) above and sentence (8A) in chapter 2 are examples of topicalization with the FOCUS being identified with the OBJ. OBJ is therefore a legal bottom function of FOCUS.

3.2.2.3.3 The OBJ_θ Function

An example of topicalization is shown below with the FOCUS filling the missing OBJ_θ function:

- (38) 英文 啊 我 教 佢
jing1man2 aa3 ngo5 gaau3 keoi5
 English PART 1.SG teach 3.SG
 ‘It is English that I teach him/her.’

Given that this sentence is acceptable, OBJ_θ is concluded as another function which

can be identified with a FOCUS.

3.2.2.3.4 The COMP Function

In (39) below, the FOCUS is linked to the COMP function (note that the particle following the topicalized phrase is different from the one in (35)).

- (39) 你 買咗 本 書 啊 我 以為
nei5 maai2zo2 bun2 syu1 aa3 ngo5 ji5wai4
 2.SG buy.PERF CL book PART 1.SG think
 ‘I thought YOU HAD BOUGHT THE BOOK.’

Unlike the case of TOPIC topicalization, COMP can appear in the sentence-initial position as a FOCUS.

3.2.2.3.5 The XCOMP Function

Sentences (40a) and (40b) are topicalization constructions with the FOCUS being linked to the XCOMP functions involving SUBJ control and OBJ control respectively.

- (40) a. 去 睇 戲 架 佢 應承咗 我
heoi5 tai2 hei3 gaa3 keoi5 jing1sing4zo2 ngo5
 go see movie PART 3.SG promise.PERF 1.SG
 ‘To go to the movies, s/he has promised me.’
- b. 買 嗰 本 書 架 我 提咗 佢
maai5 go2 bun2 syu1 gaa3 ngo5 tai4zo2 keoi5
 buy DEFCL book PART 1.SG remind.PERF 3.SG
 ‘To buy that book, I have reminded him/her.’

Both constructions are acceptable and XCOMP is thus another legitimate bottom function of FOCUS.

3.2.2.3.6 The OBL Function

Example (41) is an instance of topicalization where the FOCUS is identified with the OBL.

- (41) **喺 張 枱 啊 我 放 咗 本 書**
hai2 zoeng1 toi2 aa3 ngo5 fong3zo2 bun2 syu1
 on CL table PART 1.SG put.PERF CL book
 ‘It is on the table that I have put the book.’

This is again an acceptable construction and OBL is therefore a function which can be extracted to become the FOCUS of a clause.

3.2.2.3.7 The ADJ Function

The following is an example where the sentence-initial FOCUS is identified with the ADJ function of the clause.

- (42) **琴日 啊 我 去 睇 戲**
kam4jat6 aa3 ngo5 heoi3 tai2 hei3
 yesterday PART 1.SG go see movie
 ‘It was yesterday that I went to the movies.’

Given the acceptability of this sentence, ADJ is concluded as another legal bottom function of FOCUS.

3.2.2.3.8 The PRED Function

Sentences (43a) and (43b) are two examples of topicalization with the FOCUS being identified with the PRED, transitive in (43a) and intransitive in (43b):

- (43) a. 買咗 啦 我 個 本 書
maai5zo2 laa1 ngo5 go2 bun2 syu1
 buy.PERF PART 1.SG DEF CL book
 ‘I HAVE already BOUGHT that book.’
- b. 瞓咗 啦 佢
fan3zo2 laa1 keoi5
 sleep.PERF PART 3.SG
 ‘S/he’s already ASLEEP.’

Unlike the case of TOPIC constructions which only allow the TOPIC to be linked to an intransitive PRED, a FOCUS can be linked to both transitive and intransitive PREDs.

It has been illustrated above that all of the grammatical functions examined, SUBJ, OBJ, OBJ_θ, COMP, XCOMP, OBL, ADJ and PRED, are legitimate bottom functions of FOCUS in Cantonese. The following functional uncertainty equation is proposed:

$$(44) \quad (\uparrow \text{FOCUS}) = (\uparrow \{ \text{SUBJ, OBJ, OBJ}_{\theta}, \text{COMP, XCOMP, OBL, ADJ, PRED} \})$$

Compare this functional uncertainty equation with the one for TOPIC topicalization in (36), repeated as (45).

$$(45) \quad (\uparrow \text{TOPIC}) = (\uparrow \{ \text{SUBJ, OBJ, OBJ}_{\theta}, \text{XCOMP, OBL, ADJ, PRED} \})$$

Note that there is an asymmetry between the two equations. COMP is a legitimate bottom function of FOCUS, but not of TOPIC. Though they are both topicalized phrases, a TOPIC and a FOCUS behave differently with respect to the grammatical functions to which they can be identified.

In the above sections, the essence of the functional device for characterizing topicalization has been presented and two functional uncertainty equations for

topicalization in Cantonese have been established by identifying the legal bottom functions of TOPIC and FOCUS. The next section discusses another approach to topicalization, the informational approach.

3.3 An Independent I-Structure for Discourse Information

King (1997) identifies the limitations of using the standard functional annotations to characterize discourse functions with illustrations from Russian and English data and postulates an independent i-structure for accommodating discourse information. The limitations of the functional approach and the solutions to the problems will be presented in section 3.3.1. In section 3.3.2, the topicalization constructions in Cantonese will be investigated to show how the informational approach complements the functional approach to give a satisfactory account.

3.3.1 Limitations of the Functional Account and Some Solutions

King (1997) points out the problems of licensing discourse focus by functional annotations. She states that a wrong scope may be resulted if the element designating the discourse focus is an f-structure head. She illustrates this problem with a Russian example where the PRED is the discourse focus. She mentions two alternatives to characterize the discourse focus and points out their weaknesses. The first proposal is to annotate the PRED with $\downarrow \in (\uparrow \text{FOC})$ ¹⁰. This annotation will however lead to the wrong conclusion that the whole f-structure is the discourse focus since the annotation to the PRED is passed to the entire f-structure. Another alternative is to specify that it is only the PRED which is the discourse focus with the annotation $(\downarrow \text{PRED}) \in (\uparrow \text{FOC})$. The problem of this approach is that the arguments of the PRED are included inside the discourse focus with the PRED itself, which is not the intended

¹⁰ The set symbol '∈' is employed since multiple instances of discourse focus are allowed in Russian.

interpretation. In order to resolve this scoping problem, King employs the annotation ‘PRED FN’ which refers to the basic meaning of the PRED (Kaplan and Maxwell 1996 cited in King 1997), excluding its argument structure.

Apart from the scoping problem discussed above, King (1997) notices the mismatch between f-structure and i-structure constituents through examining the it-cleft constructions in English. The clefted part of a construction, though being obliged to contain the discourse focus, is not obliged to contain *only* the discourse focus. The phrase being clefted is an f-structure constituent. It however does not necessarily correspond to an i-structure constituent since it may consist of both the discourse focus and the background information. Based on the observation that an i-structure element does not always correspond to an f-structure element, King (1997) introduces a separate i-structure to encode discourse information, independent of the f-structure.

This section has outlined the limitations of the functional account and some solutions to the problems. In the next section, topicalization constructions in Cantonese will be investigated in order to show whether the two problems also exist in Cantonese and how the informational account complements the functional account in characterizing TOPIC and FOCUS.

3.3.2 Characterizing Topicalization in Cantonese

Problems similar to the two mentioned by King (1997), the mismatch between f-structure and i-structure constituents and the failure of functional annotations to account for constructions with the PRED being the discourse focus, can also be found in Cantonese. A few examples will be given below to illustrate the problems.

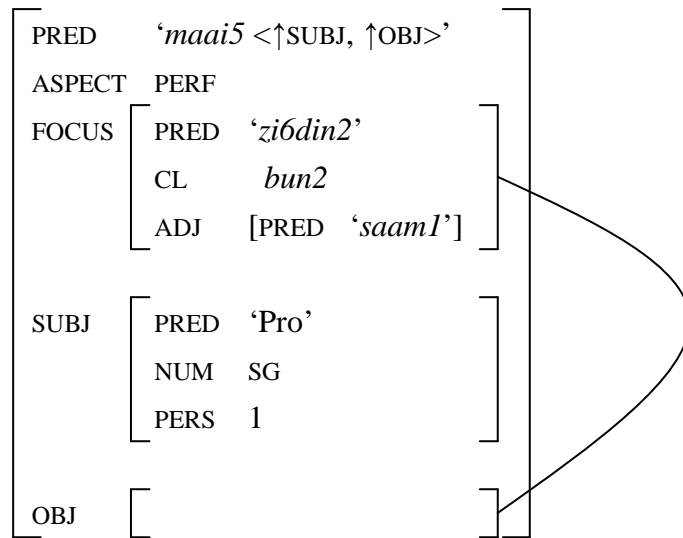
As mentioned in chapter 2, FOCUS in Cantonese, though being obliged to contain some [+New] information, does not necessarily contain only [+New]

information. This phenomenon has a similar effect as the English it-cleft constructions in that it demonstrates the mismatch between f- and i-structure constituents. Consider the question-answer pair in example (9) in chapter 2 again, repeated as (46):

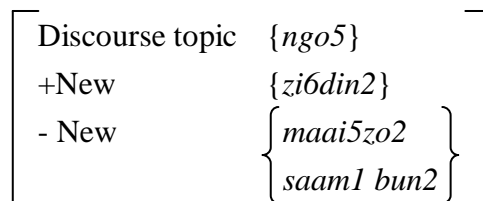
- (46) Q: 你 買咗 三 本 咩 啊?
nei5 maai5zo2 saam1 bun2 me1 aa3?
 2.SG buy.PERF three CL what PART
 ‘What are the three books that you have bought?’
- A: [三 本 字典] 啊 我 買咗
[saam1 bun2 zi6din2] aa3 ngo5 maai5zo2
 three CL dictionary PART 1.SG buy.PERF
 ‘It is three DICTIONARIES that I have bought.’

In the answer, the OBJ *saam1 bun2 zi6din2* is topicalized to become the FOCUS and is thus an f-structure constituent. Given the question asking three copies of what (*saam1 bun2 me1*) have been bought, only *zi6din2* is the [+New] information and thus an i-structure constituent. This example illustrates that an f-structure element does not always correspond to an i-structure element and this is where an independent i-structure is necessary. The f-structure and i-structure of the construction are shown below:

(47) F-structure of (46A)



(48) I-structure of (46A)



The i-structure here differs from King’s (1997) in a few ways. ‘Discourse topic’ is used instead of ‘TOP’ or ‘TOPIC’ to show that *ngo5* is the phrase designating the discourse topic. ‘TOPIC’ is reserved for *topicalized* discourse topic and appears in the f-structure. The same applies to ‘FOCUS’ and the symbol ‘FOC’ or ‘FOCUS’ is therefore avoided in the i-structure. ‘+New’ is used instead to denote discourse focus (see chapter 2). For consistency reason, ‘-New’, instead of ‘BCK’, is employed to refer to background information.

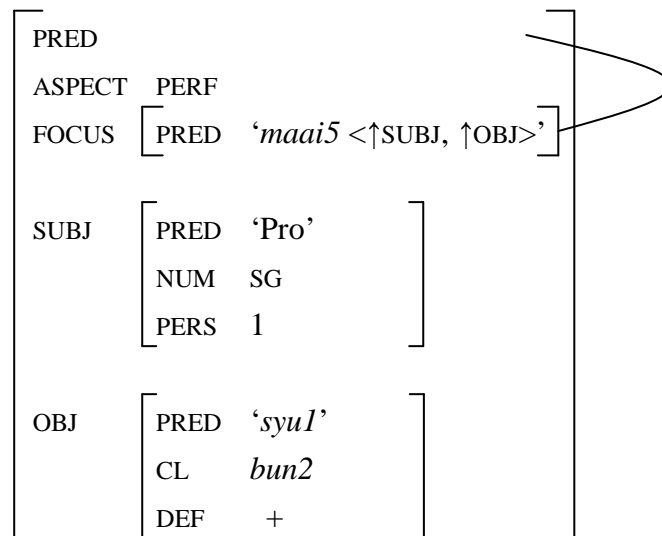
King (1997), after introducing an independent i-structure, raises the question of whether there should be any discourse function information in the f-structure. This study takes Bresnan and Mchombo’s (1987) suggestion that TOPIC and FOCUS are *grammaticized* discourse functions and therefore should be present in the f-structure.

Now Consider example (43a) again, repeated here as (49). In this construction, only the PRED is topicalized to become the FOCUS.

- (49) 買咗 啦 我 個 本 書
maai5zo2 laa1 ngo5 go2 bun2 syu1
 buy.PERF PART 1.SG DEF CL book
 ‘I HAVE already BOUGHT that book.’

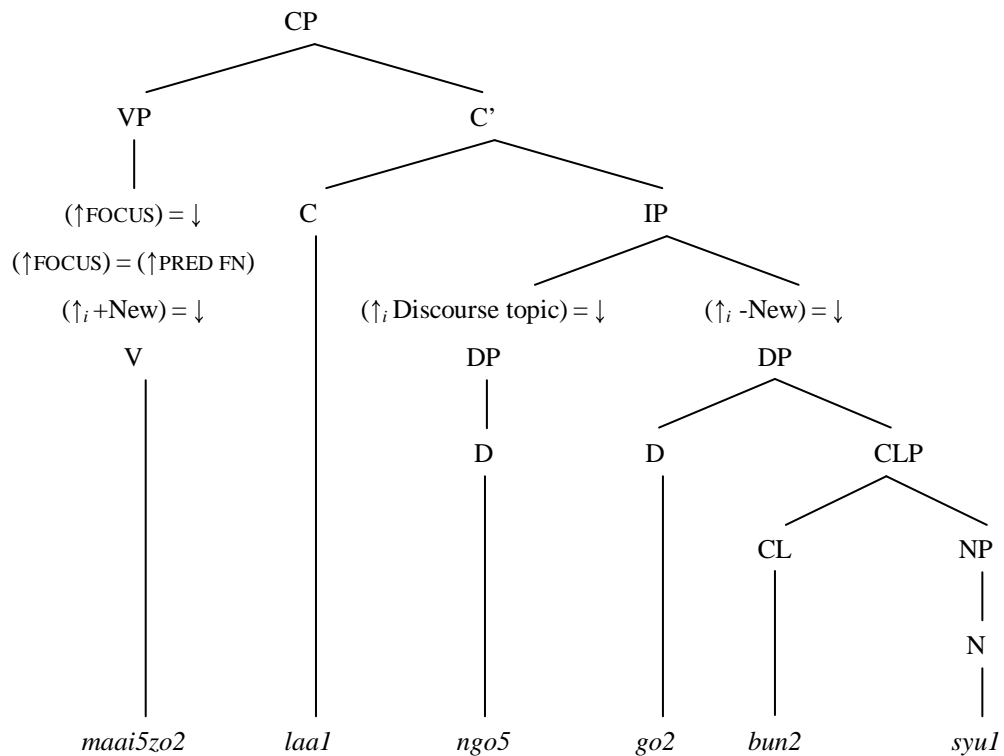
If the equation (\uparrow FOCUS) = (\uparrow PRED) is annotated to the V node *maai5zo2*, the FOCUS includes not only the PRED, but also its arguments, SUBJ and OBJ, as shown in the following f-structure:

- (50) Tentative f-structure of (49)



The ‘PRED FN’ notation solves the scoping problem:

(51) Annotated c-structure of (49)¹¹

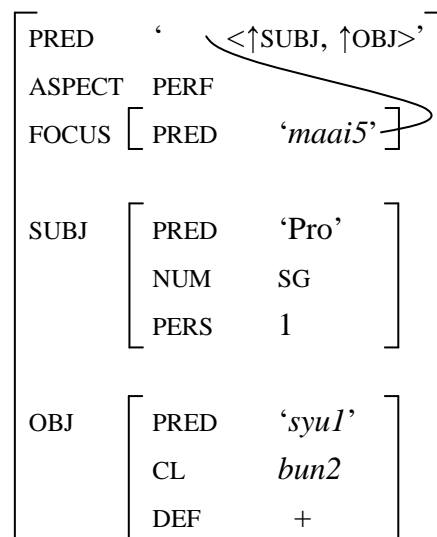


Due to space limitations, only the relevant annotations are shown here. This study will not discuss whether multiple TOPICS and multiple FOCUSES are allowed in Cantonese and, therefore, the ‘=’ symbol is used instead of the ‘∈’ symbol. Consider the V *maai5zo2*. The ‘(↑FOCUS) = ↓’ annotation specifies that the FOCUS of the sentence is the present V node. The ‘(↑FOCUS) = (↑PRED FN)’ indicates that the FOCUS of the sentence is at the same time the PRED of the matrix f-structure, that is *maai5*, excluding its argument. The ‘(↑i+New) = ↓’ designates that the present V node carries [+New] information. ‘PRED FN’ is not necessary for the DP *go2 bun2 syul* since it is the whole DP which bears [-New] information and it takes no argument. By the same token, the DP *ngo5* only needs the down-arrow to refer to the whole DP as the discourse topic.

