The writing of this thesis owes much to many people, above all the committee members of my thesis committee: Prof. Gu Yang, Prof. Virginia Yip and Prof. Gladys Tang. I am immensely indebted to them.

First and foremost, I wish to express my deepest gratitude to Prof. Gu Yang, my thesis supervisor. She has been an endless source of support and inspiration. Her scholarly expertise in formal syntax and Chinese linguistics has influenced me in various respects. In particular, I would like to thank her for painstakingly commenting on my superficial observation and sketchy ideas when I was writing this thesis. The feedback I received from her has enabled me to clarify my thought and refine my arguments in this study. Without her constant encouragement, insightful suggestions and constructive criticisms, this thesis could not have been produced.

I am also obliged to the other two committee members: Prof. Virginia Yip and Prof. Gladys Tang for their guidance and intellectual stimulation. I have benefited immeasurably from their teaching of applied linguistics and also from discussing with them some of the issues considered in this thesis. My sincere gratitude goes to Prof. Stephen Matthews who has taught me various topics on
language typology and semantics. Thanks also go to Prof. Jiang King-Ping for introducing to me different issues in phonology.

I am grateful for having the chance to be friends with and learn from my classmates: Sarah Lee, Prudence Lau and Catherine Wong. They have offered me tremendous support and unconditional help throughout my study.

For earlier guidance and encouragement, I would like to thank Prof. Tang Sze-wing and Dr. Gu Gang. They have introduced to me their own work on various topics in generative grammar. I also want to thank Prof. Thomas Lee who first taught me linguistics.

Thanks are also due to the staff of my department: Sirwinnie Cheung, Judy Leung and Manna Wong for their care and support. In particular, I wish to express my special thanks to Sirwinnie for her kind help and professional assistance at various stages of my study. I will cherish the memory of her.

For their encouragement and help, I want to thank Felix Sze, Leo Wong, Ben Au Yeung, Simon Huang, Dr. Wang Lidi, Scholastica Lam, Angel Chan, Queenie Wai, Peppina Lee, Tommi Leung, Prof. Cathy Wong, Carine Yiu, Maki Iida, Uta Lam, Gene Chu, Kenny Chu, Fiona Keung, Kevin Wong, Olive Cheung, Michelle Wong, Geoffrey Cheung, Lisa Lam, Sam Chan, Eliza Chan, Winnie Chan, Kian Iu, Pang Po-ming, Justin Wan, Nicholas Chan, Gwai Go, Kenneth Chu, Winne Wong and the team members of the Social Work Team, CUHK.

Parts of this thesis were developed from a paper which I presented at the Postgraduate Research Forum on Language and Linguistics 2001. My thanks go
to the participants there for their valuable comments. I would also like to thank the members of the organizing committee: Hu Jianhua, Dr. Zhang Wei and Nishida Fuminobu for their guidance and support during the organization of the Forum.

Finally, I thank my parents and brother. I owe a great deal to them for their love, care, patience and encouragement. Without their support, I could not have completed this study. I dedicate this thesis to them.
# Table of Contents

## Acknowledgements

Page  

## Abbreviations and Symbols

Page vi

## Abstract (English)

Page vii

## Abstract (Chinese)

Page viii

### Chapter One  Characteristics of Compounds: An Overview

1.0 Introduction  

Page 1

1.1 General Properties of Compounds  

1.1.1 Lexical Properties  

Page 4

1.1.2 Phrasal Properties  

Page 12

1.1.3 Complexity of the Status of Compounds  

Page 16

1.2 Relation between Morphology and Syntax  

1.2.1 Lexicalist Approach  

Page 18

1.2.2 Syntactic Approach  

Page 19

1.2.3 Parallel Approach  

Page 20

1.2.4 A Note on the Lexicon  

Page 21

1.3 Research Objectives  

Page 24

1.4 Summary and Organization of the Thesis  

Page 29

### Chapter Two  Notions of Word, Compound and Phrase

2.0 Introduction  

Page 31

2.1 Criteria of Wordhood  

Page 32

2.2 Distinction between Compounds and Phrases  

2.2.1 Introduction  

Page 39

2.2.2 Syntactic Aspect  

Page 43

2.2.3 Semantic Aspect  

Page 45

2.2.4 Morphological and Phonological Aspects  

Page 48

2.2.5 Summary  

Page 50

2.3 The Lexical Integrity Hypothesis (LIH)  

Page 52

2.4 Further Consideration on the Nature of Compound  

2.4.1 Compounding and Other Combinatory Processes  

Page 60

2.4.2 Status of the Constituents  

Page 62

2.4.3 Degree of Separability  

Page 64

2.4.4 Definitions of Compounds  

Page 66

2.5 Summary  

Page 69
**Abbreviations and Symbols**

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASP</td>
<td>Aspect marker</td>
</tr>
<tr>
<td>CL</td>
<td>Classifier</td>
</tr>
<tr>
<td>PRT</td>
<td>Particle</td>
</tr>
<tr>
<td>*</td>
<td>Ungrammatical/Unavailable</td>
</tr>
<tr>
<td>?</td>
<td>Questionable/Semantically odd</td>
</tr>
<tr>
<td>X/X&lt;sup&gt;0&lt;/sup&gt;</td>
<td>A lexical item of grammatical category X</td>
</tr>
<tr>
<td>X'</td>
<td>An X-bar, projection of X&lt;sup&gt;0&lt;/sup&gt;</td>
</tr>
<tr>
<td>XP</td>
<td>An X phrase, maximal projection of X&lt;sup&gt;0&lt;/sup&gt;</td>
</tr>
<tr>
<td>SPEC</td>
<td>Specifier</td>
</tr>
<tr>
<td>COMP</td>
<td>Complement</td>
</tr>
<tr>
<td>t</td>
<td>Trace</td>
</tr>
<tr>
<td>e</td>
<td>Empty category</td>
</tr>
</tbody>
</table>
Abstract

This thesis aims at exploring the behavior and the nature of a particular type of compound in Cantonese, namely verb-object compounds (VOCs) such as *hoi-sam* “open-heart = to be happy”, *ceot-maau* “go out-cat = to cheat in an examination” and *waai-ji* “bear-doubt = to doubt somebody”. The study of the process of compounding is of linguistic interest because many compounds are observed to exhibit both morphological and syntactic properties. In this thesis, we depart from traditional analyses by showing that there are in fact two types of VOCs in Cantonese: *lexical VOCs* and *phrasal VOCs*. The former is inseparable and the latter is separable by certain syntactic constituents. In order to explain the formation of these two types of compounds, we argue that (i) two levels of representation exist prior to syntax, to wit, the lexical-semantic interface and the lexical-syntactic interface (in the spirit of Levin and Rappaport Hovav 1995) at which the lexical VOCs and phrasal VOCs are formed respectively, and (ii) the internal verb-object structure of lexical VOCs is irrelevant to syntax because the level at which they are formed is far from syntax while the internal structure of phrasal VOCs is visible to syntax because they are formed at a level which is closer to syntax.

This study shows that acknowledging the difference with respect to the levels of formation of the lexical and the phrasal VOCs in Cantonese better explains the observed difference in the behavior of these two types of Cantonese VOCs. Our findings also lead us to a deeper understanding of the status of morphology and its relation to syntax.
摘要

本論文詳細探討粵語動賓複合詞(例如：「開心」、「出貓」和「懷疑」)的表現和性質。複合詞構成過程的討論具有語言學上的研究價值，理由是不少複合詞同時顯示出詞匯及句法上的特質。本論文有別於傳統的研究，我們發現粵語中有著兩類不同的動賓複合詞，分別是詞匯複合詞和句法複合詞。前者不可被分割；後者則能被某些詞組結構分割。就上述兩種動賓複合詞的衍生方法，本論文提出以下兩點：(i) 在句法層面之前，有兩個表達式存在。即詞匯語義介面及詞匯句法介面(依據 Levin 和 Rappaport Hovav 1995 的研究)。詞匯複合詞和句法複合詞分別在這兩個介面上形成。(ii) 詞匯複合詞的動賓內部結構不受句法影響，原因是這類複合詞產生於一個遠離句法層面的介面上。然而，句法卻可看到句法複合詞的內部結構，因它們於一個較接近句法層面的介面上形成。

指出詞匯複合詞和句法複合詞形成於不同的介面上，讓我們能更清楚解釋這兩類粵語複合詞的不同表現。研究結果亦令我們更了解詞匯的性質，以及詞匯跟句法的關係。
Chapter One  Characteristics of Compounds: An Overview

1.0 Introduction

Under the current generative approach to linguistics, the grammatical system (the Language Faculty) consists of different modules and each module has its own set of rules and principles (Cook and Newson 1996, Culicover 1997, Sproat 1998). The present study is concerned with the relation between two specific modules of the grammar, namely morphology and syntax. Morphology deals with the formation and structure of words. The study of morphology "approaches morphemes as the minimal linguistic units with semantic content, and studies the relations among them" (Bybee 1985:v). Syntax concerns the combination of words into structural constituents. The study of syntax focuses on the "process whereby words are combined to form phrases which in turn are combined to form sentences" (Radford et al 1999:279). At first glance the boundary between the two modules is fairly clear. However, this is not true. In fact, there is a long-standing debate on the distinction between morphology and syntax. Issues on the status of morphology and its relation to syntax have aroused intense discussions over the past few decades (Chomsky 1970, Halle 1973, Roeper and Siegel 1978, Anderson 1982, 1985, 1988, LaPointe 1987,

The nature of morphology has been rigorously studied and most of the discussions have centered on the autonomy of the morphological module. There are two main reasons why some linguists cast doubt on the existence of an independent^4 morphological component. First, proponents of an exclusively syntactic word formation claim that morphology can be subsumed under syntax and so it should not be regarded as a separate module. They argue that syntactic principles, which are independently motivated, are powerful enough to account for the richness of word formation phenomena (Fabb 1984, Roeper 1988, Baker 1988a, Reuland 1990). Another reason for some linguists to question the status of morphology is that outputs of certain word formation processes are observed to exhibit both lexical and phrasal properties. Compounds, the subject of this research, constitute an output of this kind; they behave like both words and phrases. The existence of this kind of lexical items which have a dual nature leads some linguists to think that morphology may not exist as an autonomous module (Pesetsky 1985, Sproat 1985).

This thesis attempts to explore the autonomic status of morphology and the relation between morphology and syntax. Compounding has been chosen as the focus of the present study, because it reflects both lexical and syntactic properties. A close examination of the process of compounding will definitely help us understand more about the nature of morphology and its relation to syntax. In the
following chapters we will show that compounding takes place at the interfaces prior to syntax. The idea that a grammatical process may involve more than one component is not novel in the literature. Many studies argue that grammatical operations can take place at the interfaces of different components of the grammatical system (as observed in Bayer and Lahiri 1990 for Bengali cliticization; Ackema and Schoorlemmer 1994 for Dutch middle construction, Levin and Rappaport Hovav 1995 for English causativization; Ralli and Stavrou 1997, Smirniotopoulos and Joseph 1998 for Greek compounding). Following this line of research, the present study investigates the interface properties of a particular type of compound, namely the verb-object compounds in Cantonese and provides a theoretical account of the nature and formation of this type of compound. It is hoped that this study can bring a deeper understanding of the notions of word, compound and phrase, and shed new light on the relation between morphology and syntax.

1.1 General Properties of Compounds

A compound is generally assumed to be a word formed from two or more words (Selkirk 1982, Matthews 1991, Spencer 1991, 1998, Tang 1991, Payne 1997, among others). But the fact is that there is no satisfactory definition of compound in the linguistic literature so far. The nature of compound remains unclear. Another problematic issue related to the process of compounding is that
it resembles both syntactic phrase formation and word formation (Bybee 1985, Matthews 1991, Spencer 1991). In order to have a preliminary understanding of the characteristics of a compound, a general survey of the major lexical and phrasal properties of compounds will be given in sections 1.1.1 and 1.1.2 respectively. The definition of compound will be dealt with in Chapter Two.

1.1.1 Lexical Properties

In general, there are four main properties of compounds that make them resemble words. First, compounds—like words—often have noncompositional or idiosyncratic meaning. Their meanings are not always predictable from a summation of the meaning of their parts as noted by Bybee (1985). For instance, a hot dog is not a kind of dog but a kind of food and it may not necessarily be hot. Now consider some Cantonese examples taken from Lo and Tam (1996):

(1) a. *ceot-maau*

    go out – cat

    “to cheat in an examination”

b. *sik-sei maau*

eat – dead cat

    “being accused falsely”
c. laai-maau

link – cat

“to shirk responsibilities”

The meanings of compounds in (1a) to (1c) are not compositional. Although they all contain the word *maau* “cat”, they have nothing to do with cats. Further examples of Cantonese compounds which have idiosyncratic meanings are given in (2a) to (2s) below:

(2) a. lei-pou leave – score “be outrageous”
b. bok-meng risk – life “be diligent”
c. ce-haa pull – shrimp “to pant”
d. gau-fo save – fire “to put out fire”
e. zuu-gai run – chicken “to lose a chance”
f. bong-seoi weight – water “to pay”
g. daa-zam hit – needle “to inject somebody with drug”
h. daa-gei hit – machine “to play TV or computer games”
i. daa-bou hit – book “to update a bankbook”
j. bou-seoi replenish – water “to give extra money”
k. haang-syun walk – ship “to be a sailor”
l. zoek-cou wear – grass “to escape”
The compounds in (2a) to (2s) are semantically opaque. One cannot get the meanings of these compounds from their constituents.

Another resemblance between a compound and a word is the property of nonreferentiality exhibited by them. First, let us consider some examples concerning words. The examples given in (3) are taken from Postal (1969):

(3) a. Chomsky likes Chomsky-ites/*him-ites.

b. All Bloomfieldians like Bloomfield/*him.

From (3), one can see that anaphoric rules cannot apply to refer to a subpart of a word. Sproat (1988) argues that words are anaphoric islands, showing that the proper subparts of a word are not referential. Now we turn to compounds. It is a well-known fact that a noun in a compound has a generic rather than a referential function. This can be shown by the examples in (4):
Consider the compound *hoksaang kek se* "student drama club" in example (4a) first. The non-heads of the compound, i.e. *hoksaang* "student" and *kek* "drama" never refer to specific objects. Neither *hoksaang* "student" nor *kek* "drama" will pick out any specific student or drama. Similarly, in (4b) *Cengnin Manhok Zoeng* "Youth Literary Awards", *cengnin* "youth" does not point to any definite young person; *manhok* "literature" does not refer to any particular kind of genre. Both terms are used generically. In (4c), *suksaang wui* "resident association" refers to the people living in the hostel in general but not any specific resident. Apart from the Cantonese examples given in (4), the nonreferentiality of the compounded noun in compounds can be further illustrated by an English example like *garbage man*. Downing (1977) notes that it is not the case that every man who takes out the
garbage is a garbage man, because the word man in the compound is generic rather than referential.

A third property that makes compounds resemble words is that compounds exhibit lexical integrity. Lexical integrity refers to the phenomenon that a lexical item cannot be split up by other words or phrases. This can be illustrated by the Cantonese word maa-faan “troublesome” as in (5) below:

(5) a. maa-faan

  adjective: troublesome

  noun: trouble or problem

  verb: to trouble somebody

b. maa-gwai-faan (where gwai is an infix)

  “downright troublesome”

c. *maa nei faan (where nei is a word meaning “you”)

  Intended meaning: “to trouble you” as in maa-faan nei

d. *maa houdaaige faan (where houdaaige is a phrase meaning “very big”)

  Intended meaning: “a big trouble” as in houdaaige maa-faan

An infix can enter a single morpheme (word). Since gwai is an infix used to emphasize adjectives in Cantonese (Matthews and Yip 1994), it can be infixed into the word maa-faan “troublesome” as shown in (5b). In contrast, insertion of a word (as in 5c) or a phrase (as in 5d) is not allowed, because a lexical item, like
maa-faan, is treated as a single unit in syntax. The word maa-faan does not allow anything which is a syntactic constituent to enter into it. Compounds also observe lexical integrity, as exemplified in (6) and (7):

(6) a. gwaan-sam

   close – heart

   “to care about somebody or something”

   b. *gwaan nei sam

   close you heart

   (intended meaning: “to care about you” as in gwaan-sam nei)

(7) a. zau-haumun

   enter – back door

   “to attain one’s goal by means of improper ways or personal connection” (Lo and Tam 1996:173)

   b. *zau zingfuge haumun

   enter government’s back door

   (intended meaning: “to attain one’s goal by means of improper ways or personal connection with government officers”)

In (6), a word nei “you” is inserted into the compound gwaan-sam. In (7), a phrase zingfuge “government’s” is inserted into the compound zau-haumun. The results of the two insertions are ungrammatical as indicated in (6b) and (7b) respectively. This can be attributed to the lexical integrity of the compounds
which does not allow the constituents of a lexical item to be split up by other words or phrases.

The last major lexical property concerning compounding is its limited productivity. Matthews (1991) points out that in English we have compounds like blackbird and bluebird, blackthorn and whitethorn. However, strings with a similar structure such as whitebird and redbird, greenthorn and pinkthorn do not exist. He suggests that these facts demonstrate that compounding, like word formation, is not fully productive. Cantonese has a similar situation. Consider the compounds in (8) below:

(8) a. saunganjyun “cashier”
    b. wuwaijyun “security guard”
    c. gausaangjyun “lifeguard”

The Cantonese compounds in (8) are names for different occupations with the structure: NATURE OF THE JOB + JYUN where jyun means “person”. However, this formula is not fully productive. If one is talking about people who work in the field of citgai “design” or ginzuk “architecture”, one does not call them *citgaijyun or *ginzukjyun. Another formula: NATURE OF THE JOB + SI where si means “person” has to be employed, thereby obtaining citgaisi “designer” and ginzuki “architect”. These facts show that compounding is not fully productive.