¹¹ The subscript ‘i’ indicates non-f-structure projections (King 1997).

After solving the scoping problem in the annotated c-structure, it has to be decided how the f-structure of this construction looks like. It should be noted that this Cantonese example is different from King's (1997) Russian contrastive discourse focus example in that the PRED in the Cantonese example is a FOCUS (i.e., a topicalized function) while the one in Russian is an *in situ* contrastive discourse focus with no word order change. The FOCUS in the Cantonese construction, being a grammaticized discourse function, should be included in the f-structure. The following f-structure is proposed:

(52) Revised f-structure of (49)



Since a FOCUS has to be linked to a missing function and the FOCUS here is the PRED FN, there should be a representation identifying the FOCUS with only the PRED but not its arguments. This f-structure is proposed where the FOCUS is linked to the core meaning of the PRED, excluding its arguments, by connecting the FOCUS to the empty slot of the PRED in front of the subcategorization frame.

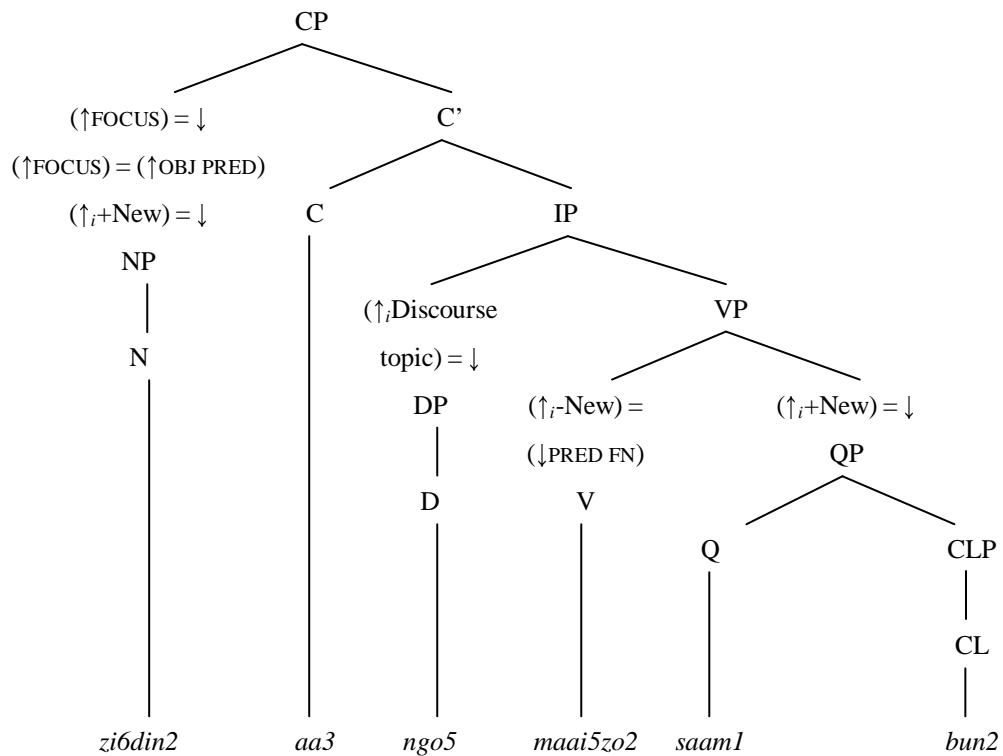
Another example is given below to illustrate the application of an independent i-structure and the ‘PRED FN’ annotation. Consider the question-answer pair (10) in

chapter 2 again, repeated as (53):

- (53) Q: 你 買咗 咩 啊?
nei5 maai5zo2 me1 aa3?
 2.SG buy.PERF what PART
 ‘What have you bought?’
- A: 字典 啊 我 買咗 三 本
zi6din2 aa3 ngo5 maai5zo2 saam1 bun2
 dictionary PART 1.SG buy.PERF three CL
 ‘I have bought three DICTIONARIES.’

In this example, *saam1 bun2 zi6din2* is the OBJ and the [+New] information of the clause. Only the NP *zi6din2* is topicalized to become the FOCUS, leaving the Q⁰ *saam1* and the CL⁰ *bun2 in situ*. The annotation (\uparrow FOCUS) = (\uparrow OBJ) fails to account for this construction since it gives the wrong interpretation of the entire OBJ being identified with the FOCUS. The alternative (\uparrow FOCUS) = (\uparrow OBJ PRED) is employed here as shown in the following annotated c-structure:

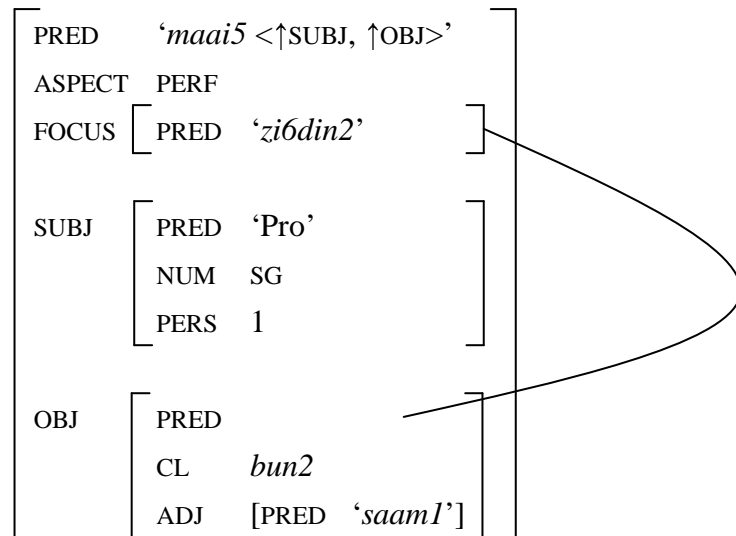
(54) Annotated c-structure of (53A)



Since the OBJ PRED has no subcategorization frame, the annotation ‘(↑FOCUS) = (↑OBJ PRED)’ represents the correct scoping by stating that the FOCUS is at the same time the PRED of the OBJ of the matrix f-structure. The ‘PRED FN’ at the V node is necessary to show that only the core meaning of the PRED of the V is [-New] information, excluding its SUBJ and OBJ.

The following f-structure is proposed for the construction:

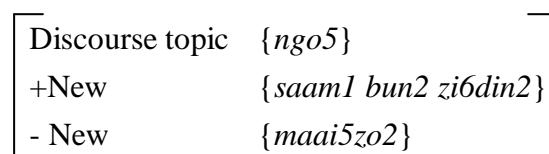
(55) F-structure of (53A)



In this f-structure, the FOCUS is linked to the empty PRED of the OBJ, while the CL and the ADJ remain in the OBJ. Compare this account with Kuhn’s (1999) account of split NPs in German. Kuhn (1999) links in the f-structure an empty TOPIC to the OBJ. The present study assumes that a topicalized constituent appears inside the f-structure of the TOPIC/FOCUS. Therefore, with *zi6din2* being topicalized, it appears inside the f-structure of the FOCUS. *Saam1* and *bun2* which remain in the canonical OBJ position appear inside the f-structure of the OBJ.

Finally, the i-structure of (53A) is shown below:

(56) I-structure of (53A)



Again, the f-structure constituent, *zi6din2*, which is the FOCUS, and the i-structure constituent, *saam1 bun2 zi6din2*, which is the [+New] information, can be captured

clearly in two separate levels of representation.

3.4 Summary

This chapter has analyzed topicalization in Cantonese within an LFG approach. The c-structure configuration of TOPIC, FOCUS and E-TOP constructions have been presented. A functional account has been proposed to characterize the topicalization phenomenon, resulting in the following two functional uncertainty equations:

$$(57) \quad (\uparrow\text{TOPIC}) = (\uparrow\{\text{SUBJ, OBJ, OBJ}_\theta, \text{XCOMP, OBL, ADJ, PRED}\})$$

$$(58) \quad (\uparrow\text{FOCUS}) = (\uparrow\{\text{SUBJ, OBJ, OBJ}_\theta, \text{COMP, XCOMP, OBL, ADJ, PRED}\})$$

It has however been shown that a purely functional approach fails to give a satisfactory account of the phenomenon. An independent i-structure (King 1997) has been employed to encode discourse information, separated from the f-structure, and the ‘PRED FN’ notation (Kaplan and Maxwell 1996) has been adopted to give a correct scoping of grammaticized discourse functions and discourse information. It has been suggested that the grammaticized discourse functions, TOPIC and FOCUS, should be included in the f-structure and the unification between a topicalized phrase and the PRED FN can be represented in the f-structure by connecting the topicalized phrase to the empty slot of the PRED only, with its arguments being excluded.

In the next chapter, an OT-LFG account will be given to the phenomenon of topicalization in Cantonese.

Chapter 4 TOPIC and FOCUS: An OT-LFG Account of OBJ Topicalization

OBJ is one of the legitimate bottom functions of TOPIC and FOCUS in Cantonese and it has been shown in chapters 2 and 3 that there are instances where only part of the OBJ is topicalized. This chapter examines OBJ topicalization constructions in Cantonese within an OT-LFG framework, proposing constraint hierarchies for the language¹².

This chapter is organized as follows. Section 4.1 presents the methodology. Section 4.2 outlines the properties of Cantonese topicalization with an OBJ bottom, according to which the relevant constraints are identified and described in section 4.3. The constraints are ranked in section 4.4 based on a detailed analysis of the syntactic phenomenon. Section 4.5 summarizes the chapter.

4.1 Methodology

The acceptability of the constructions discussed in this chapter is verified by the results of part III of the questionnaire survey (see appendix I for the questionnaire and appendix II for the results; see chapter 2 for the details of the survey). With the aid of the judgment of my informants on the data, some properties of OBJ topicalization in Cantonese are outlined, based on which the relevant constraints are identified and ranked against each other. Several tableaux are then presented to demonstrate how the optimal outputs are selected and the constraint rankings are revised whenever necessary.

¹² Part of this chapter has been presented by the author at the Fifth Cambridge Postgraduate Conference in Language Research (Fung 2007).

4.2 Properties of OBJ Topicalization in Cantonese

Some properties of OBJ topicalization in Cantonese will be outlined in this section based on the results of the questionnaire survey. The illustrations of these properties will be presented in the succeeding sections throughout the discussion of constraints and their rankings.

When the phrase bearing the OBJ function is at the same time the discourse topic of the sentence, it can either be extracted to the initial position to become the TOPIC or stay in its canonical post-verbal position. The case is similar for FOCUS. The OBJ carrying [+New] information can either be topicalized to become the FOCUS or stay in its canonical position. A FOCUS may contain some [-New] information together with the [+New] element and it is not obligatory that all [+New] information appear inside the FOCUS (see chapter 2).

OBJ is one of the grammatical functions which can be topicalized to become a TOPIC or a FOCUS and it is not necessary that the entire OBJ be topicalized. With an OBJ consisting of a Q^0 , a CL^0 and an NP, the NP can be topicalized to become the TOPIC or the FOCUS, being separated from the CL^0 . It is however impossible to extract the CLP (consisting of a CL^0 and an NP), leaving the Q^0 *in situ*. Whenever the CL^0 is topicalized, it has to be topicalized together with its NP complement. Constructions with the NP complement being left *in situ* are ill-formed.

Based on these observations on OBJ topicalization in Cantonese, the relevant constraints will be identified in the next section.

4.3 The Relevant Constraints

According to the above observations, the following constraints are found to be relevant to the current analysis:

- (1)
- a. **SUBJ-L**: SUBJ aligns left in the clause (Morimoto 2001)
 - b. **DISTOPIC-L**: Phrase designating the discourse topic aligns left in the sentence
 - c. **Abut-OBJ(V-HD)**: Abut Edge₁ of OBJ with Edge₂ of V⁰ (Morimoto 2001)
 - d. **Abut-NP(CL-HD)**: Abut Edge₁ of NP with Edge₂ of CL⁰
 - e. **Abut-CLP(Q-HD)**: Abut Edge₁ of CLP with Edge₂ of Q⁰
 - f. **[+New]-FOCUS \wedge FOCUS-L**: When there exists some [+New] information in the sentence, there must be a FOCUS and the FOCUS aligns left in the sentence
 - g. ***TOPICALIZE-FC**: Do not topicalize functional categories

Each of the constraints will be discussed in the following sections. Section 4.3.1 looks at the alignment constraints (i.e., SUBJ-L and DISTOPIC-L). Section 4.3.2 investigates the abutment constraints (i.e., Abut-OBJ(V-HD), Abut-NP(CL-HD) and Abut-CLP(Q-HD)). The [+New]-FOCUS \wedge FOCUS-L and *TOPICALIZE-FC constraints are discussed in sections 4.3.3 and 4.3.4 respectively.