Another aspect showing the unproductiveness of compounding is its limited
recursiveness. For instance, the only clearly recursive compound type in English is the NN\%N\textsuperscript{11} combination (a compound noun formed from nouns). Selkirk (1982) uses the compound noun towel rack to show that the compounding with respect to the formation of compound nouns is recursive in English:

(9) a. towel rack
b. bathroom towel rack
c. bathroom towel rack designer
d. bathroom towel rack designer training

Selkirk (1982) states that the compound towel rack in (9a) can be expanded by adding bathroom to become the compound bathroom towel rack in (9b). This compound may be further expanded to another compound by adding the noun designer to form bathroom towel rack designer as in (9c). One can see in (9d) that this compound may be expanded in a similar fashion to form bathroom towel rack designer training. However, other types of combinations will rarely be recursive. Fabb (1998) suggests that the unproductive nature of the process of compounding provides an answer to the question as to why some rules for building compounds (as in Selkirk 1982)\textsuperscript{12} are manifested by very few actual compounds. For instance, compounds of the VN\%N type (a compound noun consisting of a verb and a noun) like swearword, scrubwoman and rattlesnake are very unproductive in English.
1.1.2 Phrasal Properties

In the literature, many studies regard compounds as words. As such the internal structure of compounds will not be relevant to syntax, because syntax treats lexical items (including compounds) as single units. But one of the peculiar properties of compounds is that they sometimes behave like phrases; they seem to have an internal structure.

Compounds resemble syntactic structures in various ways. This section reviews two major phrasal properties exhibited by compounds. First, compounds have a constituent structure. Matthews (1991) shows that there are layers in compounds, which are like layers of syntactic constructions. He uses fruit-juice carton for illustration. Fruit-juice is compounded of fruit and juice. This compound in turn combines with carton to form another compound fruit-juice carton. The compound [N fruit-juice] is embedded within the compound: [N [N fruit-juice] carton]; just like phrases are embedded within phrases in syntax: [DP the juice of [DP the fruit]]. The structural embedding of the compound fruit-juice carton and the phrase the juice of the fruit can be represented hierarchically as the tree diagrams in (10) and (11) respectively:
Comparing the tree structures in (10) and (11), one can see that a compound resembles a phrase in that it has a constituent structure.

Further evidence for a hierarchical structure of compounds comes from the interpretation of multiple-word (three or more words) compounds. Typical examples can be found in Cantonese, as shown in (12) below:

(12) a. Jatbun jamngok lousi  
    Japanese music teacher

b. [Jatbun [jamngok lousi]]
   [Japanese [music teacher]]
c. [[Jatbun jamngok] lousi]

[[Japanese music] teacher]

The compound noun *Jatbun jamngok lousi* in (12a) is interpreted as “a music teacher whose nationality is Japanese” if one assigns it with the structure of (12b). But (12a) can also mean “a teacher of Japanese music” if it has the structure of (12c). One can see that the interpretation of a compound depends on the way it is built up. This suggests that compounds have a hierarchical structure like a phrase rather than a flat structure.

The second aspect of compounding that mimics syntax can be seen from the semantic (meaning) relations holding between the constituents of compounds. For instance, consider the examples from Matthews (1991):

(13) a. She is a good book-keeper.

b. She is good at keeping the books.

Sentence (13b) is a paraphrase of sentence (13a). In (13a), *book* in the compound *book-keeper* is understood to be in relation to *keeper* in the same way as *books* are understood in relation to *keeping* in the verb phrase *keep the books* in (13b).

In other words, a parallel between the compound and a corresponding phrase in terms of meaning is observed. Here one can see that the meaning relation holding between the constituents of a compound can be the same as that one can
Spencer (1991) discusses the semantic relations between the components of compounds in detail. He suggests that head-modifier, predicate-argument, and appositional relations which are held between the constituents of phrases exist between constituents of compounds. The head-modifier relation can be found in endocentric compounds (i.e. compounds with head). For instance, in compound nouns English major, Chinese major and Music major, major is the head. The modifier constituents are English, Chinese and Music respectively, and they attribute a property to the head, since they specify the subject of the major. In exocentric compounds (i.e. compounds with no head), the predicate-argument relation can be observed. Examples of such compounds include, in English, pickpocket, cutthroat and sawbones, and in Cantonese, ji-tou “cure-belly = to eat” and saai-meng “shine-life = to show off”. These compounds consist of a predicate-type constituent (pick, cut, saw, ji “cure” and saai “shine”) and an argument-type constituent (pocket, throat, bones, tou “belly” and meng “life”). The appositional relation is found in appositional compounds which are “formed by a simple conjunction of two elements, without any further dependency holding between them” (Spencer 1991:311). Fabb (1998) points out that the constituents in these compounds equally share head-like characteristics. Examples include English compounds: mother-child, Austria-Hungary, student-prince and Cantonese compounds such as hoi-gwaan “open-close = switch” and maaau-teon “spear-shield = contradiction”. To sum up, semantic
relations between the constituents of phrases are also held between the constituents of compounds.

1.1.3 Complexity of the Status of Compounds

In sections 1.1.1 and 1.1.2 we have shown that compounds exhibit both lexical and phrasal properties. This dual nature of compounds has made the study of their nature very difficult, because it is hard for us to decide whether we are looking at morphology or syntax or both. Another way of expressing this is to say that in the investigation of the process of compounding, we are, in fact, trying to understand the complex relation between the morphological and the syntactic components. In the next section, we review some major approaches in the study of the relation between morphology and syntax.

1.2 Relation between Morphology and Syntax

As we have shown in sections 1.1.1 and 1.1.2, compounds resemble both words and phrases. This particular nature of compounds has naturally led one to ask questions like “What is the nature of the process of compounding?” Intuitively, compounding, a word formation process, should be a morphological process and take place in morphology. But as we can recall, in section 1.1.2 we have reviewed different phrasal properties of compounds. In fact, many
morphosyntactic analyses argue that compounding takes place in syntax (Fabb 1984, Sproat 1985, Roeper 1988, Lieber 1988). Consequently, another set of questions regarding compounding arises: "Is it possible that a morphological process like compounding takes place in syntax? If so, what is the mechanism of this process?"

From a theoretical point of view, the study of compounding is not only an examination of the formation of compounds, but also an exploration of the very nature of morphology. Specifically, we are concerned with the autonomy of morphology—whether it is an independent module in the linguistic system or not. At the same time, we are investigating the relation between morphology and other modules in the grammatical system, in particular, syntax. All these are fundamental issues which are related to the study of morphology. Among these issues, the question of the relation between the morphological and syntactic components receives the greatest attention from linguists.

Various positions with respect to the relation between morphology and syntax have been argued for in the literature. Basically, there are three positions: 14 (i) morphology is strictly distinct from syntax (Aronoff 1976, 1994, Di Sciullo and Williams 1987, Anderson 1992, Bresnan and Mehonmo 1995, Ackema 1998, 1999, Bisetto and Scalise 1999, among others), (ii) morphology is subsumed under syntax (Fabb 1984, Sproat 1985, Baker 1988a, Lieber 1988, Roeper 1988, Drijkoningen 1994, among others), and (iii) morphology is independent but it can enter syntax (Borer 1988, 1991, 1997, Shibatani and
1.2.1 Lexicalist Approach

One of the most important assumptions underlying much of the work in the current studies of morphology is the Lexical Integrity Hypothesis, the LIH, which is developed in the seventies (Chomsky 1970, Jackendoff 1972, LaPointe 1979) and has been formulated in various versions over the years (Lexicalist Hypothesis, Strong Lexicalist Hypothesis, Lexical Integrity Hypothesis, and Thesis of Atomicity). The essence of all these formulations of the Hypothesis is that syntax is blind to the internal structure of words and cannot affect it. As noted by Spencer (1991), the Lexicalist Hypothesis is the logical consequence of the division between morphology and syntax. In other words, lexicalists hold that morphology, which determines the well-formedness of words, is an independent module, having its own set of rules and principles, and this module cannot be subsumed under other components of the grammatical system such as syntax. In Bisetto and Scalise’s (1999) study of Italian compounds, for instance, they examine the nature of the compounds, e.g. capostazione “station master” by carrying tests such as head deletion (*haao licenziato il capostazione e il __ reparto “*the station master and the department __ have been fired”) and
topicalization ("stazione, hanno licenziato il capo — "*station, (they) have fired
the master") and conclude that morphological formation cannot be accounted for
by syntax.