4.3.1 The Alignment Constraints

SUBJ-L and DISTOPIC-L are alignment constraints. They are proposed based on Generalized Alignment (McCarthy and Prince 1993 cited in Morimoto 2001: 155):

- (2)
- Align (Cat₁, Edge₁, Cat₂, Edge₂) = *def*
- \forall Cat₁ \exists Cat₂ such that Edge₁ of Cat₁ and Edge₂ of Cat₂ coincide.
- Where Cat₁, Cat₂ \in ProsCat \cup GramCat
- Edge₁, Edge₂ \in {Right, Left}

Cat₁ and Cat₂ are some prosodic categories or grammatical categories. Edge₁ and Edge₂ are the right or left edge. According to the above definition, Edge₁ of Cat₁ has to coincide with Edge₂ of Cat₂. A scalar interpretation of alignment constraints is assumed in the present study so that every element (to be defined below) intervening between the specified edges incurs one violation to the constraint (Kuhn 2003).

Align (SUBJ, L, β , L), which is abbreviated as SUBJ-L, requires the left edge of the SUBJ to coincide with the left edge of β , where β is the natural domain for grammatical function. Natural domain is defined as follows (Sells 1998: 20):

- (3) The immediate mother M of an element E in representation R is the immediately superior constituent domain containing E . M constitutes the natural domain for E .

Therefore, the natural domain for SUBJ is the immediate outer f-structure containing the SUBJ. In order to satisfy SUBJ-L, the f-structure of SUBJ has to align with this immediate outer f-structure. F-alignment (Sells 1998: 11) explains how two f-structure elements align:

- (4) For two f-structure elements f_1 and f_2 , f_1 *left-aligns* with f_2 if and only if the left edge of the maximal node that maps onto f_1 shares a left edge with the maximal node that maps on to f_2 .

Hence, SUBJ-L requires the left edge of the maximal projection mapping onto the f-structure of the SUBJ to share a left edge with the maximal projection mapping onto the immediate outer f-structure containing the SUBJ (Morimoto 2001).

Since one of the main interests of this thesis is to find out when and how a

discourse topic is topicalized to become a TOPIC, the DISTOPIC-L constraint is proposed, which designates where the phrase representing the discourse topic should occur in a sentence, to be differentiated from TOP-L (Sells 2000) which deals with the position of the TOPIC function. DISTOPIC-L requires that the left edge of the phrase designating the discourse topic align with the left edge of its natural domain. Discourse topic is a discourse notion, instead of a grammatical function. Therefore, it has a different natural domain from SUBJ. Though a discourse topic usually stretches across a number of utterances, what is being dealt with in this study is the phrase *in a sentence* which designates the discourse topic. Therefore, the natural domain for discourse topic is defined as the immediate sentence containing the phrase corresponding to the discourse topic. DISTOPIC-L thus demands that the left edge of the phrase designating the discourse topic share the left edge of the immediate sentence containing that phrase.

4.3.2 The Abutment Constraints

The general definition of abutment is as follows (Morimoto 2001: 164):

- (5) $ABUT(C_1, Edge_1, C_2, Edge_2) =_{def}$
 C_1 abuts with C_2 if and only if $Edge_1$ of C_1 shares $Edge_2$ of C_2 where $Edge_1 \neq Edge_2$.
 Abbreviated as $Abut-C_1(C_2)$.

In the above definition, C_1 and C_2 refer to some categories and $Edge_1$ and $Edge_2$ refer to the left edge and the right edge respectively or vice versa. The constraint is satisfied if the opposite edges of the two categories are adjacent to each other. The number of elements between $Edge_1$ of C_1 and $Edge_2$ of C_2 is the number of violations

incurred to this constraint.

Morimoto (2001: 165) in addition proposes the GF Abutment:

- (6) For $F \in \text{F-Cat}$, $C_1, C_2 \in \text{C-Cat}$; $\text{Edge}_1 \neq \text{Edge}_2$
 $\text{Abut}(F, \text{Edge}_1, C_1, \text{Edge}_2) =_{\text{def}}$
 F abuts with C_1 if and only if (i) $\exists C_2 \in \phi^{-1}(F)$, and (ii) Edge_1 of C_2 shares Edge_2 of C_1 (Abbreviated as $\text{Abut-F}(C_1)$)

C_1 is a c-structure category and C_2 is the inverse image of a grammatical function F . F abuts with C_1 if and only if Edge_1 of F 's inverse image C_2 shares Edge_2 of C_1 . Morimoto (2001) illustrates the GF Abutment by introducing the following constraint:

- (7) $\text{Abut-OBJ}(V\text{-HD})$: OBJ abuts with V^0

This constraint requires the inverse image of the OBJ function to abut with the verbal head. Note that the inverse image of a GF refers to the whole c-structure phrase to which the GF is mapped. With respect to the information domain, Choi (2001: 35) suggests that ‘a domain aligns if and only if every member of the domain aligns’. This concept also applies to alignment and abutment constraints. Therefore, $\text{Abut-OBJ}(V\text{-HD})$ is satisfied if and only if the entire inverse image of the OBJ abuts with the verbal head. Only abutting part of the OBJ violates the constraint. Every element intervening between the verbal head and the inverse image of the OBJ or part of the OBJ incurs one violation to this constraint. Therefore, the following example violates this constraint twice by having *aa3* and *ngo5* between the V^0 , *maai56zo2*, and *zi6din2*, which is part of the OBJ, *saam1 bun2 zi6din2*.

- (8) 字典 啊 我 買咗 三 本
zi6din2 aa3 ngo5 maai5zo2 saam1 bun2
 dictionary PART 1.SG buy.PERF three CL
 ‘I have bought three DICTIONARIES.’

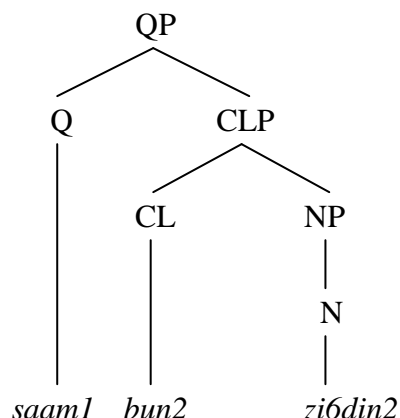
It should be noted that, though the input f-structure is only an underspecified f-structure without specifying to which grammatical functions the GFS refer (see chapter 1 for the structure of an OT-LFG input), constraints with reference to some specific GFS such as SUBJ and OBJ are still valid since it is assumed that the mapping between the a-structure and the f-structure has already been taken care of by some a-to f-structure correspondence constraints (see Asudeh 2001 and Bresnan 2000 for details of the constraints) before evaluating the candidates with respect to the constraints involving specific GFS.

The Abut-OBJ(V-HD) constraint has been applied to Cantonese (Lam 2004) and will also be adopted in the present study. Apart from this constraint, there are some other abutment constraints which are relevant to the current analysis. Based on the observations in section 4.2, the following two constraints are proposed.

- (9) Abut-NP(CL-HD): Abut Edge₁ of NP with Edge₂ of CL⁰
 (10) Abut-CLP(Q-HD): Abut Edge₁ of CLP with Edge₂ of Q⁰

The Abut-NP(CL-HD) constraint is introduced based on the observation that an NP can be topicalized leaving its CL⁰ *in situ*. Extracting a CLP (consisting of a CL⁰ and an NP), separating it from its Q⁰ is however impossible. This illegitimacy has to be accounted for by the Abut-CLP(Q-HD) constraint. The following c-structure of the phrase *saam1 bun2 zi6din2* (three-CL-dictionary meaning *three dictionaries*) illustrates how these constraints work:

(11) C-structure of *saam1 bun2 zi6din2*



This phrase satisfies Abut-CLP(Q-HD) since the left edge of the CLP, *bun2 zi6din2*, abuts with the right edge of the Q^0 , *saam1*. This constraint is violated by, for example, **saam1 ngo5 bun2 zi6din2* (three-1.SG-CL-dictionary), where the CLP, *bun2 zi6din2*, and the Q^0 , *saam1*, is separated by the D^0 , *ngo5*. The phrase in (11) also satisfies Abut-NP(CL-HD) with the left edge of the NP *zi6din2* abutting with the right edge of the CL^0 , *bun2*. **Bun2 saam1 zi6din2* (CL-three-dictionary), for instance, incurs one violation to this constraint by separating the NP *zi6din2* and the CL^0 *bun2* by the Q^0 , *saam1*.

In the above discussion on alignment and abutment constraints, it is stated that the number of violations incurred is the number of *elements* intervening between edges. The word *element* is interpreted as a categorial unit in this study. Example (8A) in chapter 2, repeated here as (12), illustrates what an element refers to:

- (12) 三 本 字典 啊 我 買咗
saam1 bun2 zi6din2 aa3 ngo5 maai5zo2
 three CL dictionary PART 1.SG buy.PERF
 ‘It is three dictionaries that I have bought.’

In this example, the sentence-initial QP which consists of three elements including a Q^0 (*saam1*), a CL^0 (*bun2*) and an N^0 (*zi6din2*), and the particle (C^0) *aa3* standing in front of the SUBJ *ngo5* incur in total four violations to the SUBJ-L constraint.

4.3.3 [+New]-FOCUS \wedge FOCUS-L

The [+New]-FOCUS \wedge FOCUS-L constraint is derived particularly for FOCUS topicalization. The two ‘FOCUS’ in the constraint refer to a topicalized discourse focus. This constraint requires that there be a FOCUS in the sentence whenever there exists some [+New] information *and* the FOCUS should occur in the leftmost position of the sentence. One violation is incurred when there is some [+New] information but no FOCUS in the sentence *or* when the FOCUS does not align left. This constraint is introduced based on Choi’s (2001) [+New] feature and NEW-L constraint. In the following, it will be explained why it is derived, instead of employing NEW-L directly.

Choi (2001: 34) proposes the I-s/C-s Alignment Constraints which deal with the correspondence between i-structure and c-structure (see chapter 2 for the definition of ‘Prom’ and ‘New’):

- (13) PROM-L: [+Prom] aligns left in the clause
- (14) PROM-R: [+Prom] aligns right in the clause
- (15) NEW-L: [+New] aligns left in the clause
- (16) NEW-R: [+New] aligns right in the clause

Every [-New] element occurring on the left of [+New] incurs one violation to NEW-L and every [-New] element occurring on the right of the [+New] incurs one violation to NEW-R. The same rule applies to PROM-L and PROM-R.

Choi (1999) argues that topicalization is employed in English to encode information prominence instead of newness. PROM-L is then the relevant constraint for characterizing English topicalization. It has been shown in chapter 2 that FOCUS in Cantonese can encode discourse focus bearing either [+Prom] or [-Prom] feature and it is the [+New] feature rather than the [+Prom] feature which licenses FOCUS topicalization. Choi's (2001) NEW-L, however, cannot be directly applied to Cantonese. Recall that, although the [+New] feature licenses FOCUS topicalization, a FOCUS may also at the same time contain some [-New] information and it is not obligatory for the [+New] element to appear in the leftmost position of the sentence. Neither is it necessary that all [+New] information appear inside the FOCUS (see chapter 2 for details). Based on these observations, the conjunctive constraint [+New]-FOCUS \wedge FOCUS-L is introduced, which states that there must be a FOCUS when there exists some [+New] information in the sentence *and* the FOCUS should align left. This constraint conjunction does not require that the [+New] information should align left or that the FOCUS should contain only [+New] information. Neither does it demand that all [+New] information should appear inside the FOCUS. One violation is incurred when either or both of the constraints are failed in accordance with the principle of constraint conjunction (Crowhurst and Hewitt 1997:7):

- (17) Constraint Conjunction: A candidate *Cand* passes a conjunction $A \wedge B$ iff *Cand* passes constraint *A* *and* *Cand* passes constraint *B*.

When [+New]-FOCUS is violated (i.e., there is some [+New] information but no FOCUS), FOCUS-L becomes irrelevant since there is no FOCUS to align. In other words, there is no instance where both constraints are violated. Therefore, one violation is incurred to [+New]-FOCUS \wedge FOCUS-L when either of the constraints is violated.

4.3.4 *TOPICALIZE-FC

*TOPICALIZE-FC is introduced based on the observation that sentence (18) is well-formed while sentence (19) is ill-formed:

(18) 字典 啊 我 買咗 三 本
zi6din2 aa3 ngo5 maai5zo2 saam1 bun2
 dictionary PART 1.SG buy.PERF three CL
 ‘I have bought three DICTIONARIES.’

(19) *三 本 啊 我 買咗 字典
**saam1 bun2 aa3 ngo5 maai5zo2 zi6din2*
 three CL PART 1.SG buy.PERF dictionary

Broadwell (1999) proposes the *NP extraction constraint to account for the fact concerning interrogative constructions in English and San Dionicio Ocotepc Zapotec that ‘it is ungrammatical to attempt to extract either a determiner or a possessive from the NP without pied-piping the NP’ (1999: 8), which says ‘[t]he specifier of NP must be a sister of N’ (1999: 8). While the phenomenon which Broadwell is trying to account for is similar to the one in Cantonese discussed here, the *NP extraction constraint is not applicable since CL^0 can in fact be separated from the NP as in (18).

While the topicalized phrase in (18) is an NP, the topicalized elements Q^0 and CL^0 in (19) do not form a maximal projection. This however does not necessarily lead to ungrammaticality. Consider example (43a) in chapter 3 again, repeated here as (20).

(20) 買咗 啦 我 嗰 本 書
maai5zo2 laa1 ngo5 go2 bun2 syu1
 buy.PERF PART 1.SG DEF CL book
 ‘I HAVE already BOUGHT that book.’

In this example, the FOCUS *maai5zo2* is a V, which is not a maximal projection either. The acceptability of this construction shows that a topicalized element is not necessarily a maximal projection.

Another difference between constructions (18) and (19) is that the topicalized phrase in (18) is an NP which takes no complement. In (19), it is a Q^0 and a CL^0 which are topicalized, with the NP complement of the CL^0 , *zi6din2*, being left *in situ*. Leaving the complement *in situ*, again, cannot explain its unacceptability. In (20), the V *maai5zo2* is topicalized to become the FOCUS, also leaving its DP complement *go2 bun2 syu1* in its canonical OBJ position. Therefore, a complement can in fact be left *in situ* in FOCUS topicalization in Cantonese.

Though the complement of the topicalized CL^0 and V^0 in sentences (19) and (20) respectively are both left *in situ*, the two constructions differ in the syntactic category of the topicalized element. In (20), the stranded CLP is the complement of the topicalized V^0 , which is a lexical category, while the stranded NP in (19) is the complement of the topicalized CL^0 , which is a functional category. Based on this observation, the *TOPICALIZE-FC constraint is proposed, which disallows topicalizing only a functional category with the complement being left *in situ*. One violation is incurred to this constraint for every topicalized functional category of which the complement stays in its canonical position. Therefore, sentence (19) incurs one violation to this constraint by topicalizing the CL^0 *bun2* without its NP complement *zi6din2*.

It should however be noted that since Cantonese is a pro-drop language, the NP *zi6din2* can be dropped provided that it can be recovered from the context. If this NP in (19) is dropped, the sentence becomes grammatical:

- (21) 三 本 啊 我 買咗
saam1 bun2 aa3 ngo5 maai5zo2
 three CL PART 1.SG buy.PERF
 ‘It is three (dictionaries) that I have bought.’

To conclude, what *TOPICALIZE-FC prohibits is that the functional category is topicalized with the complement being left *in situ*. The functional category is free to be topicalized when the complement which can be recovered from the context is dropped.

In the above sections, the constraints relevant to topicalization in Cantonese have been discussed. These constraints will be ranked against each other based on a more detailed analysis of the data in the next section.

4.4 The Constraint Rankings

In this section, the constraint rankings for TOPIC and FOCUS topicalization in Cantonese will be proposed. Sections 4.4.1 and 4.4.2 establish the subhierarchies for TOPIC and FOCUS topicalization respectively. Section 4.4.3 revises the subhierarchies through examining the evaluation of several inputs.

4.4.1 The Subhierarchy for TOPIC Topicalization

The TOPIC topicalization constructions in Cantonese will be examined presently. The relevant constraints will be investigated and ranked against each other based on the results of the questionnaire survey.

SUBJ-L and DISTOPIC-L interact to decide whether it is the SUBJ or the phrase designating the discourse topic which comes first. When the SUBJ of a clause is at the same time the discourse topic, the canonical SVO order satisfies both SUBJ-L and DISTOPIC-L. This is however not always the case. Consider the following example:

- (22) Q: 本 字典 呢?
bun2 zi6din2 ne1?
 CL dictionary PART
 ‘Where’s the dictionary?’
- A1: 本 字典 啊 我 掉咗
bun2 zi6din2 aa4 ngo5 deu6zo2
 CL dictionary PART 1.SG throw away.PERF
 ‘As for the dictionary, I have thrown it away.’
- A2: 我 掉咗 本 字典
ngo5 deu6zo2 bun2 zi6din2
 1.SG throw away.PERF CL dictionary
 ‘I have thrown away the dictionary.’

The question establishes *bun2 zi6din2* as the discourse topic. In (22A1), *bun2 zi6din2*, which is at the same time the OBJ of the clause, is topicalized to become the TOPIC. This construction satisfies DISTOPIC-L by aligning the discourse topic on the left periphery. It however violates SUBJ-L thrice by having three elements, *bun2*, *zi6din2* and *aa4*, to the left of the SUBJ *ngo5* and violates Abut-OBJ(V-HD) twice by separating the OBJ *bun2 zi6din2* and the V⁰ *deu6zo2* with two elements, *aa4* and *ngo5*. Keeping the phrase designating the discourse topic in the OBJ position as in (22A2), which maintains the SUBJ alignment and the abutment between the OBJ and the V⁰ on the one hand and forsakes the left alignment of the discourse topic on the other hand, is also possible. Since there are two possible outputs, the notion of free ranking becomes relevant (Kager 1999: 406):

- (23) Interpretation of free ranking of constraints C₁, C₂: Evaluation of the candidate set is split into two subhierarchies, each of which selects an optimal output. One subhierarchy has C₁ >> C₂, and the other C₂ >> C₁.

The topicalization construction wins over the canonical structure when DISTOPIC-L outranks both SUBJ-L and Abut-OBJ(V-HD). For the canonical construction to win, DISTOPIC-L should be outranked by either SUBJ-L or Abut-OBJ(N-HD), or both. The following shows the possible subhierarchies:

(24) Topicalization construction (22A1) wins:

- a. DISTOPIC-L >> SUBJ-L, Abut-OBJ(V-HD)

Canonical construction (22A2) wins:

- b. SUBJ-L >> DISTOPIC-L >> Abut-OBJ(V-HD)
 c. Abut-OBJ(V-HD) >> DISTOPIC-L >> SUBJ-L
 d. SUBJ-L, Abut-OBJ(V-HD) >> DISTOPIC-L

Whether it is (24b), (24c) or (24d) which is responsible for selecting the canonical structure requires more data and analyses to confirm. The present study concentrates on the constraint ranking for topicalization and the subhierarchy for selecting the canonical construction will not be examined here. Therefore, only subhierarchy (24a) will be considered in the succeeding discussion.

Now consider the following example:

- (25) Q: 你 買咗 幾多 本 字典 啊?
nei5 maai5zo2 gei2do1 bun2 zi6din2 aa3
 2.SG buy.PERF how many CL dictionary PART
 ‘How many dictionaries have you bought?’
- A: 字典 啊 我 買咗 三 本
zi6din2 aa4 ngo5 maai5zo2 saam1 bun2
 dictionary PART 1.SG buy.PERF three CL
 ‘As for dictionaries, I have bought three.’

Assume that the NP *zi6din2* is the discourse topic (in fact, the discourse topic can also be the SUBJ *nei5*, depending on the larger context). This NP is topicalized in (25A) in order to satisfy DISTOPIC-L. Two violations are incurred to SUBJ-L with *zi6din2* and *aa4* appearing in front of the SUBJ *ngo5*, two to Abut-OBJ(V-HD) with *aa4* and *ngo5* standing between part of the OBJ *zi6din2* and the V⁰ *maai5zo2*, and four to Abut-NP(CL-HD) with *aa4*, *ngo5*, *maai5zo2* and *saam1* standing between the NP *zi6din2* and the CL⁰ *bun2*. For this topicalization construction to be selected as the optimal output, DISTOPIC-L should dominate SUBJ-L, Abut-OBJ(V-HD) and Abut-NP(CL-HD):

(26) DISTOPIC-L >> SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD)

The Abut-CLP(Q-HD) constraint cannot be ranked against DISTOPIC-L since there is no instance where the Q⁰ is separated from the CLP in order to satisfy DISTOPIC-L. If *saam1 bun2 zi6din2* (*three dictionaries*), which is indefinite, is the OBJ of a clause, there is no point in making the CLP *bun2 zi6din2*, which gives a definite interpretation, the TOPIC, leaving the Q⁰ *saam1* behind, since the definiteness of the OBJ and the TOPIC does not converge and the uniqueness condition is violated (see chapter 1 for the definition of the condition). If it is to be conveyed that the speaker has bought three copies of a *particular* dictionary, the following construction can be used:

(27) 本 字典 啊 我 買咗 三 本
bun2 zi6din2 aa4 ngo5 maai5zo2 saam1 bun2
 CL dictionary PART 1.SG buy.PERF three CL
 ‘As for the dictionary, I have bought three copies of it.’

The sentence-initial CLP *bun2 zi6din21* refers to a particular dictionary which is supposed to be known to the participants of the conversation. *Saam1 bun2* at the end of the clause conveys the meaning of *three copies*. It can however be noticed that the Q^0 still abuts with its CLP *bun2 zi6din2*.

Another situation where the Q^0 does not abut with its CLP is when the Q^0 is topicalized, leaving the CLP *in situ*. There is however no instance where only the Q^0 (e.g. *saam1*) is the discourse topic, but not its CLP complement (e.g. *bun2 zi6din2*). To conclude, in no case is Abut-CLP(Q-HD) violated in order to satisfy DISTOPIC-L or vice versa and the two constraints thus cannot be ranked against each other. The ranking of the Abut-CLP(Q-HD) constraint will be left to the discussion of FOCUS topicalization in the next section.

To summarize this section, the following subhierarchy is proposed based on the observations on TOPIC topicalization in Cantonese:

- (28) DISTOPIC-L >> SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD)

4.4.2 The Subhierarchy for FOCUS Topicalization

This section looks at the FOCUS topicalization constructions in Cantonese. All the constraints relevant to the phenomenon will be examined in an attempt to establish a constraint ranking.

Consider question-answer pair (9) in chapter 2 again, repeated here as (29).

- (29) Q: 你 買咗 三 本 咩 啊?
nei5 maai5zo2 saam1 bun2 me1 aa3?
 2.SG buy.PERF three CL what PART
 ‘What are the three books that you have bought?’

A: [三 本 字典] 啊 我 買咗
 [saam1 bun2 zi6din2] aa3 ngo5 maai5zo2
 three CL dictionary PART 1.SG buy.PERF
 ‘It is three dictionaries that I have bought.’

Given the question, *zi6din2* in the answer becomes the only phrase carrying the [+New] information. The answer is a topicalization construction where the FOCUS *saam1 bun2 zi6din2* containing the [+New] phrase *zi6din2* occurs in the leftmost position of the sentence. Four violations are incurred to SUBJ-L with four elements, *saam1*, *bun2*, *zi6din2* and *aa3*, appearing to the left of the SUBJ *ngo5*. Two violations are incurred to Abut-OBJ(V-HD) with *aa3* and *ngo5* intervening between the OBJ *saam1 bun2 zi6din2* and the V⁰ *maai5zo2*.

The canonical structure with no FOCUS is also an acceptable answer to (29Q):

(30) 我 買咗 三 本 字典 啊
 ngo5 maai5zo2 saam1 bun2 zi6din2 aa3
 1.SG buy.PERF three CL dictionary PART
 ‘I have bought three dictionaries.’

The [+New]-FOCUS \wedge FOCUS-L constraint is violated once here as there is no FOCUS. Similar to the case of TOPIC topicalization discussed above, there is more than one subhierarchy which is responsible for the two possible outputs. Again, this thesis will only concentrate on the subhierarchy in favour of the topicalization structure. [+New]-FOCUS \wedge FOCUS-L should outrank both SUBJ-L and Abut-OBJ(V-HD) in order to have the FOCUS topicalization construction as the optimal output:

(31) [+New]-FOCUS \wedge FOCUS-L \gg SUBJ-L, Abut-OBJ(V-HD)

Sentence (32) is yet another possible answer to question (29Q).

- (32) 字典 啊 我 買咗 三 本
zi6din2 aa3 ngo5 maai5zo2 saam1 bun2
 dictionary PART 1.SG buy.PERF three CL
 ‘I have bought three DICTIONARIES.’

In this construction, only part of the OBJ is identified with the FOCUS. It again violates Abut-OBJ(V-HD) twice by having part of the OBJ *zi6din2* separated from the V^0 *maai5zo2* by *aa3* and *ngo5*. Besides, it incurs four violations to Abut-NP(CL-HD) as the NP *zi6din2* is separated from the CL^0 *bun2* by *aa3*, *ngo5*, *maai5zo2* and *saam1*. For this construction to be the optimal output, [+New]-FOCUS \wedge FOCUS-L has to dominate Abut-NP(CL-HD). The following subhierarchy is concluded:

- (33) [+New]-FOCUS \wedge FOCUS-L \gg SUBJ-L, Abut-OBJ(V-HD),
 Abut-NP(CL-HD)

There are two FOCUS topicalization constructions which are acceptable as an answer to question (29Q). It will be demonstrated in section 4.4.3 whether this subhierarchy suffices to select two optimal outputs.

Not all FOCUS topicalization constructions with the [+New] information *zi6din2* appearing in the FOCUS are legitimate. Sentence (34) is an ill-formed example.

- (34) *本 字典 啊 我 買咗 三
**bun2 zi6din2 aa3 ngo5 maai5zo2 saam1*
 CL dictionary PART 1.SG buy.PERF three

[+New]-FOCUS \wedge FOCUS-L is satisfied by the presence of the FOCUS *bun2 zi6din2*.

Though Abut-OBJ(V-HD) is violated as *bun2 zi6din2*, being part of the OBJ, is separated from the V^0 *maai5zo2*, it is not the reason for the unacceptability since it is shown in (29A) and (32) above that this constraint can be violated in order to satisfy [+New]-FOCUS \wedge FOCUS-L and the number of violations incurred by sentences (29A), (32) and (34) to this constraint are the same as the elements intervening between (part of) the OBJ and the V^0 are the same (i.e., *aa3* and *ngo5*). SUBJ-L cannot explain the ungrammaticality either since the well-formed sentence (29A) above incurs even one more violation to this constraint than sentence (34) by having one more element, *saam1*, in front of the SUBJ. The other constraint which is violated by construction (34) is Abut-CLP(Q-HD) as the CLP *bun2 zi6din2* and the Q^0 *saam1* are separated by *aa3*, *ngo5* and *maai5zo2*. It is therefore suggested that Abut-CLP(Q-HD) is the reason for the construction's unacceptability and should be ranked higher than [+New]-FOCUS \wedge FOCUS-L. The subhierarchy is revised as follows:

- (35) Abut-CLP(Q-HD) \gg [+New]-FOCUS \wedge FOCUS-L \gg SUBJ-L,
Abut-OBJ(V-HD), Abut-NP(CL-HD)

Finally, the *TOPICALIZE-FC constraint will be examined. Consider the following two question-answer pairs.

- (36) Q: 你 買咗 三 本 咩 啊?
nei5 maai5zo2 saam1 bun2 me1 aa3?
2.SG buy.PERF three CL what PART
'What are the three books that you have bought?'
- A: 字典 啊 我 買咗 三 本
zi6din2 aa3 ngo5 maai5zo2 saam1 bun2
dictionary PART 1.SG buy.PERF three CL
'I have bought three DICTIONARIES.'

- (37) Q: 你 買咗 幾多 本 字典 啊?
nei5 maai5zo2 gei2do1 bun2 zi6din2 aa3
 2.SG buy.PERF how many CL dictionary PART
 ‘How many dictionaries have you bought?’
- A: *三 本 啊 我 買咗 字典
**saam1 bun2 aa3 ngo5 maai5zo2 zi6din2*
 three CL PART 1.SG buy.PERF dictionary

In both answers, the [+New] information is topicalized in an attempt to satisfy [+New]-FOCUS \wedge FOCUS-L. Both constructions incur the same number of violations to Abut-OBJ(V-HD) by having two elements, *aa3* and *ngo5*, intervening between part of the OBJ (*zi6din2* and *saam1 bun2* respectively) and the V^0 *maai5zo2*. Construction (36A) incurs four violations to Abut-NP(CL-HD) by having *aa3*, *ngo5*, *maai5zo2* and *saam1* intervening between the NP *zi6din2* and the CL^0 *bun2* and two violations to SUBJ-L by having *zi6din2* and *aa3* in front of the SUBJ *ngo5*. Sentence (37A) violates Abut-NP(CL-HD) thrice with *aa3*, *ngo5* and *maai5zo2* appearing between the NP and the CL^0 and violates SUBJ-L thrice with *saam1*, *bun2* and *aa3* standing to the left of the SUBJ. The critical difference between the two sentences is that, as mentioned in section 4.3.4, in (36A), it is the lexical projection NP which takes no complement that is topicalized, while in (37A), it is the functional category CL^0 without its complement which is topicalized. Therefore, construction (37A) violates *TOPICALIZE-FC by topicalizing a functional category with its complement being left *in situ*. Based on the ungrammaticality of this construction, it is proposed that *TOPICALIZE-FC dominates [+New]-FOCUS \wedge FOCUS-L. The subhierarchy in (38) is concluded for FOCUS topicalization in Cantonese:

- (38) Abut-CLP(Q-HD), *TOPICALIZE-FC >> [+New]-FOCUS \wedge FOCUS-L >>
 SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD)

In the above two sections, a subhierarchy has been proposed for TOPIC topicalization and FOCUS topicalization in Cantonese respectively. The next section demonstrates how candidates are evaluated by these subhierarchies.