1.2.2 Syntactic Approach

Many studies (Fabb 1984, Sproat 1985, Baker 1988a) assume that syntax is
responsible for word formation and deny morphology its status as an
independent module. These studies aim to extend the highly articulated models
of syntax to the domain of word formation, arguing that the reduction of
morphology to syntax can minimize the complexity of the grammar and thus
maximizing the learnability\(^\text{16}\) of it. Consider, for instance, Baker’s (1988a)
analysis which is the most rigorous syntactic account of morphological
formations. In his study, it is argued that noun incorporation\(^\text{17}\) is derived via a
syntactic process, namely head-to-head movement. His analysis of noun
incorporation in Southern Tiwa can be illustrated as (14) below (the diagram is
taken from Ackema (1999:24)):
The diagram in (14) is a tree representation of the sentence involving noun incorporation: Yede a-seuan-mū-ban. (where ban indicates past tense) “You saw that man.” (14) shows that the structure is derived by adjourning the N⁰ of the NP to the V⁰. With this model, Baker is able to account for the phenomenon of noun incorporation in terms of syntax.

1.2.3 Parallel Approach

Instead of explaining morphological phenomena either from a lexical perspective or a syntactic point of view, some linguists propose that while maintaining an independent morphological (or word formation) component, word formation processes can, in fact, take place at different levels. Specifically, morphology is not ordered prior to syntax but is parallel to it. Thus, words can be formed (i) before syntax, (ii) in syntax and (iii) after syntax. This approach to
word formation is known as Parallel Morphology (Borer 1988, 1991, 1997, Spencer 1991). In Borer’s (1988) study of compounds and constructs in Hebrew, she makes use of the idea of parallel morphology to account for the syntactic properties of construct state nominals. According to her analysis, Hebrew constructs (e.g. נְפַל הַיָּלָדָה “the girl’s scarf”) and compounds (e.g. בְּיֵית חוֹלִים “hospital”) are morphological formations. However, construct state nominals in Hebrew show several syntactic properties, for instance, they are productive and semantically transparent, but Hebrew compounds are not productive and are semantically opaque. Borer argues that the idea of parallel morphology provides a straightforward explanation to the difference in the syntactic behavior between compounds and constructs in Hebrew. Specifically, she proposes that the two types of structures are both derived by the same word formation component but at different levels of the grammatical system: compounds are formed at D-structure while constructs are formed after D-structure, i.e. at syntax. This difference in the level of formation is responsible for the different syntactic properties (e.g. productivity and semantic compositionality) observed in compounds and construct state nominals in Hebrew.

1.2.4 A Note on the Lexicon

This study deals with the nature and boundary of the modules in the grammatical system. The purpose of this section is to clarify the confusion
concerning the notions of two linguistic levels: morphology and the lexicon. Many studies on morphology have been devoted to the characterization of the nature of the lexicon and its relation to the morphological component (Lieber 1980, Perlmutter 1988, Packard 1990, Gu 1996, Cruse 2001). One of the common claims in the literature is that morphology is operative in the lexicon, meaning that morphological processes take place inside the lexicon. Note that some linguists, however, maintain a restrictive use of the term “lexicon”. In her study of word structure, Selkirk (1982) points out that two senses of the concept of the lexicon can be identified:

(15)

[T]he rules of word structure form part of what one may call the *lexical component* or simply the *lexicon* (understood in a broad sense). As it is viewed here and in most earlier theories, the lexical component contains a variety of subcomponents. First, it contains a list of freely occurring lexical items (which I will assume to be words, in English). We may call this *dictionary* (or *lexicon*, in the restricted sense). Second, it contains a list of the bound morphemes of the language. This, together with the dictionary proper, I will call the *extended dictionary*. Third, the lexical component includes the set of rules characterizing the possible morphological structures of a language

Aronoff (1994) shares a similar view with respect to Selkirk’s restricted use of the term “lexicon” in (15). He suggests that the term “lexicon” should be understood as referring only to the list of all idiosyncratic signs, regardless of their category or complexity. In a discussion on lexicalization and productivity,
Aronoff and Anshen (1998:237) clearly distinguish the lexicon from morphology:

(16)

[T]he two systems do have a great deal to do with one another, for two simple reasons. The first is that they serve the same role in a language: both provide words. This overlap has even led some linguists to say that morphology is “in the lexicon” (Jensen and Strong-Jensen 1984), although in doing so, these linguists are using the term “lexicon” in a much broader and different sense, to mean the source of all words, actual and potential, rather than in the narrow sense of a list of unpredictable items that we have inherited from the traditional grammar and from Bloomfield (Bloomfield 1933, Zwicky 1989, Aronoff 1994). The second reason is that morphology and the lexicon are interdependent. Most centrally, the morphology, which forms words from words, finds the words that it operates on (its bases) in the lexicon.

Since our concern in this study is the nature of morphology and the interaction between morphology and syntax, we will not get into the details of the controversy about the relation of morphology and the lexicon. The present study, following Aronoff and Anshen (1998), regards the lexicon of a language as a list of arbitrary signs (e.g. words, idioms, affixes) that a speaker knows, and assumes that morphology operates on items taken from the lexicon. The advantage of adopting a restricted sense of the term “lexicon” is that this view of the relation between the lexicon and the morphological component is not subject to the debate on the organization of the lexicon that complicates much of the literature on the nature of morphology.
1.3 Research Objectives

The present study follows the tradition of generative grammar in that it assumes that the grammatical system is composed of various distinct modules or components (Cook and Newson 1996, Culicover 1997, Sproat 1998). Specifically, it assumes that “the well- or ill-formedness of an expression is determined not by a single monolithic set of rules, but rather by a set of modules (or components), each formally independent of the other, and each with its own set of rules or principles that must be satisfied” (Sproat 1998:335).

On this basis, we are going to deal with the issues related to the status of morphology and its relation to syntax by conducting a detailed examination on the process of compounding as it manifests the complexity of words and phrases. The present study focuses on a particular type of compound: *verb-object compounds* in Cantonese. Compounding is a very productive means of word formation process in this language, as exemplified in (17) below:\(^9\)

(17)

a. Subject-Predicate Compounds

\[
\begin{array}{ccc}
\text{dei-zan} & \text{earth – vibrate} & \text{“earthquake”} \\
\text{tau-tung} & \text{head – pain} & \text{“headache”} \\
\text{meng-fu} & \text{life – bitter} & \text{“unfortunate”} \\
\text{nin-heng} & \text{year – light} & \text{“young”} \\
\text{sam-jyun} & \text{heart – soft} & \text{“soft-hearted”}
\end{array}
\]
b. Coordinate Compounds

- **tiu-gin**: entry – item
  - “condition”
- **hoi-gwaan**: open – close
  - “switch”
- **gaau-sau**: teach – transfer
  - “professor”
- **maau-teon**: spear – shield
  - “contradiction”
- **ji-kaau**: follow – lean
  - “to depend on somebody”

c. Synthetic Compounds

- **jin-gau-juyn**: study – study – person
  - “researcher”
- **sing-haak**: travel – guest
  - “passenger”
- **tai-fun-kaa**: take – money – card
  - “ATM card”
- **gei-si-bou**: record – matter – book
  - “notebook”
- **long-saam-gaa**: airing – clothes – rack
  - “clothes-horse”