4.4.3 Revising the Subhierarchies

After proposing two subhierarchies for Cantonese, it will be demonstrated in this section how the subhierarchies select the optimal outputs. Various inputs will be evaluated and the subhierarchies will be revised whenever necessary.

King (1997) mentions that every constituent is assigned a discourse label in the i-structure. It is therefore normal to find discourse topic, [+New] and [-New] information coexisting in an i-structure. Though it is indeed possible for some constructions to have TOPIC as well as FOCUS (see example (2) in chapter 2 and example (5) in chapter 3) considering the DISTOPIC-L and [+New]-FOCUS \wedge FOCUS-L constraints, this kind of construction is not commonly found in the language. Since a language is assumed to have one single hierarchy (subhierarchies are possible where options exist), it is reasonable and desirable to combine the above-established subhierarchies for TOPIC and FOCUS constructions in Cantonese. Due to space limitations, however, this will be left for future research (see chapter 5). The inputs given in the following will therefore be independently evaluated against the subhierarchy for TOPIC topicalization and the one for FOCUS topicalization and constructions with multiple topicalized phrases will not be included in the candidate set.

The constraint ranking for TOPIC constructions will be examined first. As

mentioned in chapter 1, the present study assumes an underspecified f-structure and an i-structure for the input. Consider the following input:

(39) a. Input - F-structure

PRED	<i>'deu6 <x, y>'</i>		
ASPECT	PERF		
GF ₁	PRED	<i>'Pro'</i>	x
	NUM	SG	
	PERS	1	
GF ₂	PRED	<i>'zi6din2'</i>	y
	CL	<i>bun2</i>	
	DEF	+	

b. Input - I-structure

Discourse topic	{ <i>bun2 zi6din2</i> }
+New	{ <i>ngo5</i> <i>deu6zo2</i> }

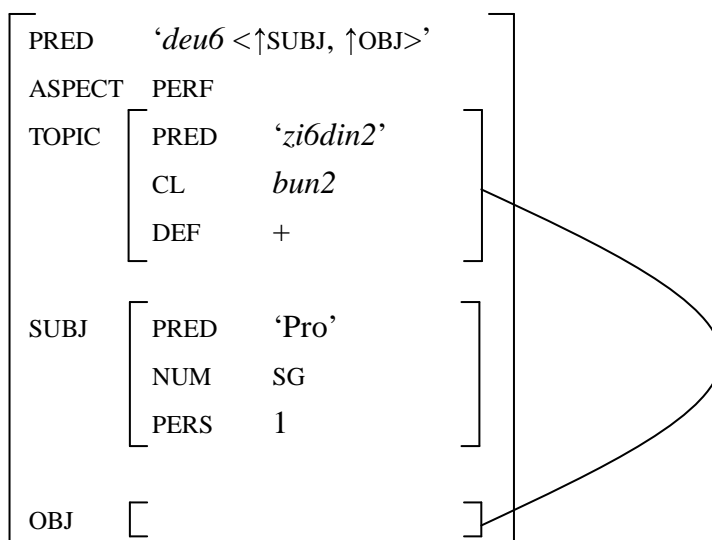
The tableau in (40) shows how the optimal output is selected with subhierarchy (28) established for TOPIC topicalization.

(40) Tableau of input (39) with subhierarchy (28)

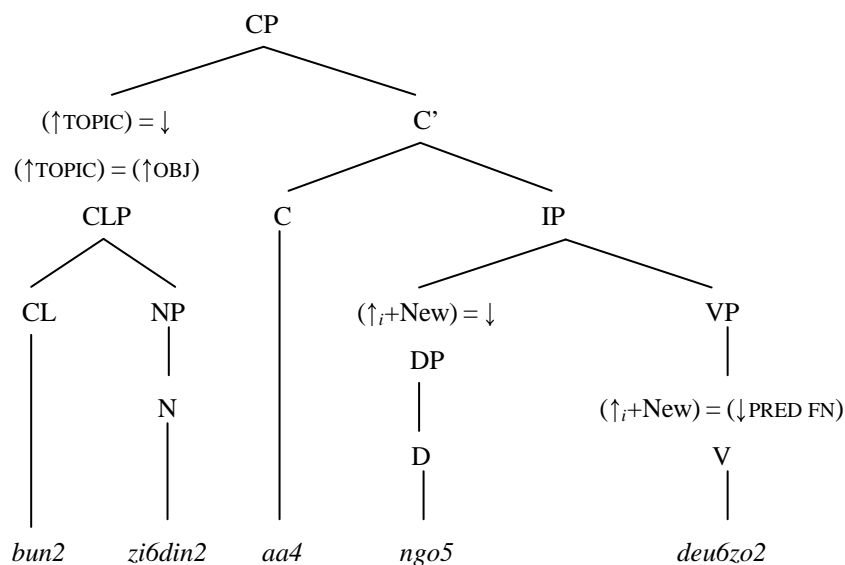
	DisTOPIC-L	SUBJ-L	Abut-OBJ(V-HD)	Abut-NP(CL-HD)
a. <i>ngo5 deu6zo2 bun2 zi6din2</i> 1.SG throw away.PERF CL dictionary	*!*			
b. <i>bun2 zi6din2 aa4 ngo5 deu65zo2</i> CL dictionary PART 1.SG throw away.PERF		** *	**	

Since the constraint ranking considered here is for selecting the topicalization construction, candidate (40a) with the canonical structure is ruled out by violating the highest-ranked DISTOPIC-L even though it is well-formed. Candidate (40b) becomes the optimal output since the highest-ranked constraints it violates (SUBJ-L and Abut-OBJ(V-HD)) are lower-ranked than the one violated by candidate (40a) (DISTOPIC-L). The optimal output has the following fully-specified f-structure and annotated c-structure (the i-structure of the output is the same as that of the input):

(41) a. F-structure of (40b)



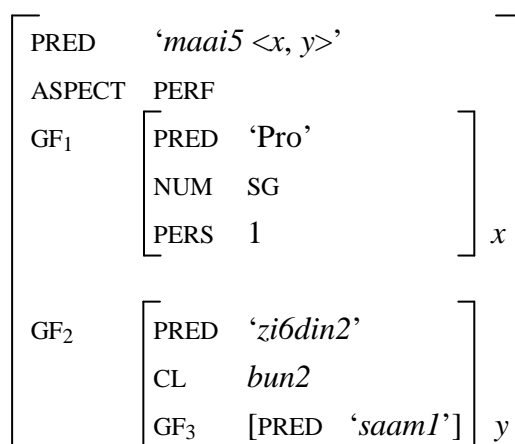
b. Annotated c-structure of (40b)



In f-structure (41a), GF₁ and GF₂ are replaced by ‘SUBJ’ and ‘OBJ’ respectively and the variables *x* and *y* by ‘↑SUBJ’ and ‘↑OBJ’. An additional function TOPIC is added as the phrase designating the discourse topic is topicalized to satisfy DISTOPIC-L. In the annotated c-structure (41b), the ‘PRED FN’ notation above the V is necessary to exclude its arguments from the [+New] domain (see chapter 3 for details).

Now consider another input:

(42) a. Input - F-structure



b. Input - I-structure

Discourse topic	{ngo5}
+New	{saam1}
- New	{ maai5zo2 bun2 zi6din2}

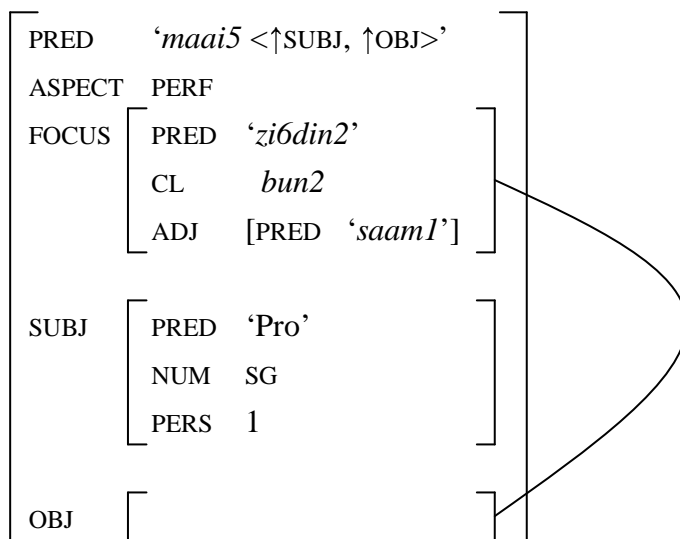
Since the discourse topic is at the same time the SUBJ *ngo5* (note that the mapping of *ngo5* to SUBJ is determined by the a- to f- structure correspondence constraints (Asudeh 2001; Bresnan 2000) as mentioned in section 4.3.2), it is obvious that the canonical SVO structure, which satisfies both DISTOPIC-L and SUBJ-L is the optimal output with respect to the subhierarchy for TOPIC topicalization. The following discussion will then concentrate on the evaluation against the subhierarchy for FOCUS topicalization. The tableau below shows the evaluation:

(43) Tableau of input (42) with subhierarchy (38)

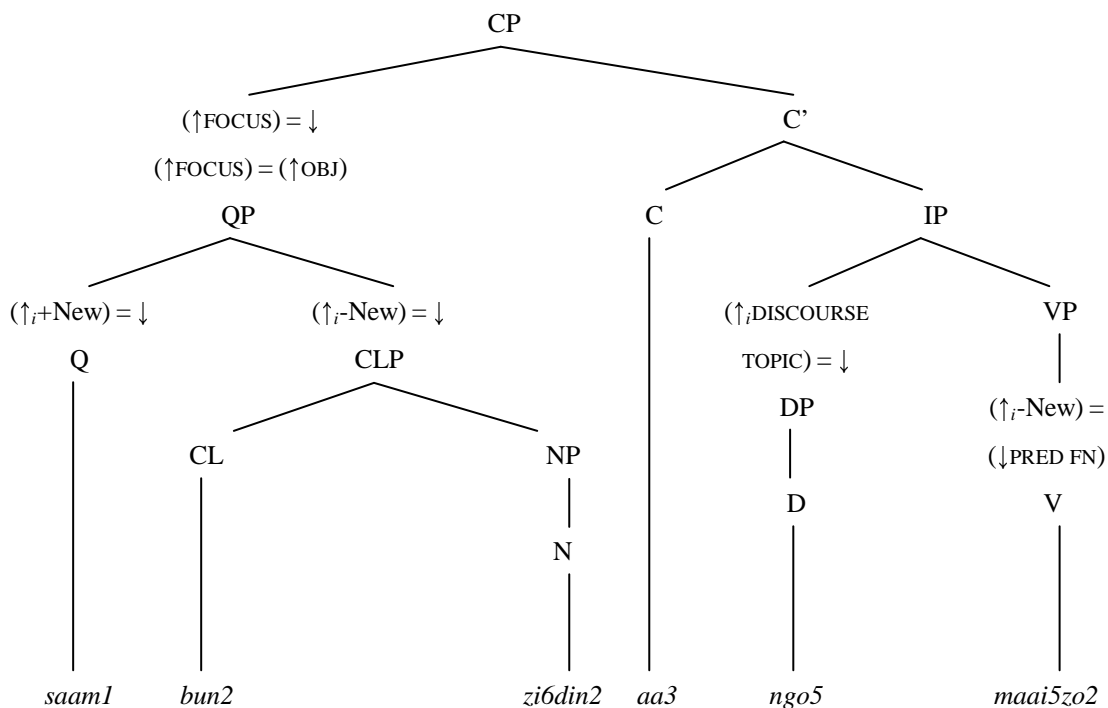
	Abut-CLP(Q-HD)	*TOPICALIZE-FC	[+New]-FOCUS ^ FOCUS-L	SUBJ-L	Abut-OBJ(V-HD)	Abut-NP(CL-HD)
a. <i>ngo5 maai5zo2 saam1 bun2 zi6din2</i> 1.SG buy.PERF three CL dictionary			*!			
b. <i>saam1 aa3 ngo5 maai5zo2 bun2 zi6din2</i> three PART 1.SG buy.PERF CL dictionary	*!* *			**	**	
c. <i>saam1 bun2 aa3 ngo5 maai5zo2 zi6din2</i> three CL PART 1.SG buy.PERF dictionary		*!		**	**	**
d. ☞ <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF				**	**	

Candidate (43a) violates [+New]-FOCUS \wedge FOCUS-L by containing no FOCUS. Candidate (43b) violates Abut-CLP(Q-HD) by separating the CLP *bun2 zi6din2* and the Q⁰ *saam1*. Candidate (43c), by topicalizing the CL⁰ with the Q⁰, satisfies Abut-CLP(Q-HD). It however violates *TOPICALIZE-FC as the NP complement is not topicalized together with the functional category CL⁰. Candidate (43d) is selected as the optimal output by incurring violations only to the lower-ranked constraints. The fully specified f-structure and the annotated c-structure of this output are shown below:

(44) a. F-structure of (43d)



b. Annotated c-structure of (43d)



Now, consider the f-structure input in (42a) with another i-structure input:

(45) Input - I-structure

DISCOURSE TOPIC	{ngo5}
+New	{zi6din2}
- New	{maai5zo2 saam1 bun2}

With the discourse topic being the SUBJ *ngo5*, the SVO structure wins again with respect to the TOPIC subhierarchy. The following tableau shows the evaluation of this input against the subhierarchy for FOCUS constructions:

(46) Tableau of input (45) with subhierarchy (38)

	Abut-CLP(Q-HD)	*TOPICALIZE-FC	[+New]-FOCUS \wedge FOCUS-L	SUBJ-L	Abut-OBJ(V-HD)	Abut-NP(CL-HD)
a. <i>ngo5 maai5zo2 saam1 bun2 zi6din2</i> 1.SG buy.PERF three CL dictionary			*!			
b. <i>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2</i> dictionary PART 1.SG buy.PERF three CL				**	**	** *!*
c. <i>bun2 zi6din2 aa3 ngo5 maai5zo2 saam1</i> CL dictionary PART 1.SG buy.PERF three	*!			** *	**	
d. <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF				** **	**	

The canonical construction (46a) again violates [+New]-FOCUS \wedge FOCUS-L by having no FOCUS. Candidate (46c) violates Abut-CLP(Q-HD) by topicalizing the CLP *bun2 zi6din2*, separating it from the Q⁰ *saam1*. Since SUBJ-L, Abut-OBJ(V-HD) and Abut-NP(CL-HD) are on the same stratum of the subhierarchy, candidate (46d) wins over candidate (46b) by having fewer cumulative violations to the three constraints.

According to the results of the questionnaire survey, both candidates (46b) and (46d) are acceptable given that the [+New] information falls on the NP *zi6din2*. The subhierarchy has to be revised since it only selects (46d) as the optimal output. The number of violations to SUBJ-L and Abut-NP(CL-HD) differ for the two candidates. With SUBJ-L outranking Abut-NP(CL-HD), candidate (46b) becomes optimal. With Abut-NP(CL-HD) outranking SUBJ-L, candidate (46d) is optimal. The tableaux below show how the two optimal outputs are selected with different subhierarchies:

(47) Revised tableau of input (45) with SUBJ-L outranking Abut-NP(CL-HD)

	Abut-CLP(Q-HD)	*TOPICALIZE-FC	[+New]-Focus ^ Focus-L	Abut-OBJ(V-HD)	SUBJ-L	Abut-NP(CL-HD)
a. <i>ngo5 maai5zo2 saam1 bun2 zi6din2</i> 1.SG buy.PERF three CL dictionary			*-			
b. <i>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2</i> dictionary PART 1.SG buy.PERF three CL				**	**	**
c. <i>bun2 zi6din2 aa3 ngo5 maai5zo2 saam1</i> CL dictionary PART 1.SG buy.PERF three	*.			**	**	
d. <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF				**	*.*	

(48) Revised tableau of input (45) with Abut-NP(CL-HD) outranking SUBJ-L

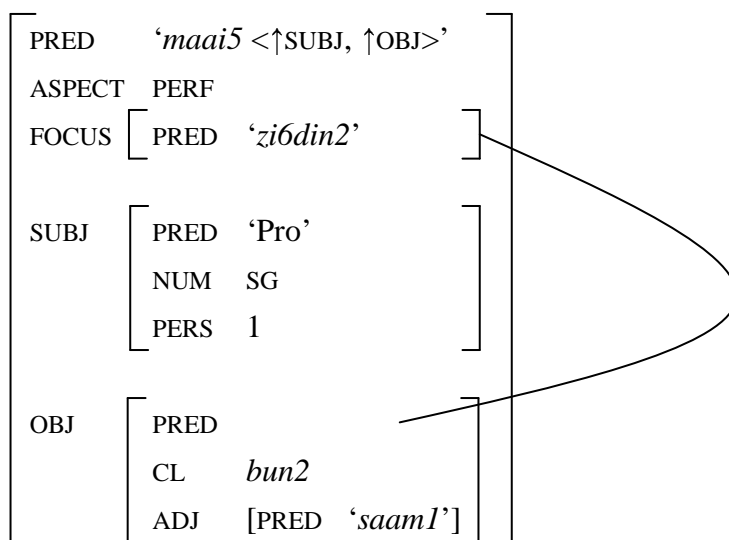
	Abut-CLP(Q-HD)	*TOPICALIZE-FC	[+New]-Focus ^ Focus-L	Abut-OBJ(V-HD)	Abut-NP(CL-HD)	SUBJ-L
a. <i>ngo5 maai5zo2 saam1 bun2 zi6din2</i> 1.SG buy.PERF three CL dictionary			*-			
b. <i>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2</i> dictionary PART 1.SG buy.PERF three CL				**	*.*	**
c. <i>bun2 zi6din2 aa3 ngo5 maai5zo2 saam1</i> CL dictionary PART 1.SG buy.PERF three	*.			**	**	**
d. <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF				**		**

It has been assumed that SUBJ-L, Abut-OBJ(V-HD) and Abut-NP(CL-HD) are on the

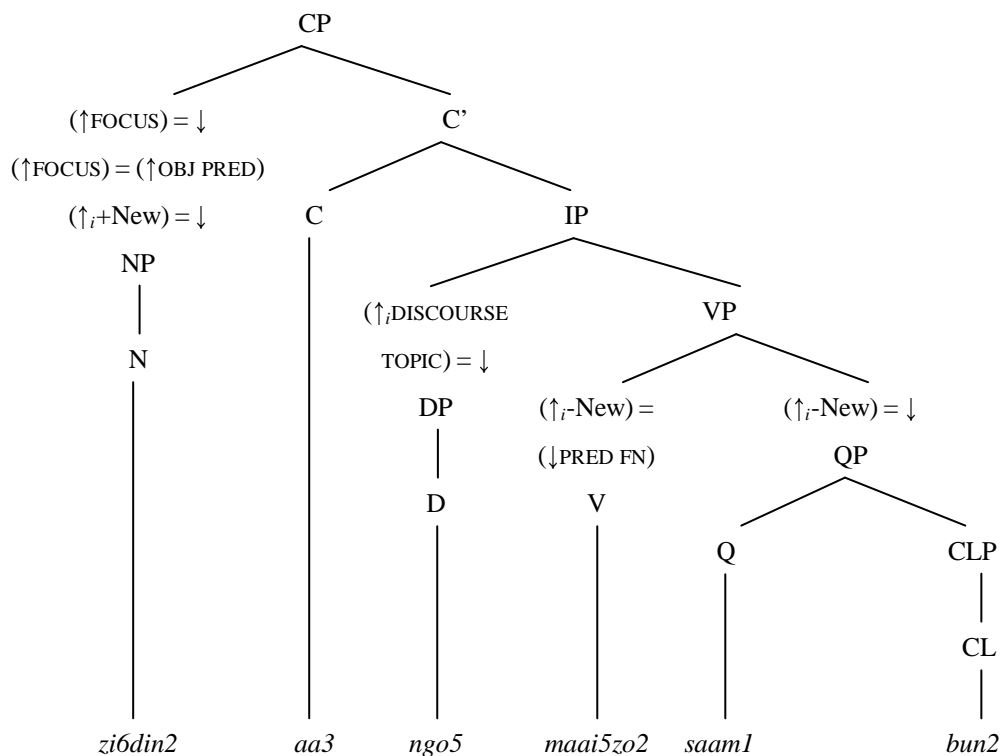
same stratum. According to Tesar and Smolensky (1998), all constraints are on the highest level of the hierarchy at the initial stage for constraint demotion. Therefore, when Abut-NP(CL-HD) in (47) or SUBJ-L in (48) is demoted, Abut-OBJ(V-HD) stays with the higher-ranked constraint for further demotion if necessary.

The fully-specified f-structure and annotated c-structure of output (47b) are shown below:

(49) a. F-structure of (47b)

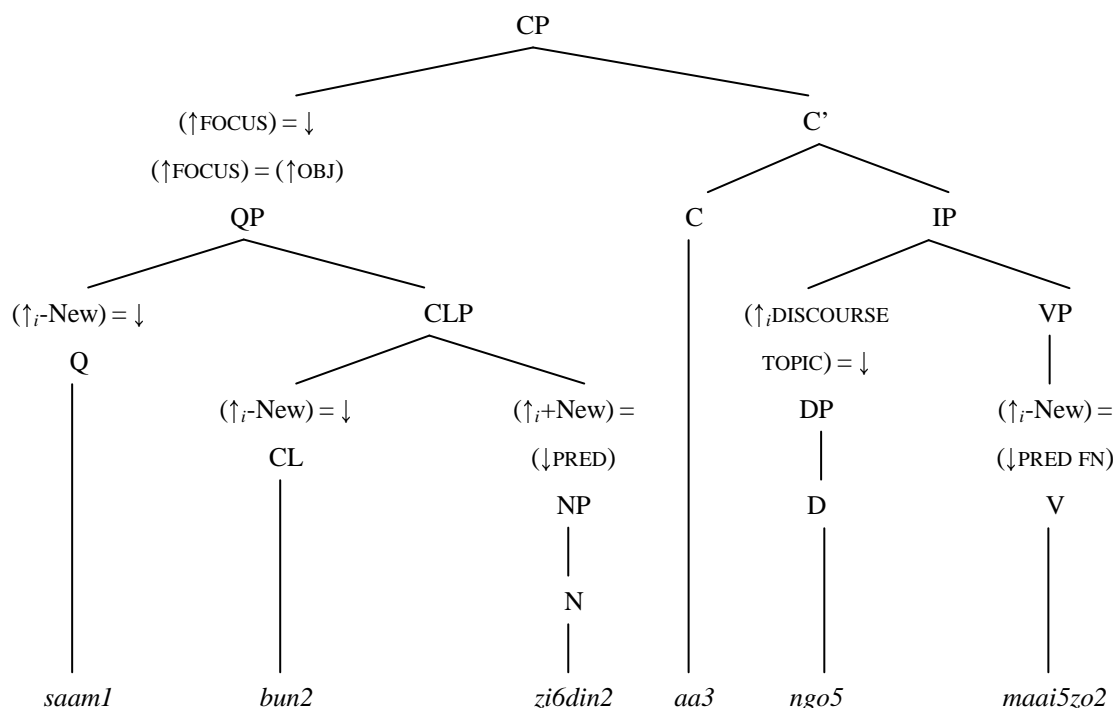


b. Annotated c-structure of (47b)



The annotated c-structure of output (48d) is as follows (see diagram (44a) for the f-structure):

(50) Annotated c-structure of (48d)



Through examining the evaluation of several inputs with respect to the subhierarchies for TOPIC and FOCUS topicalization in Cantonese, this section has demonstrated how the established rankings select the optimal outputs and revised the subhierarchy for FOCUS topicalization, resulting in one more subhierarchy. The next section summarizes the chapter.

4.5 Summary

In this chapter, an OT-LFG account has been proposed for OBJ topicalization in Cantonese. Some properties of the syntactic phenomenon have been presented, based on which the relevant constraints have been distinguished. The constraints have been ranked against each other and one subhierarchy and two subhierarchies have been proposed for TOPIC topicalization and FOCUS topicalization respectively as below:

- (51) The subhierarchy for TOPIC topicalization:
 DISTOPIC-L >> SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD)
- (52) The two subhierarchies for FOCUS topicalization:
- a. Abut-CLP(Q-HD), *TOPICALIZE-FC >> [+New]-FOCUS \wedge FOCUS-L >>
 Abut-OBJ(V-HD), SUBJ-L >> Abut-NP(CL-HD)
 - b. Abut-CLP(Q-HD), *TOPICALIZE-FC >> [+New]-FOCUS \wedge FOCUS-L >>
 Abut-OBJ(V-HD), Abut-NP(CL-HD) >> SUBJ-L

The next chapter concludes the study by summarizing the above chapters and suggesting areas for future research.

Chapter 5 Summary and Conclusions

In this chapter, a summary of the main findings is provided and some conclusions are drawn. Section 5.1 summarizes the study by addressing the research questions raised in chapter 1. Section 5.2 concludes the research.

5.1 Summary of the Research

The findings of the above chapters will be summarized in this section, addressing the three research questions outlined in chapter 1, repeated below:

- (i) How should TOPIC and FOCUS in Cantonese be defined?
- (ii) How should TOPIC and FOCUS topicalization in Cantonese be characterized with Functional Uncertainty in LFG? Does the functional approach suffice?
- (iii) What are the OT constraints relevant to Cantonese OBJ topicalization and how should they be ranked against each other?

5.1.1 Definition of FOCUS in Cantonese

While Chafe's (1976) definition of Chinese-style topic applies to TOPIC in Cantonese, there seems to be no definition in the literature which satisfactorily describes FOCUS in Cantonese. In order to define the concept, the information features of Cantonese FOCUS have been investigated. In chapter 2, it has been proved based on the results of the questionnaire survey that the [+New] feature is what licenses FOCUS topicalization in Cantonese. A FOCUS is not obliged to be purely

[+New]. [-New] information may co-exist with the [+New] information. While [+New] is the characteristic feature of a FOCUS, not all [+New] elements are required to appear inside the FOCUS. Given these observations, FOCUS in Cantonese has been defined as a topicalized phrase which contains the entire or part of the discourse focus (i.e., [+New] information).

5.1.2 An LFG Account of Topicalization

Chapter 3 provides an LFG account of topicalization in Cantonese, presenting a functional and an informational approach to the phenomenon. In the functional approach, two functional uncertainty equations have been established:

- (1) a. $(\uparrow\text{TOPIC}) = (\uparrow\{\text{SUBJ, OBJ, OBJ}_\theta, \text{XCOMP, OBL, ADJ, PRED}\})$
 b. $(\uparrow\text{FOCUS}) = (\uparrow\{\text{SUBJ, OBJ, OBJ}_\theta, \text{COMP, XCOMP, OBL, ADJ, PRED}\})$

According to these two equations, TOPIC and FOCUS, though being both topicalized functions, differ in the grammatical functions to which they can be linked.

With a purely functional account, Cantonese topicalization has been shown to experience scoping problems and display mismatches between f- and i-structure constituents. An informational approach has been employed to complement the functional approach, which postulates an independent i-structure (King 1997) responsible for representing discourse information so that, for instance, which element in the FOCUS is [+New] and which [-New] can be clearly captured. It has been proposed that the scoping problem of topicalizing only the PRED of a matrix f-structure or the PRED of a grammatical function should be resolved in the f-structure by connecting the topicalized phrase to the missing PRED only, excluding its arguments (for topicalizing the PRED of a matrix f-structure) or attributes (for

topicalizing the PRED of an OBJ).

5.1.3 Subhierarchies of Constraints for Cantonese

In chapter 4, an OT-LFG account of Cantonese OBJ topicalization has been presented, seeking to determine the constraint subhierarchies for TOPIC and FOCUS topicalization in the language. One subhierarchy and two subhierarchies are proposed for the two syntactic phenomena:

(2) The subhierarchy for TOPIC topicalization:

DIS_{TOPIC}-L >> SUBJ-L, A_{but}-OBJ(V-HD), A_{but}-NP(CL-HD)

(3) The subhierarchies for FOCUS topicalization:

- a. A_{but}-CLP(Q-HD), *TOPICALIZE-FC >> [+New]-FOCUS \wedge FOCUS-L >> A_{but}-OBJ(V-HD), SUBJ-L >> A_{but}-NP(CL-HD)
- b. A_{but}-CLP(Q-HD), *TOPICALIZE-FC >> [+New]-FOCUS \wedge FOCUS-L >> A_{but}-OBJ(V-HD), A_{but}-NP(CL-HD) >> SUBJ-L

5.1.4 Similarities and Differences between TOPIC and FOCUS

The following table summarizes the major similarities and differences between TOPIC and FOCUS in Cantonese:

(4) Similarities and Differences between TOPIC and FOCUS in Cantonese

	TOPIC	FOCUS
Similarities		
<i>Syntactic requirements</i>	Topicalized function: a sentence-initial constituent being identified with a missing constituent in the sentence (Rosén 1998)	
<i>Syntactic position</i>	[SPEC, CP]	
<i>Particles</i>	Optionally followed by a particle	
Differences		
<i>Information features</i>	'limit[s] the applicability of the main predication to a certain restricted domain' and 'sets a spatial, temporal, or individual framework within which the main predication holds' (Chafe 1976: 50)	[+New] information: new, informative or unknown information (Choi 2001)
<i>Particles</i>	Followed by particles such as <i>aa4</i> and <i>ne1</i>	Followed by particles such as <i>aa3</i> , <i>laa1</i> and <i>gaa3</i> .
<i>Functional Uncertainty Equations</i>	(↑TOPIC) = (↑{SUBJ, OBJ, OBJ _θ , XCOMP, OBL, ADJ, PRED}) (Cannot be identified with the COMP function)	(↑FOCUS) = (↑{SUBJ, OBJ, OBJ _θ , COMP, XCOMP, OBL, ADJ, PRED}) (Can be identified with the COMP function)

5.2 Concluding Remarks and Future Research Directions

This study has presented an LFG account of topicalization and an OT-LFG account of OBJ topicalization in Cantonese. Functional uncertainty equations and constraint subhierarchies have been proposed. These findings demonstrate the applicability of the frameworks to the Cantonese language. In the following, areas for future research will be suggested.