d. Verb-object Compounds

- **gwo-sou**: pass – sum
  - “to transfer money”
- **sai-coeng**: wash – intestine
  - “colon hydrotherapy”
- **ji-tou**: cure – belly
  - “to eat”
- **saai-meng**: shine – life
  - “to show off”
- **tung-deng**: unblock – roof
  - “to burn the midnight oil”
However, unlike Mandarin Chinese compounds which are much better studied (Chao 1968, H.-M. Liu 1986, Y.-F. Li 1990, S.-X. Liu 1990, Chang 1991, Jin 1991, Cheng 1992, Dai 1992, Gu 1992, S.-F. Huang 1998, Shi 1998, Gu and Shen 2001, Yang 2001), Cantonese compounds have not drawn much attention from linguists. Take the subject of this research: verb-object compounds as an example. There is not much discussion on this type of compound in Cheung’s (1972) study on Cantonese grammar in Hong Kong. He has only devoted a small section describing the phenomenon which he calls the ionization in which the verb and the object of a verb-object compound are separated. In Matthews and Yip’s (1994) grammar of Cantonese, a useful summary of the properties exhibited by verb-object compounds is offered. So far there is no theoretical analysis on this type of compound. Hence, in the present study we are going to examine verb-object compounds in Cantonese in the hope of finding out more salient features of the Cantonese compounds and in turn understanding more about the nature of the morphological component. The reason for choosing Cantonese verb-object compounds as the research subject is that these compounds demonstrate a wide range of lexical and phrasal properties which can help illuminate our understanding of the nature of compounds. We have given some examples of Cantonese verb-object compounds (VOCs) in (17d) already. Additional examples are provided in (18) below. A detailed review of this type of compound will be offered in Chapter Three.
<table>
<thead>
<tr>
<th>Verb</th>
<th>Action/Instrument</th>
<th>Translation</th>
</tr>
</thead>
<tbody>
<tr>
<td>maai-daan</td>
<td>bury – bill</td>
<td>“to bill”</td>
</tr>
<tr>
<td>zau-tong</td>
<td>run – class</td>
<td>“to skip class”</td>
</tr>
<tr>
<td>maai-fau</td>
<td>sell – port</td>
<td>“to sell things to foreign countries”</td>
</tr>
<tr>
<td>aau-caai</td>
<td>twist – firewood</td>
<td>“to twist one’s ankle”</td>
</tr>
<tr>
<td>jam-caa</td>
<td>drink – tea</td>
<td>“to go to a restaurant”</td>
</tr>
<tr>
<td>caau-ce</td>
<td>fry – car</td>
<td>“to crash one’s car in an accident”</td>
</tr>
<tr>
<td>cong-baan</td>
<td>hit – board</td>
<td>“to make a mistake”</td>
</tr>
<tr>
<td>tok-saucang</td>
<td>support – elbow</td>
<td>“to refuse”</td>
</tr>
<tr>
<td>hok-ce</td>
<td>learn – car</td>
<td>“to learn driving”</td>
</tr>
<tr>
<td>gam/gei</td>
<td>press – machine</td>
<td>“to get money from ATM”</td>
</tr>
<tr>
<td>dou-daan</td>
<td>pour – egg</td>
<td>“to spoil (a party)”</td>
</tr>
<tr>
<td>pek-paaau</td>
<td>chop – cannon</td>
<td>“to quit”</td>
</tr>
<tr>
<td>pek-zau</td>
<td>chop – wine</td>
<td>“to drink a lot”</td>
</tr>
<tr>
<td>pek-jau</td>
<td>chop – friend</td>
<td>“to fight with weapons”</td>
</tr>
<tr>
<td>suk-bei</td>
<td>shrink – skin</td>
<td>“to reduce the budget”</td>
</tr>
<tr>
<td>juk-sau</td>
<td>move – hand</td>
<td>“to fight with somebody”</td>
</tr>
<tr>
<td>ceoi-gai</td>
<td>blow – chicken</td>
<td>“to seek help”</td>
</tr>
<tr>
<td>to-maa</td>
<td>drag – horse</td>
<td>“to seek help”</td>
</tr>
<tr>
<td>lo-syun</td>
<td>take – ship</td>
<td>“to get a scholarship”</td>
</tr>
<tr>
<td>siu-zi</td>
<td>burn – fat</td>
<td>“to keep fit”</td>
</tr>
<tr>
<td>zong-daangung</td>
<td>install-spring</td>
<td>“to set somebody up”</td>
</tr>
</tbody>
</table>
This thesis aims at exploring the nature of the compounding process by examining Cantonese VOCs like those in (17d) and (18). The focus of the study is on the dual nature of these compounds and the status of morphology. Specifically, we attempt to answer the following research questions:

1/ What are the characteristics of VOCs in Cantonese?
2/ Why do Cantonese VOCs exhibit both lexical and phrasal properties?
3/ What is the status of morphology in the grammatical system?
4/ What is the relation between morphology and syntax?

The descriptive aspect of Cantonese VOCs (research question #1) will be dealt with in detail. We will characterize the morphological and the syntactic behavior of this particular type of VOC. After that we will move on to exploring relevant theoretical aspects. To seek for an answer to research question #2, we will propose our analysis on the nature and formation of this type of compound. Basing on our analysis, we will then present our view on the nature of morphology and its relation to syntax as an attempt to answer research questions #3 and #4.
1.4 Summary and Organization of the Thesis

In this chapter the main morphological and syntactic properties of compounds have been reviewed. The study of the process of compounding is of linguistic interest because compounds, the products of word formation, exhibit both lexical and phrasal properties. This salient feature of compounds has raised theoretical issues in the following aspects: the autonomy of the morphological component, the interaction of morphology and syntax, and the existence of interface(s) between different modules of the grammatical system. The present study is an attempt to resolve these theoretical issues.

The organization of this thesis is as follows. In Chapter Two we begin with an examination of the notion of word. Since the definition of compounds relies heavily on the notion of word, an understanding of the nature of word is a prerequisite. It is demonstrated that although many problems center around the definition of word, the Lexical Integrity Hypothesis (LIH) can be used as a primary criterion to define wordhood. Chapter Three discusses the properties of verb-object compounds (VOCs) in Cantonese. We look at different analyses which try to explain the nature of verb-object compounds. It is shown that existing analyses cannot adequately account for the behavior of VOCs. We then propose that Cantonese VOCs should be classified into types: one type being inseparable and the other type separable. In Chapter Four, issues concerning the nature of the separable and inseparable Cantonese VOCs and their formation processes are dealt with. In particular, we show that the difference in relation to
the syntactic properties of the two types of Cantonese VOCs can be accounted for if we assume that they are formed at different interface levels. Chapter Five concludes the thesis with a brief summary of the proposed analysis.
Chapter Two  Notions of Word, Compound and Phrase

2.0 Introduction

The focus of the present study is verb-object compounds in Cantonese. Before we discuss the properties of this particular type of compound, we need to explore a more basic concept: “word”¹ for the very reason that in the literature compounds are generally argued to be words (i.e. lexical items) or formed from words. There is massive literature on the question of how to define wordhood (Bloomfield 1933, Di Sciullo and Williams 1987, Baker 1988b, Zwicky 1990b, Matthews 1991, Tsui 1999, Packard 2000, Feng 2000, 2001, among many others). While it is not our purpose to reconcile the various notions, we shall need a working definition for word. There are two aims for this chapter. First, we examine different criteria (e.g. morphological, syntactic, semantic and phonological criteria) used to define the notion of word. Second, we evaluate accounts which are concerned with the distinction between phrases and compounds. In particular, we focus on a principle which is assumed in many morphosyntactic studies, namely the Lexical Integrity Hypothesis (LIH). We deal with the definition of compound at the end of this chapter.

¹
2.1 Criteria of Wordhood

The notion of word has been characterized from a variety of perspectives: morphology, syntax, phonology, semantics and so on (Anderson 1992, Packard 2000). The results along these various lines of research do not always coincide. That explains why the definition of wordhood has been the subject of much controversy. This section does not aim to reconcile the various notions of word. Rather it discusses some widely-cited criteria in the literature.

Criteria used to define what a word (lexical item) is can be sought at different levels: morphology, semantics, phonology as well as syntax. A morphological definition of word can be used for highly inflecting languages (Anderson 1985, Matthews 1991). Word boundaries in highly inflecting languages can be determined by locating obligatory inflectional morphemes which close up a word. In Latin, for instance, inflectional morphemes always close up the word. Take *bono:s libro:s* “good books” as an example. Here two morphological words can be defined. *bono:s* consists of the base *bon* and an ending of case, number and gender of adjectives. *libro:s* consists of the base *libr* and an ending of case and number (Dai 1992). Another morphological feature of words is the fixed ordering of the constituents (Matthews 1991, Mithun and Corbett 1999, Cruse 2001). We can see that it is a cross-linguistic phenomenon that the order of stem(s) and inflection(s) of a word is fixed. Mithun and Corbett (1999) note that even in languages with some freedom of word order such as Mohawk, the morpheme order within a word cannot be rearranged.
Morphological criteria may not be that decisive when one considers languages with little or no inflection. Chinese, for example, is argued to have a meager Infl (C.-T. Huang 1982). So inflection in general cannot be used to determine the word boundaries of this language.\

Those who study the semantics of words regard anaphoric islandhood as a general property of words (Postal 1969, Sproat 1988, Wang 1994, 1998). It is argued that words are anaphoric islands and tend to be referentially opaque, in that it is impossible to refer to their parts. We have already seen some examples in Chapter One. Specifically, no parts of a morphological formation may be used as an antecedent for anaphoric co-indexation. The difference of anaphoricity between a word and a phrase can be illustrated by the examples in (1) which are taken from Spencer (1991:42). In (1a) tea is part of a phrase while in (1b) tea is part of a word:

(1)  
\begin{enumerate}  
\item a. a pound of tea  
\item b. a teapot  
\end{enumerate}

As Spencer notes, in (1a) tea refers to a particular kind of stuff (i.e. tea) and the meaning of the whole expression is determined compositionally, i.e. (1a) refers to one pound of something and that something must be tea. But for (1b), there is no sense in which the constituent tea in the word teapot actually makes reference to the stuff tea in determining the reference of the whole expression.
Consequently, one can refer back to the *tea* in (1a) using a pronoun as in (2a).

But this is impossible in the case of *teapot* as shown in (2b):

(2) a. He took the pound of tea, and put two spoonfuls of it, into a teapot.

b. ?He took the teapot and poured it into the cup.

In (2a) *it* refers to *tea*. This is not the case with (2b). We cannot say that (2b) means “He poured the tea into the cup.” without any preceding context.

Another semantic property of word is that semantic interpretation rules cannot see a subpart of a word. This can be exemplified by the expressions in (3) below:

(3) a. jat faai luksik ge hakbaan
   one CL green GE black-board
   “a green blackboard”

b. a brown blackbird

Although the words *hakbaan* “blackboard” and *blackbird* contain a constituent which is a color term meaning black, they are compatible with a color adjective which does not mean black. The reason is that in each case the adjective modifies the whole compound noun regardless of its internal structure. So we will not interpret (3a) *jat faai luksik ge hakbaan* “a green blackboard” as a black
board which is green nor (3b) a brown blackbird as a black bird which is brown.