5.2.1 Functional Uncertainty Equations

In the functional uncertainty equations established in the present study, the possible bottom functions of TOPIC and FOCUS have been presented. The legitimate body paths have however not been included. A questionnaire was distributed to 40 Cantonese native speakers, which contained topicalization constructions with different bottom functions and different body paths such as ‘COMP COMP’ and ‘COMP XCOMP’. The constructions become very complicated when the number of functions in the path increases and the acceptability of the constructions indicated by the informants varies greatly so that a conclusion could hardly be drawn. In order to confirm the acceptability of constructions with different body paths, a survey of a much larger scale has to be conducted and it will be left for future research.

5.2.2 The Role of TOPIC and FOCUS Particles

All of the topicalization constructions discussed in chapter 4 have a particle following the TOPIC or FOCUS. The existence of this particle leads to one more violation to certain constraints such as SUBJ-L, Abut-OBJ(V-HD), Abut-NP(CL-HD) and Abut-CLP(Q-HD). In fact, the particle is not always compulsory. For instance, the TOPIC topicalization construction (22A1) in chapter 4 is acceptable with or without the particle as shown below:

- (5) 本 字典 (啊) 我 掉咗 喇
bun2 zi6din2 (aa4) ngo5 deu6zo2 laa3
 CL dictionary PART 1.SG throw away.PERF PART
 ‘As for the dictionary, I have thrown it away.’

The particle following the FOCUS in example (8A) in chapter 2 is also optional:

- (6) 三 本 字典 (啊) 我 買咗
saam1 bun2 zi6din2 (aa3) ngo5 maai5zo2
 three CL dictionary PART 1.SG buy. PERF
 ‘It is three dictionaries that I have bought.’

Constructions with a particle following the topicalized phrase will become suboptimal with one more violation to constraints such as SUBJ-L than their no-particle counterparts, which contradicts the fact that both constructions are acceptable. Research can be done in the future on exploring the relevant constraints and constraint rankings which are responsible for accommodating both constructions, with and without a particle.

5.2.3 Towards One Hierarchy for Cantonese

As mentioned in chapter 4, a language is assumed to have one constraint hierarchy (subhierarchies are allowed where options exist) and the subhierarchies established for TOPIC and FOCUS constructions should be combined into one hierarchy for the Cantonese language. Combining the subhierarchies requires the analysis of constructions involving coexistence of TOPIC and FOCUS. Questions such as when it is possible to have both discourse functions grammaticized and which function should precede which have to be considered. Law (2003) proposes that a syntactic topic should precede a syntactic focus. Example (2) given in chapter 2, where the FOCUS precedes the TOPIC, is however also possible. More studies have to be done in order to confirm the possible relative orders of TOPIC and FOCUS. It is also worth researching whether the bottom functions of TOPIC and FOCUS stay the same in constructions where both discourse functions are grammaticized.

References

- Ackerman, Farrell and Gert Webelhuth. 1999. A Lexical-Functional Analysis of Predicate Topicalization in German. *The American Journal of Germanic Linguistics and Literature*, 11(1).
- Aissen, Judith L. 1992. Topic and Focus in Mayan. *Language*, 68(1):43-80.
- Asudeh, Ash. 2001. Linking, optionality, and ambiguity in Marathi. In Peter Sells (ed). *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CA.: CSLI Publications
- Bresnan, Joan. 2000. Optimal Syntax. In Joost Dekkers, Frank van der Leeuw and Jeroen van de Weijer (eds). *Optimality Theory*. New York: Oxford University Press.
- Bresnan, Joan. 2001. *Lexical-Functional Syntax*. Malden, Mass.: Blackwell Publishers.
- Bresnan, Joan and Sam A. Mchombo. 1987. Topic, pronoun, and agreement in Chichewa. *Language*. 63(4), 741–782.
- Bresnan, Joan and Jonni M. Kanerva. 1989. Locative inversion in Chichewa: a case study of factorization in grammar. *Linguistic Inquiry*, 20: 1-50. Reprinted in Stowell and Wehrli. (eds). 1992. *Syntax and the Lexicon* (Syntax and Semantics Series No. 26). New York: Academic Press, 53-101.
- Broadwell, G. A. 1999. Focus alignment and optimal order in Zapotec. *Proceedings of the 35th Chicago Linguistics Society*.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects and topics. In Charles N. Li (ed). *Subject and Topic*. New York: Academic Press.
- Cheung, Yam-Leung. 1997. A Study of Right Dislocation in Cantonese. Unpublished M.Phil Thesis. Hong Kong: The Chinese University of Hong Kong.

- Choi, Hye-Won. 1997. Information structure, phrase structure, and their interface. In Miriam Butt and Tracy Holloway King (eds). *Proceedings of the LFG97 Conference*. Stanford: CSLI Publications.
- Choi, Hye-Won. 1999. *Optimizing Structure in Context: Scrambling and Information Structure*. Ph.D. thesis, Stanford University.
- Choi, Hye-Won. 2001. Phrase structure, information structure and resolution of mismatch. In Peter Sells (ed). *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CA: CSLI Publications.
- Crowhurst, Megan and Mark Hewitt. 1997. *Boolean Operations and Constraint Interaction in Optimality Theory*. Ms., University of North Carolina, Chapel Hill and Brandeis University.
- Falk, Yehuda. 2001. *Lexical-functional grammar: an introduction to parallel constraint-based syntax*. Stanford, CA: CSLI Publications.
- Fung, Cauvis Suet-Man. 2006. TOPIC and FOCUS in Cantonese. Paper presented at Annual Research Forum, Linguistic Society of Hong Kong. The Hong Kong Polytechnic University, Dec 2-3.
- Fung, Cauvis Suet-Man. 2007. OBJECT Topicalization in Cantonese. Paper presented at the Fifth Cambridge Postgraduate Conference in Language Research. The University of Cambridge, Mar 20-21.
- Green, Melanie and Philip J. Jaggar. 2003. Ex-situ and in-situ focus in Hausa: syntax, semantics and discourse. In J. Lecarme (ed). *Research in Afrosiatic Grammar II*. Amsterdam: John Benjamins.
- Gregory, Michelle L. and Laura A. Michaelis. 2001. Topicalization and left-dislocation: a functional opposition revisited. *Journal of Pragmatics*, 33: 1665-1706.
- Grimshaw, Jane 1991. *Extended projection*. Ms. Brandeis University, Waltham, Mass.
- Halliday, Michael A. K. 1967. Notes on transitivity and theme in English part II. *Journal of Linguistics*, 3: 199-244.

- Her, One-Soon. 1989. Focus devices in Chinese revisited. Unpublished Manuscript. ECS, Inc.
- Her, One-Soon. 1991. Topic as a Grammatical Function in Chinese. *Lingua*. 84(1), 1-23.
- Huang, Chu-Ren. 1992. Certainty in Functional Uncertainty. *Journal of Chinese Linguistics*, 20(2): 247-287.
- Kager, Rene. 1999. *Optimality Theory*. Cambridge, U.K.; New York: Cambridge University Press.
- Kaplan, Ronald M. and Annie Zaenan. 1989. Long-distance dependencies, constituent structure, and functional uncertainty. In Mark R. Baltin and Anthony S. Kroch (eds). *Alternative Conceptions of Phrase Structure*. Chicago: University of Chicago Press.
- Kaplan, Ronald M., and Joan Bresnan. 1982. Lexical-Functional Grammar: A Formal System for Grammatical Representation. In Joan Bresnan (ed). *The Mental Representation of Grammatical Relations*. Cambridge, MA: The MIT Press.
- Kaplan, Ronald M. and John T. Maxwell. 1988. An algorithm for functional uncertainty. Proceedings of the 12th Conference on Computational Linguistics Volume 1. New York: Association for Computational Linguistics.
- Kaplan, Ronald M. and John T. Maxwell. 1996. LFG Grammar Writer's Workbench. XEROX PARC.
- King, Tracy Holloway. 1995. *Configuring Topic and Focus in Russian*. Stanford, California: CSLI Publications.
- King, Tracy Holloway. 1997. Focus domains and information-structure. In Miriam Butt and Tracy Holloway King (eds). Proceedings of the LFG97 Conference. Stanford: CSLI Publications.
- Kiparsky, Paul. 1987. *Morphology and grammatical relations*. Stanford, CA.: Stanford University Department of Linguistics unpublished ms.
- Kiss, Katalin É (ed). 1995. *Discourse Configurational Languages*. Oxford: Oxford University Press.

- Krapova, Iliyana. 2004. Word order in Topic-Focus structures in the Balkan languages. *România Orientale XVII*, 139-162.
- Krifka, M. 1997. Investigations into the Nature of Focus. Series of three lectures. City University of Hong Kong and Chinese University of Hong Kong, May 19-21.
- Kuhn, Jonas. 1999. The syntax and semantics of split NPs in LFG. In F. Corbin, C Dobrovie-Sorin and J. M. Marandin (eds). *Empirical Issues in Formal Syntax and Semantics 2, Selected Papers from the Colloque de Syntaxe et Sémantique à Paris (CSSP 1997)*. Thesus, The Hague, 145-166.
- Kuhn, Jonas. 2003. *Optimality-theoretic syntax: a declarative approach*. Stanford, Calif.: CSLI Publications.
- LaPolla, Randy J. 1988. Topicalization and the question of lexical passives in Chinese. Paper presented at the Third Annual Ohio State University Conference on Chinese Linguistics, Ohio State University, May 13-14.
- Lam, Olivia Si-Ching. 2004. Aspects of the Cantonese Verb Phrase: Order and Rank. M.Phil. Thesis. Hong Kong: The University of Hong Kong.
- Lambrecht, Knud. 1994. *Information structure and sentence form: topic, focus and the mental representations of discourse referents*. Cambridge: Cambridge University Press.
- Law, Ann. 2003. Right Dislocation in Cantonese as a Focus-Marking Device. UCL Working Papers in Linguistics 15.
- Lee, On-Man. 2003. The Subject Function in Cantonese. M.Phil. Thesis. Hong Kong: University of Hong Kong.
- Lee, Yat-Mei. 2002. Complement functions in Cantonese: a Lexical-Functional Grammar approach. M.Phil Thesis. Hong Kong: University of Hong Kong.
- Li, Charles N. and Sandra A. Thompson. 1976. Subject and Topic: A new typology of language. In Charles N. Li (ed). *Subject and Topic*. New York: Academic Press.
- Li, Charles N. and Sandra A. Thompson. 1981. *Mandarin Chinese: a Functional Reference Grammar*. Berkeley: University of California Press.

- Li, Daqin. 2001. Guanxihua dui huatihua de yingxiang: hanyu huati jieou ge'an fenxi [Influence of relativization on topicalization: A case study of topicalization in Chinese]. *Dangdai yuyanxue*, 3(2): 127-131.
- Linguistic Society of Hong Kong. 2003. *Jyutjyu pingjam zibiu* [Guide to LSHK Cantonese Romanization of Chinese Characters]. Hong Kong: Linguistic Society of Hong Kong.
- Lu, Yingshun. 2000. Guanyu zhuti de lai yuan wanti [A note on the source of the topic in Mandarin Chinese]. *Journal of Fuyang Teachers College (Social Science)*, 78: 6-7, 10.
- Matthews, Stephen and Virginia Yip. 1994. *Cantonese: A Comprehensive Grammar*. London: Routledge.
- McCarthy, John J. and Alan S. Prince. 1993. Generalized Alignment. In G. Booij and J. van Marle (eds). *Yearbook of Morphology*, 6: 79–153. Dordrecht, Kluwer.
- Morimoto, Yukiko. 2001. Verb raising and phrase structure variation in OT. In Peter Sells (ed). *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CA.: CSLI Publications, 129-196.
- Pan, Haihua and Jianhua Hu. 2003. Head noun movement, focus, and topicalization in Mandarin Chinese. *Language Science and Technology Monograph Series No. 1: Chinese Syntax and Semantics*, 119-156.
- Paul, Waltraud. 2002. Sentence-internal topic in Mandarin Chinese: the case of object preposing. *Language and Linguistics (Academia Sinica, Taipei)*, 3(4): 695-714.
- Paul, Waltraud. 2005. Low IP area and left periphery in Mandarin Chinese. *Recherches linguistiques de Vincennes*, 33: 111-134.
- Prince, Alan and Paul Smolensky. 1993. Optimality Theory: constraint interaction in generative grammar. *RuCCs Technical Report #2*. Piscataway, NJ: Centre for Cognitive Science, Rutgers University.
- Rochemont, M. and P. Culicover. 1990. *English Focus Constructions and the Theory of Grammar*. Cambridge: Cambridge University Press.

- Rosén, Victoria. 1998. Topics and Empty Pronouns in Vietnamese. Doctoral Dissertation. Department of Linguistics, University of Bergen.
- Sells, Peter. 1998. Optimality and Economy Of Expression in Japanese and Korean. In Noriko Akatsuka et al. (eds). *Japanese/Korean Linguistics*. Stanford, CA.: CSLI Publications. 7: 499-514.
- Sells, Peter. 2000. Alignment Constraints in Swedish Clausal Syntax. MS., Stanford University.
- Sells, Peter (ed). 2001. *Formal and Empirical Issues in Optimality Theoretic Syntax*. Stanford, CA.: CSLI Publications.
- Svolacchia, Marco, Lunella Mereu and Annarita Puglielli. 1995. Aspects of discourse configurationality in Somali. In Katalin É. Kiss (ed). *Discourse Configurational Languages*. Oxford: Oxford University Press.
- Tesar, Bruce and Paul Smolensky. 1998. Learnability in Optimality Theory. *Linguistic Inquiry*, 29(2): 229-268.
- Xu, Jinfeng. 1994. "Topicalisation" in Chinese. *Zhengzhou gongxueyuan xuebao (Sheke ban)*, 1: 67-73.
- Xu, Liejiong and Danqing Liu. 1998. Huati de Jiegou yu Gongneng [The Structure and Function of Topic]. In *Contemporary Chinese Linguistics*. Shanghai: Shanghai Jiaoyu Chubanshe.
- Yu, Hao-peng. 2005. Hanyu 'de' hou mingci duanyu zhutihua de zhiyue tiaojian [The topicalization of NP/DP after 'de' in Chinese]. *Journal of Zhengzhou Institute of Aeronautical Industry Management (Social Science Edition)*, 24(3):93-96.
- Zhang, Shuzheng. 1996. Guanyu zhuyu ji huati [A note on the subject and topic in Mandarin Chinese]. *Journal of Shandong University (Philosophy and Social Sciences)*, 3:45-50.
- Zaenan, Annie. 1980. Extraction rules in Icelandic. PhD Dissertation. Harvard University. Reprinted by Garland Press, New York, 1985.
- Zaenan, Annie. 1983. On syntactic binding. *Linguistic Inquiry*, 14: 469-504.

Appendix I The Questionnaire

QUESTIONNAIRE ON TOPICALIZATION IN CANTONESE

I am an M. Phil. student in the Faculty of Arts at the University of Hong Kong. I am doing research on the phenomenon of topicalization in Cantonese.