When one focuses solely on the semantics of the structures in question, there can be indeterminacy between words and phrases. Consider idioms for instance. English phrases like pull somebody's leg “to make a joke by telling someone something that is not actually true”, live from hand to mouth “precariously, spending money as soon as it is received” and have one's heart in one’s mouth “be badly frightened” are called idioms because their meanings are idiosyncratic and not compositional. In this respect, they resemble words in that they exhibit idiomaticity of meaning. However, unlike words which are treated by syntax as single lexical units, these idioms must be analyzed syntactically, because they observe syntactic rules (Matthews 1991). For instance, the verb in each idiom above must be able to inflect for tense. And for the idiom have one's heart in one's mouth, there must be agreement between the subject of the possessor of the heart and the possessor of the mouth, as shown by the contrast in (4) below:

(4) a. John has his heart in his mouth.
   b. * Mary has her heart in my mouth.

(4a) is grammatical, because his month refers to the month of John (who is the possessor of the heart). In (4b), however, the possessor of the month (i.e. I) is not the same with the possessor of the heart (i.e. Mary), so the sentence is
ungrammatical. In a nutshell, the distinction between words and idioms will not be clear, if one only relies on the semantic criterion: idiomaticity of meaning.

Phonological criteria are often used for the determination of word boundaries. Matthews (1991) points out that words tend to be a unit of phonology as well as grammar. He uses several languages to illustrate this characteristic of words. He notes that, for example, in Egyptian Arabic, a word serves as the domain for stress. A word like *kitáb* "book" is accented finally because the final syllable is long while words like *kátáb* "he wrote" and *kátábib* "clerk" are accented initially because the final syllable is short. Another piece of evidence showing a word is a phonological unit is that word boundaries in some languages are marked by phonological phenomena such as vowel harmony, stress or tone sandhi. These criteria give us a notion of phonological word. In English the phonological rule of devoicing assimilation applies within a word (Dai 1992). For instance, *cat's* is one phonological word, as evidenced by the applicability of the sandhi rule: /s/ in *cat's* is voiceless, as opposed to /s/ in *cat is*, which is voiced. Note that there is a discrepancy between criteria drawing from different levels: a phonological word *cat's* corresponds to *cat is* which consist of two morphological words.

Finally, we turn to criteria drawn from syntax. An often-cited syntactic definition of word, which is proposed by Bloomfield (1933) is that words are minimal free forms. By minimal free form, Bloomfield means it is the smallest unit (i.e. cannot be divided further) which can appear on its own. For instance,
bona fide can be divided into bona and fide, so it should be a phrase according to Bloomfield’s definition. But bona and fide cannot be further divided and so each of them is a minimal free form, i.e. a word (lexical item). As Matthews (1991) notes, this definition of word has often been criticized. For one thing, some free forms are not likely to occur alone. For instance, prepositions in both English (e.g. on, at, in) and Cantonese (e.g. hoeng “towards”) normally do not appear on their own. Moreover, some forms which are not free forms can occur alone. It is possible for one to imagine a scenario in which parts of words can appear alone (the example below is taken from Matthews 1991:210-211):

(5) A: Did you say revise or devise?

B: Re.

Re is not a word, however, it can appear alone as an answer to a question as shown in (5). One can see that the idea that words as minimal free forms seems intuitively obvious, but the criterion is not precise enough.

Summing up the above observations of the properties of words, we find that there is no single universal criterion that can be used to define the notion of word. But each of the criteria we have discussed does reflect a specific property of lexical items. In the next section we will review three important works on Chinese grammar. We will see how these studies utilize different properties of lexical items to distinguish compounds from phrases.
2.2 Distinction between Compounds and Phrases

It has been noted in the literature that the characterization of the nature of compound, and the distinction between compounds and phrases are very difficult (Li and Thompson 1981, Matthews and Yip 1994). There are two important factors at work here. For one thing, compounds involve the notion of word. A compound is generally argued to be a lexical item just like a word (Quirk et al 1985, Matthews 1991, Spencer 1991, Tang 1991, Matthews and Yip 1994, Payne 1997, among others). Many linguists further argue that compounds are formed from words (Selkirk 1982, Spencer 1991, 2001, Fabb 1998). However, there is no consensus on the very nature of word itself. That means when one attempts to characterize the nature of compound, one will inevitably make reference to a rather fuzzy definition of word. The second factor is the dual nature of compound. Compounds resemble both words and phrases in that they demonstrate lexical as well as syntactic properties (Matthews 1991, Spencer 1991, Carstairs-McCarthy 1992, Fabb 1998, Ackema 1999). In other words, when one tries to distinguish a compound from a phrase, one is dealing with two items which both show syntactic properties. Consequently, the distinction between compounds and phrases becomes much more difficult.

In this section, we focus on studies which are concerned with this kind of difficult tasks. The starting point is a review of the work of Matthews (1991), for he provides a very general and useful discussion on the criteria of
compoundhood. Then we evaluate analyses which deal with the distinction between compounds and phrases.

2.2.1 Introduction

Matthews (1991) points out that orthography is not a reliable guide for distinguishing a compound from a phrase. This can be seen from the example girl friend. This compound noun can be written in any of three ways: as two words: girl friend, as one word hyphenated: girl-friend or unhyphenated: girlfriend. The indeterminacy of orthography in the identification of the boundary between compounds and phrases leads one to employ other criteria.

First, consider morphological criteria. Matthews mentions that many linguists use the inflection test to check whether a string is a compound or not. Specifically, if the first constituent element of the item in question inflects as a separate unit, then the item is not a compound. For instance, heir apparent is not a compound, because heir itself is still inflected in the plural form heirs apparent. Another similar example is commander-in-chief. The word commander will be inflected to indicate plurality as in commanders-in-chief. So one will not regard commander-in-chief as a compound. Morphological criteria may not be that useful in determining word boundaries in languages with little or no inflection such as Chinese (C.-T. Huang 1982, Gu 1995) and Cantonese (Chin 2001).6

The unpredictability of meaning of compounds is often used as a semantic
criterion of compoundhood. Consider the phrase *black bird* and the compound *blackbird*. The meaning of *black bird* is compositional. Any bird which is black is a *black bird*. On the other hand, the meaning of the compound *blackbird* is not predictable. It is a particular species of bird. Female blackbirds are, in fact, brown in color. One can see that unpredictability of meaning is able to distinguish compounds from phrases. But this semantic criterion has its problems. The main problem lies in the fact that the meaning of idiom phrases, just like the case in compounds, is not transparent. English phrases *kick the bucket* means “die” and *skeleton in the closet* means “something shameful and kept secret” are called idioms, because they have idiosyncratic (i.e. noncompositional) meanings. In short, unpredictability or idiomaticity of meaning is not a compound-specific (or word-specific) property.

Phonological criteria can also be employed in judging whether a particular string is a compound or a phrase. Matthews notes that compounds in English are stressed on their first constituent. For example, the compound *blackbird* has its stress on *black* (as opposed to the phrase *black bird* which is stressed on *bird*). Another example is *mental hospital*. The string is written as two words, but by its stress pattern, one can tell that it is a compound because the string has a primary stress on *men-* , not *hos-*. Stress pattern seems to be a useful tool in distinguishing a compound from a phrase, however, Matthews points out a problem tied up with this phonological criterion. Some exocentric compounds such as *red admiral* “a type of butterfly that has black wings with bright red
marks on them” have phrase-like stress. In other words, phonologically we say a particular string, for example, *red admiral* or *black belt*, is a phrase because it is stressed like a phrase but semantically we know that it is not a phrase because it contains no head.8

As for the syntactic criterion, Matthews states that the obvious test for compounds is the separation test. Specifically, a compound should not be separable. He uses this criterion to explain why phrasal verb are not regarded as compounds. Consider examples in (6) below:

(6) a. I put out the fire.
   
b. I put the fire out.

In (6a), *put out* means “extinguish” and this meaning is not predictable from that of *put* and *out*. One may want to say *put out* could be a compound. But consider also (6b). The meaning of (6b) is the same as (6a) and the meaning “extinguish” also comes from *put out*. But in (6b), *put* and *out* are separated by the fire. Therefore, he argues that although *put out* has its meaning corresponding to a lexical word, it is not a compound due to the separability it demonstrates. A related phenomenon is reduction in coordination (conjunction). When two compounds are conjoined, one should not be able to delete part of a compound as coordination is a syntactic operation, as exemplified in (7) below:
From (7), one can see that it is unacceptable to conjoin the compounds *sickbeds and deathbeds* with a deletion of the common part “beds”: *sick and deathbeds* or *death and sickbeds*. However, Matthews notes that conjunction reduction does occur with some compounds. For example, *macro-economic studies* and *micro-economic studies* can be conjoined to form *macro- and micro-economic studies*. This demonstrates that syntactic criteria may not be decisive in certain situations.

Matthews concludes that no criterion mentioned above is irrelevant in the distinction between compounds and phrases. Each of them reflects certain properties of compounds. But no single criterion can be used to define compoundhood. To have a better view on what relevant criteria can be used to define compounds, we are going to review studies which discuss the distinction between compounds and phrases. Since there is no in-depth discussion on the compound-phrase distinction in those important studies on Cantonese grammar: Chao (1947), Cheung (1972) and Matthews and Yip (1994). We will shift our focus to analyses on Mandarin Chinese, a language having a close relationship to Cantonese. In sections 2.2.2 to 2.2.4 below, we review three important studies on
Chinese grammar, focusing on their discussion on the nature and the identification of compound.

2.2.2 Syntactic Aspect

In Chao’s (1968) classic study on Chinese grammar, he discusses various dimensions in which compounds may be classified. The following list (taken from C.-T. Huang 1984) is a summary of the criteria for the identification of compounds proposed in his study:

(8) a. Part of the item is a bound form.

b. Part of it is neutral-toned.

c. The meaning of the whole is not compositional of its parts.

d. The internal structure is exocentric.

e. The parts are inseparable from each other.

According to Chao’s analysis, an item is identified as a compound if one or more of the above criteria are met. The list in (8) has been criticized by different linguists (Chi 1984, C.-T. Huang 1984, H.-M. Liu 1986, Packard 2000 who largely bases on C.-T. Huang 1984). We will focus on the theoretical objections to these criteria offered in C.-T. Huang’s (1984) study. With the exception of (8c), C.-T. Huang reduces all the criteria in (8) to one single criterion which
serves to distinguish the domain of syntax from that of morphology. The criterion is Lexical Integrity Hypothesis (LIH) which will be discussed in detail in section 2.3.

The gist of this criterion is that syntactic rules cannot affect the internal structure of a morphological item. Consider (8a) the bound form criterion first. C.-T. Huang points out that neutral-toned elements are usually bound forms, so (8b) can be reduced to (8a). Since a bound form by definition cannot stand alone, (8a) in turn can be reduced to (8e), the criterion of inseparability. The property of inseparability itself is, in fact, a reflection of the lexical integrity of lexical items. Consequently, Chao's criteria on compounds: (8a), (8b) and (8e) can be reduced to the LIH.

C.-T. Huang further shows that the exocentric criterion, i.e. (8d) proposed by Chao (1968) also falls under the LIH. According to Chao (1968), an item is identified as a compound if the internal structure is exocentric (i.e. without a head). This criterion does not mean that a compound cannot have a head. What it suggests is that since a phrase must have a head, an item without a head should be lexical rather than phrasal. So C.-T. Huang argues that (8d) can be captured by the LIH. He uses bang-tui\(^{10}\) “legging” as an example. This compound noun is exocentric in that the nominal constituent of this verb-object combination is not the head. If this item were a phrase, it would violate the endocentricity constraint which requires a phrase to be headed by a unique head. But if bang-tui is regarded as a compound, this string will then be treated as a single lexical unit in
terms of syntax. Its exocentric structure will no longer be relevant to the endocentricity constraint, because syntax will be inaccessible to its internal structure. Here one see that the criterion stated in (8d) relies essentially on the lexical integrity of lexical items. Therefore, (8d) can be reduced to the LIH.

Now consider (8c) the criterion of semantic noncompositionality or idiomaticity. C.-T. Huang thinks that this is not a sufficient condition for compoundhood and therefore should be excluded, because there are a lot of phrases whose meanings are idiomatic. For example, the meanings of the following idioms are noncompositional: in English *put one’s best foot forward* “get on with one’s work”, in Chinese *peng yi bizi hui* “bump one nose dust = to be let down”, in Cantonese *baan zyu sik loufu* “disguise as a pig and then eat a tiger = to be a wolf in sheep’s clothing” and *zaa geng zau meng* “squeeze one’s neck to suit one’s life = to swallow an insult”. If one only bases on (8c) to distinguish compounds from phrases, one will wrongly include these idiom phrases as compounds.

2.2.3 Semantic Aspect

In Li and Thompson’s (1981) grammar of Chinese, they explore the characteristics of different compounds at length, for instance, nominal compounds, resultative verb compounds, verb-object compounds and so on. Their study focuses on the semantic properties exhibited by Chinese compounds.
This section reviews their arguments which are related to the present study.

Li and Thompson realize that there is no clear-cut distinction between compounds and noncompounds. They (1981:45) point out that the difficulty in defining compound has much to do with the semantics of compounds. Their arguments can be summarized as follows:

\[(9)\]

- a component morpheme may come from classical Chinese and no longer functions as a free morpheme in modern spoken Mandarin
- the meanings of the component morphemes may not be related to the meaning of the entire word
- the meaning of a polysyllabic word may only indirectly be connected with the literal meanings of the component morphemes
- the meaning of a polysyllabic word may only be metaphorically related to the meanings of its component morphemes

In view of the indeterminacy concerning the nature of the component morphemes, their study does not attempt to find a precise definition of compound. So instead of providing a list of criteria of compoundhood, they only give a loose definition of compound in general and then move on to characterize the semantic properties of compounds. The definition of Chinese compound proposed by Li and Thompson (1981:46) is given below:
We may consider as compounds all polysyllabic units that have certain properties of single words and that can be analyzed into two or more meaningful elements, or morphemes, even if these morphemes cannot occur independently in modern Mandarin.

They state that according to the characterization above, *kai-guan* “open-close = switch” and *chou-yan* “extract-smoke = smoke” should be regarded as compounds and *he tang* “drink soap” a phrase. But they do not have further elaboration to justify such a distinction.

Now we turn to the characterization of semantic properties of compounds offered by Li and Thompson. They highlight that the relatedness between the meaning of a compound and those of its constituent elements can vary from being close to nonexistent. While many studies take it for granted that a compound must have an idiomaticity of meaning, Li and Thompson note that compounds exhibit various degrees of idiomaticity. Specifically, the meaning of compounds can vary from idiomatic to compositional. They classify compounds into three general types according to the connection between the meaning of the compound and the meaning of its component parts. Consider the examples given by Li and Thompson (1981:47):
Type (1) No apparent semantic connection between the meaning of the compound and the meaning of its constituents

<table>
<thead>
<tr>
<th>Compound</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>fen-liu</td>
<td>wind – flow = amorous</td>
</tr>
<tr>
<td>hua-sheng</td>
<td>flower – born = peanut</td>
</tr>
<tr>
<td>fei-zao</td>
<td>fat – black = soap</td>
</tr>
</tbody>
</table>

Type (2) A metaphorical, figurative or inferential connection between the meaning of the compound and the meaning of its constituents

<table>
<thead>
<tr>
<th>Compound</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>mao-dun</td>
<td>spear – shield = contradictory</td>
</tr>
<tr>
<td>re-xin</td>
<td>hot – heart = enthusiastic</td>
</tr>
<tr>
<td>huo-chai</td>
<td>fire – firewood = match</td>
</tr>
</tbody>
</table>

Type (3) Meaning of compound directly related or identical to the meaning of its components.

<table>
<thead>
<tr>
<th>Compound</th>
<th>Meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>xi-zao</td>
<td>wash – bath = take a bath</td>
</tr>
<tr>
<td>gan-jing</td>
<td>dry – clean = clean</td>
</tr>
<tr>
<td>jin-bu</td>
<td>advance – step = make progress</td>
</tr>
</tbody>
</table>

The chart in (11) shows that the degree of relatedness between the meaning of a compound and the meaning of its constituents varies from close (Type 3) to nonexistent (Type 1). This observation is important for a proper characterization of the nature of compound. It draws our attention to the fact that even when an item does not have an idiosyncratic meaning, it can still behave like a compound in other aspects.

2.2.4 Morphological and Phonological Aspects

In Zhu’s (1982) study, four criteria for what makes a compound are discussed. They are (i) presence of bound morpheme, (ii) neutral-toned second syllable, (iii) unexpandability of elements and (iv) noncompositionality of meaning. These criteria are very similar to those proposed by Chao (1968). Since
Chao's analysis has already been reviewed in section 2.2.2, we will not discuss the four criteria in Zhu's study in detail. Instead we will focus only on the first two criteria, because Zhu has more elaboration on them.