This questionnaire consists of three parts.

Part I

In this part, you would be presented with 21 Cantonese sentences. Please decide whether the given sentences are acceptable, strange but still acceptable, very strange, or unacceptable and put a tick in the corresponding boxes.

	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
1. 佢啊(aa4)訓咗喇				
2. 杯水啊(aa4)我飲咗				
3. 呢個問題啊(aa4)我問過佢				
4. 喺香港啊(aa4)佢住過架				
5. 同佢啊(aa4)我傾咗五分鐘 咋				
6. 你買咗本書啊(aa4)我以爲				
7. 去睇戲啊(aa4)佢應承咗我 架				
8. 買嗰本書啊(aa4)我叫過佢 啦				
9. 買呢我就會呢本書□(ge2)				
10. 喊呢我就唔會				
11. 佢啊(aa3)考第一				
12. 英文啊(aa3)我教佢				
13. 喺張枱啊(aa3)我放咗本書				
14. 喺張枱啊(aa3)我放咗本 書,唔係喺張櫈啊				

15. 琴日啊(aa3)我去睇戲				
16. 琴日啊(aa3)我去睇戲, 唔係今日啊				
17. 你買咗本書啊(aa3)我以為				
18. 去睇戲架佢應承咗我				
19. 買嗰本書架我提咗佢				
20. 買咗啦我嗰本書				
21. 翻咗啦佢				

Part II

In this part, you would be presented with 3 Cantonese sentences followed by 2 to 3 interpretations. Please decide whether the interpretations are acceptable, strange but still acceptable, very strange, or unacceptable with respect to the given sentences and put a tick in the corresponding boxes.

1. 喺德國我睇到電視報導奧運已經開始咗	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 我喺德國睇到電視報導奧運已經開始咗				
b. 我睇到電視喺德國報導奧運已經開始咗				
c. 我睇到電視報導奧運已經喺德國開始咗				

2. 喺呢間房我估佢擺咗盆花	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 我喺呢間房估佢擺咗盆花				
b. 我估佢擺咗盆花喺呢間房				

3. 喺香港我以為佢有三間屋	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 我喺香港以為佢有三間屋				
b. 我以為佢喺香港有三間屋				

Part III

In this part, you would be given 4 questions followed by 2 to 6 answers. Please indicate whether the answers are acceptable, strange but still acceptable, very strange, or unacceptable with respect to the questions and put a tick in the corresponding boxes.

1. 本字典呢?	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 本字典啊(aa4)我掉咗				
b. 我掉咗本字典				

2. 你買咗咩啊?	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 三本字典啊(aa3)我買咗				
b. 字典啊(aa3)我買咗三本				
c. 我買咗三本字典啊				

3. 你買咗三本咩啊?	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 三本字典啊(aa3)我買咗				
b. 本字典啊(aa3)我買咗三				
c. 字典啊(aa3)我買咗三本				
d. 我買咗三本字典啊				

4. 你買咗幾多本字典啊?	Acceptable	Strange, but still acceptable	Very strange	Unacceptable
a. 三本字典啊(aa3)我買咗				
b. 三本啊(aa3)我買咗字典				
c. 三本啊(aa3)我買咗				
d. 三啊(aa3)我買咗本字典				
e. 字典啊(aa4)我買咗三本				
f. 我買咗三本字典				

Thank you very much!

Appendix II Results of the Questionnaire Survey

The results of the questionnaire survey are shown in the tables below. The first column shows the constructions. The second column shows the corresponding functional uncertainty equations or descriptions of the constructions. The third column shows the mean scores of the constructions/interpretations.

Part I

Constructions	Functional Uncertainty Equations	Scores
1. 佢 啊 瞓咗 喇 <i>keoi5 aa4 fan3zo2 laa1</i> 3.SG PART sleep.PERF PART	(↑TOPIC) = (↑SUBJ)	+2.000
2. 杯水 啊 我 飲咗 <i>bui1 seoi2 aa4 ngo5 jam2zo2</i> CL water PART 1.SG drink.PERF	(↑TOPIC) = (↑OBJ)	+2.000
3. 呢 個 問題 啊 我 問過 佢 <i>nei1 go3 man6tai4 aa4 ngo5 man6gwo3 keoi</i> DEF CL question PART 1.SG ask.EXP 3.SG	(↑TOPIC) = (↑OBJ _θ)	+1.900
4. 喺 香港 啊 佢 住過 架 <i>hai2 Hoeng1gong2 aa4 keoi5 zyu6gwo3 gaa3</i> in Hong Kong PART 3.SG live.EXP PART	(↑TOPIC) = (↑OBL)	+1.600
5. 同 佢 啊 我 傾咗 五分鐘 咋 <i>tung4 keoi5 aa4 ngo5 king1zo2 ng5 fan1zung1 zaa3</i> with 3.SG PART 1.SG talk.PERF five minute PART	(↑TOPIC) = (↑OBL)	+1.950
6. 你 買咗 本書 啊 我 以為 <i>nei2 maai5zo2 bun2 syu1 aa4 ngo5 ji5wai4</i> 2.SG buy.PERF CL book PART 1.SG think	(↑TOPIC) = (↑COMP)	-0.300
7. 去 睇 戲 啊 佢 應承咗 我 架 <i>heoi3 tai2 hei3 aa4 keoi5 jing1sing4zo2 ngo5 gaa3</i> go see movie PART 3.SG promise.PERF 1.SG PART	(↑TOPIC) = (↑XCOMP)	+1.825
8. 買 個 本書 啊 我 叫過 佢 啦 <i>maai5 go2 bun2 syu1 aa4 ngo5 giu3gwo3 keoi5 laa1</i> buy DEF CL book PART 1.SG ask.PERF 3.SG PART	(↑TOPIC) = (↑XCOMP)	+1.800
9. 買 呢 我就 會 呢 本書 口 <i>maai5 nei1 ngo5 zau6 wui3 nei1 bun2 syu1 ge2</i> buy PART 1.SG zau will DEF CL book PART	(↑TOPIC) = (↑PRED)	-1.950
10. 喊 呢 我就 唔 會 <i>haam3 nei1 ngo5 zau6 m4 wui3</i> cry PART 1.SG zau not will	(↑TOPIC) = (↑PRED)	+1.900
11. 佢 啊 考 第一 <i>keoi5 aa3 haau2 dai6jat1</i> 3.SG PART exam first	(↑FOCUS) = (↑SUBJ)	+1.775
12. 英文 啊 我 教 佢 <i>jing1man2 aa3 ngo5 gaau3 keoi5</i> English PART 1.SG teach 3.SG	(↑FOCUS) = (↑OBJ _θ)	+1.925

13.	係張枱啊我放咗本書 <i>hai2 zoeng1 toi2 aa3 ngo5 fong3zo2 bun2 syu1</i> on CL table PART 1.SG put.PERF CL book	(↑FOCUS) = (↑OBL)	+1.450
14.	係張枱啊我放咗本書, 唔係 <i>hai2 zoeng1 toi2 aa3 ngo5 fong3zo2 bun2 syu1, m4hai6</i> on CL table PART 1.SG put.PERF CL book not 係張櫈啊 <i>hai2 zoeng1 dang3 aa3</i> on CL chair PART	(↑FOCUS) = (↑OBL)	+1.575
15.	琴日啊我去睇戲 <i>kam4jat6 aa3 ngo5 heoi3 tai2 hei3</i> yesterday PART 1.SG go see movie	(↑FOCUS) = (↑ADJ)	+1.075
16.	琴日啊我去睇戲, <i>kam4jat6 aa3 ngo5 heoi3 tai2 hei3,</i> yesterday PART 1.SG go see movie 唔係今日啊 <i>m4hai6 gam1jat6 aa3</i> not today PART	(↑FOCUS) = (↑ADJ)	+1.325
17.	你買咗本書啊我以為 <i>nei5 maai2zo2 bun2 syu1 aa3 ngo5 ji5wai4</i> 2.SG buy.PERF CL book PART 1.SG think	(↑FOCUS) = (↑COMP)	+1.950
18.	去睇戲架佢應承咗我 <i>heoi5tai2 hei3 gaa3 keoi5 jing1sing4zo2 ngo5</i> go see movie PART 3.SG promise.PERF 1.SG	(↑FOCUS) = (↑XCOMP)	+1.650
19.	買個本書架我提咗佢 <i>maai5 go2 bun2 syu1 gaa3 ngo5 tai4zo2 keoi5</i> buy DEF CL book PART 1.SG remind.PERF 3.SG	(↑FOCUS) = (↑XCOMP)	+1.675
20.	買咗啦我個本書 <i>maai5zo2 laa1 ngo5 go2 bun2 syu1</i> buy.PERF PART 1.SG DEF CL book	(↑FOCUS) = (↑PRED)	+1.900
21.	瞓咗啦佢 <i>fan3zo2 laa1 keoi5</i> sleep.PERF PART 3.SG	(↑FOCUS) = (↑PRED)	+1.975

Part II

1. 嗰 德國 我 睇 到 電視 報導 <i>hai2 Dak1gwok3 ngo5 tai2 dou2 din6si6 bou3dou6</i> in Germany 1.SG see dou television report 奧運 已經 開始咗 <i>Ngou3wan6 ji5ging1 hoi1ci2zo2</i> the Olympics already start.PERF	Interpretations	Scores
a. 我睇德國睇到電視報導奧運已經開始咗 I saw in Germany the TV report that the Olympics has already started.	(↑TOPIC) = (↑ADJ)	+2.000
b. 我睇到電視嗰德國報導奧運已經開始咗 I saw the TV report in Germany that the Olympics has already started.	(↑TOPIC) = (↑OBJ ADJ)	-2.000
c. 我睇到電視報導奧運已經嗰德國開始咗 I saw the TV report that the Olympics has already started in Germany.	(↑TOPIC) = (↑OBJ OBJ ADJ)	-2.000
2. 嗰 呢 間 房 我 估 佢 擺 咗 盆 花 <i>hai2 nei1 gaan1 fong2 ngo5 gu2 keoi5 baai2zo2 pun4 faa1</i> in DEF CL room 1.SG guess 3.SG put.PERF CL flower	Interpretations	Scores
a. 我嗰呢間房估佢擺咗盆花 I guess in this room that s/he has put a pot of flowers.	(↑TOPIC) = (↑ADJ)	-2.000
b. 我估佢擺咗盆花嗰呢間房 I guess that s/he has put a pot of flowers in this room.	(↑TOPIC) = (↑COMP OBL)	+2.000
3. 嗰 香港 我 以爲 佢 有 三 間 <i>hai2 Hoeng1gong2 ngo5 ji5wai4 keoi5 jau5 saam1 gaan1</i> in Hong Kong 1.SG think 3.SG have three CL 屋 <i>nguk1</i> house	Interpretations	Scores
a. 我嗰香港以爲佢有三間屋 I thought in Hong Kong that s/he had three houses.	(↑TOPIC) = (↑ADJ)	-2.000
b. 我以爲佢嗰香港有三間屋 I thought that s/he had three houses in Hong Kong.	(↑TOPIC) = (↑COMP ADJ)	+2.000

Part III

1. 本字典呢? <i>bun2 zi6din2 ne1?</i> CL dictionary PART 'Where's the dictionary?' Discourse Topic = CLP <i>bun2 zi6din2</i>	Descriptions	Scores
a. 本字典啊我掉咗 <i>bun2 zi6din2 aa4 ngo5 deu6zo2</i> CL dictionary PART 1.SG throw away.PERF	TOPIC = CLP OBJ	+1.975
b. 我掉咗本字典 <i>ngo5 deu6zo2 bun2 zi6din2</i> 1.SG throw away.PERF CL dictionary	Canonical SVO order	+1.600
2. 你買咗咩啊? <i>nei5 maai5zo2 me1 aa3?</i> 2.SG buy.PERF what PART 'What have you bought?' [+New] = QP <i>saam1 bun2 zi6din2</i>	Descriptions	Scores
a. 三本字典啊我買咗 <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF	FOCUS = QP OBJ	+1.925
b. 字典啊我買咗三本 <i>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2</i> dictionary PART 1.SG buy.PERF three CL	FOCUS = NP OBJ	+1.300
c. 我買咗三本字典啊 <i>ngo5 maai5zo2 saam1 bun2 zi6din2 aa3</i> 1.SG buy.PERF three CL dictionary PART	Canonical SVO order	+2.000
3. 你買咗三本咩啊? <i>nei5 maai5zo2 saam1 bun2 me1 aa3?</i> 2.SG buy.PERF three CL what PART 'What are the three books that you have bought?' [+New] = NP <i>zi6din2</i>	Descriptions	Scores
a. 三本字典啊我買咗 <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF	FOCUS = QP OBJ	+1.875
b. 本字典啊我買咗三 <i>bun2 zi6din2 aa3 ngo5 maai5zo2 saam1</i> CL dictionary PART 1.SG buy.PERF three	FOCUS = CLP OBJ	-1.675
c. 字典啊我買咗三本 <i>zi6din2 aa3 ngo5 maai5zo2 saam1 bun2</i> dictionary PART 1.SG buy.PERF three CL	FOCUS = NP OBJ	+1.375
d. 我買咗三本字典啊 <i>ngo5 maai5zo2 saam1 bun2 zi6din2 aa3</i> 1.SG buy.PERF three CL dictionary PART	Canonical SVO order	+2.000

4. 你 買咗 幾多 本 字典 啊? <i>nei5 maai5zo2 gei2do1 bun2 zi6din2 aa3?</i> 2.SG buy.PERF how many CL dictionary PART 'How many dictionaries have you bought?' [+New] = Q <i>saam1</i>	Descriptions	Scores
a. 三 本 字典 啊 我 買咗 <i>saam1 bun2 zi6din2 aa3 ngo5 maai5zo2</i> three CL dictionary PART 1.SG buy.PERF	FOCUS = QP OBJ	+1.975
b. 三 本 啊 我 買咗 字典 <i>saam1 bun2 aa3 ngo5 maai5zo2 zi6din2</i> three CL PART 1.SG buy.PERF dictionary	FOCUS = Q+CL OBJ	-0.575
c. 三 本 啊 我 買咗 <i>saam1 bun2 aa3 ngo5 maai5zo2</i> three CL PART 1.SG buy.PERF	FOCUS = Q+CL OBJ; NP dropped	+1.975
d. 三 啊 我 買咗 本 字典 <i>saam1 aa3 ngo5 maai5zo2 bun2 zi6din2</i> three PART 1.SG buy.PERF CL dictionary	FOCUS = Q OBJ	-1.825
e. 字典 啊 我 買咗 三 本 <i>zi6din2 aa4 ngo5 maai5zo2 saam1 bun2</i> dictionary PART 1.SG buy.PERF three CL	TOPIC = NP OBJ	+1.700
f. 我 買咗 三 本 字典 啊 <i>ngo5 maai5zo2 saam1 bun2 zi6din2 aa4</i> 1.SG buy.PERF three CL dictionary PART	Canonical SVO order	+2.000