First, Zhu states that in general an item that contains a bound morpheme is a compound. This is shown in the examples (12a) to (12c) given by Zhu below:

\[(12)\]
\[
\begin{align*}
\text{a. BB: } & \text{zhi-wu “plant”} & \text{yu-shi “bathroom”} \\
\text{b. BF: } & \text{shi-you “oil”} & \text{wei-da “great”} \\
\text{c. FB: } & \text{dian-shi “TV set”} & \text{da-yi “jacket”} \\
\text{d. FF: } & \text{tie-lu “railway”} & \text{bai-cai “cabbage”}
\end{align*}
\]

where \(B\): bound morpheme
\(F\): free morpheme

According to Zhu, the items in (12a) to (12c) are compounds because each of them contains at least one bound morpheme. In (12a), both constituents are bound morphemes; in (12b), the first element is bound and in (12c), the second element is bound. From (12a) to (12c), it is shown that the bound morpheme criterion is a sufficient condition for the identification of compounds. However, Zhu himself notes that there are exceptions to this criterion. First, one is able to see that the examples in (12d) are compounds also, but they consist of free morphemes only. Also, a structure containing a bound morpheme may not necessarily be a compound but can be a syntactic phrase, for instance, \(ye hao “It sounds great”\). In this example, \(ye “also”\) is a bound morpheme. But this expression is a phrase, not a compound.
Another criterion of compound proposed by Zhu is that the second syllable of a compound is often in neutral tone. Zhu gives a comparison of Chinese compounds and phrases by using pairs of identical strings:

(13) Compounds Phrases
a. māi-māi “business” māi mǎi “buy and sell”
b. dōng-xī “thing” dōng xī “East and West”
c. huǒ-shǎo “a kind of food” huǒ shǎo “burn something with fire”
d. dā-shǒu “guard” dā shǒu “hit somebody’s hand”

Under the column of “Compounds” in (13), each item has its second syllable neutral-toned but that is not the case under the “Phrases” column. The phonological difference between the items under the two columns reflects that the neutral-tone criterion on compound proposed by Zhu is descriptively adequate. But similar to other criteria, there are exceptions to this neutral-toned criterion. In structures like zou ba “Let’s go” and wo de “(It’s) mine” in Chinese, the second syllable is in neutral tone. But they are phrasal in nature.

2.2.5 Summary

In sections 2.2.2 to 2.2.4 we have reviewed different approaches to define compounds. Different linguists focus on different properties of a compound. C.-T.
Huang has successfully captured Chao’s criteria under the LIH. Li and Thompson highlight that the meaning of compounds can vary from idiomatic to compositional. This is contrary to the traditional view that lexical items must have idiomatic meanings while phrasal items are compositional in terms of meaning. Zhu probes into the boundness of the constituents of compounds. Traditionally, compounds are argued to be formed from words (free morphemes). Chao and Li and Thompson state that compounds can be formed from bound morphemes. Zhu shows that the bound morpheme is only a sufficient but not necessary condition for the identification of compounds.

The distinction between a compound and a phrase is a long standing issue in the study of Chinese languages (Cantonese being a closest relative to the family). The general review given in previous sections shows that there is no single criterion that can be used for the identification of compounds of Chinese languages (including Cantonese). However, that does not mean that study of the nature of the process of compounding is completely impossible. In a discussion of the distinction between compounds and noncompounds, Matthews (1991:100) concludes that “a grammarian will have to work with varying and partly fluid criteria.” What Matthews suggests is that the difficulty in defining the notion of a word (lexical item) and distinguishing compounds from phrases cannot stop us from studying the nature of word and compound. It only means that we need to formulate relatively flexible definitions for words and compounds. In section 2.3 below, we will focus on a salient feature of lexical items—lexical integrity and a
related hypothesis on wordhood which underlies much work of morphosyntactic studies, namely the Lexical Integrity Hypothesis, the LIH. The working definitions of compound we are going to propose in section 2.4 will rely heavily on this Hypothesis.

2.3 The Lexical Integrity Hypothesis (LIH)

In the previous sections, various analyses on the nature of word and compound have been reviewed. Basically, we have been concerned ourselves with the general properties of lexical items. In this section, we will examine a hypothesis which is related to a specific property of lexical items and is generally assumed in morphosyntactic studies, namely the Lexical Integrity Hypothesis (LIH). But before we go into the details of the Hypothesis, we need to have a general idea about what lexical integrity is. Let's consider the illustration given by Spencer (2001). He takes the form *cats* as an example. *Cats* comprises the root morpheme "cat" and a suffix morpheme "s" indicating plurality. We know that the "s" of *cats* is not a word in its own right, because unlike a phrase such as *the cat* which can be expanded by inserting other phrases in between: *the very black cat*, the form *cats* can never be split up this way. The "s" component cannot stand alone and must only exist as part of a word. Spencer remarks that the property of indivisibility exhibited by *cats* is one of lexical integrity.

As we noted in section 1.2.1 on the lexicalist approach to the relation
between morphology and syntax, there are various formulations regarding the notion of lexical integrity over the years. For the purpose of this study, the LIH will be stated as (14) below:

(14) The LIH

Syntactic rules cannot make reference to any part of the internal structure of a lexical item.

We will show in the rest of this section that there is a cluster of facts that may be made to follow from the LIH.

First, consider conjunction reduction in coordination structures involving a subpart of a word. The Cantonese examples in (15) and (16) below involve conjunction reduction. In (15b), the reduction involves a part of a word while (16b) contains a legitimate reduction:

(15) a. Ngo zungji waan dince tung foce.  
I like play electric-car and fire-car  
"I love to play trams and trains."

b. * Ngo zungji waan din tung foce.  
I like play electric and fire-car
(16) a. Ngo zungji waan gulonge dince tung gulonge foce.
I like play old electric-car and old fire-car
“I love to play old trams and old trains.”
b. Ngo zungji waan gulonge dince tung foce.
I like play old electric-car and fire-car
“I love to play old trams and trains.”

The contrast between the grammaticality of the conjunction reduction in (15b) and (16b) can be explained by the LIH proposed in (14). In (15b), the coordination involves the deletion of ce “car” which is part of the word dince “tram”. Since the LIH requires that no syntactic rules can affect the internal structure of a lexical item, the sentence is excluded by the LIH. In (16b), on the other hand, the deletion only involves the phrase gulonge “old”, but does not affect the internal structure of any word it modifies. The conjunction deletion in (16b) does not violate the LIH; therefore, it is legitimate.

Second, we consider a widely discussed property of word, namely the inseparability (also known as cohesiveness, unexpandability or indivisibility). Specifically, nothing can be inserted into the internal structure of a word except infixes. Consider the Cantonese examples below:
(17) a. maa-faan

   adjective: troublesome
   noun: trouble or problem
   verb: to trouble somebody

b. maa-gwai-faan (where gwai is an infix)

   “downright troublesome”

c. *maa nei faan (where nei is a word meaning “you”)

   Intended meaning: “to trouble you” as in maa-faan nei

d. *maa houdaaige faan (where houdaaige is a phrase meaning “very big”)

   Intended meaning: “a big trouble” as in houdaaige maa-faan

(17a) maa-faan is a Cantonese word meaning “troublesome”. (17b) involves an insertion of a Cantonese infix gwai. For (17c) and (17d), a word and a phrase are inserted into the word maa-faan respectively. (17b) is grammatical whereas (17c) and (17d) are not. According to the definition given in (14), one expects that the LIH will rule out (17c) and (17d), because both expressions are inserted with syntactic constituents. This prediction is borne out. The LIH correctly rejects (17c) and (17d) as they involve an insertion of items which are not infixes into the internal structure of a word.

Third, words are islands for movement (Wang 1994, 1998). Specifically, nothing can be moved out of a word. The movement of the nominal component of the verb-object sequence in (18) is grammatical but it is not allowed in the
case of (19). This is shown by the contrast in the grammaticality in (18b) and (19b) below:

(18) a. Keisat m jatdeng jiu *maai lau* gaa!
    in fact not must need buy flat PRT
    “In fact, it is not necessary to buy a flat.”

   b. *Lau* keisat m jatdeng jiu *maai* gaa!
      flat in fact not must need buy PRT
      “As far as flat is concerned, it is not necessary to buy one.”

(19) a. Keisat m jatdeng jiu *ceotgaa* gaa!
    in fact not must need go out-home PRT
    “In fact, it is not necessary to become a Buddhist monk.”

   b. *Gaa* keisat m jatdeng jiu *ceot* gaa!
      home in fact not must need go out PRT

Consider the sentences in (18) first. (18a) contains a verb-object sequence *maai lau* “buying flat” which is not a word. This can be shown by the examples in (20) in which the sequence is expanded to become larger units:
(20) a. Maai faan gaan lau.

buy ASP CL flat

“To buy a flat.”

b. Maai gaan leng lau.

buy CL nice flat

“To buy a nice flat.”

As shown in (20a) one can readily expand the string *maai lau* “buying flat” to form *Maai faan gaan lau*. “To buy a flat.” where the verb *maai* and the object *lau* is separated by an aspect marker *faan* and a classifier *gaan*. (20b) indicates that the combination can also be expanded to *Maai gaan leng lau*. “To buy a nice flat.” where a classifier *gaan* and an adjective *leng* intervene between the verb *maai* and the object *lau*. These facts show that the string *maai lau* cannot be a word. It must be a phrase, so topicalizing the nominal constituent *lau* “flat” as in (18b) does not violate the LIH. Therefore, the sentence is grammatical.

In contrast to the legitimate topicalization in (18), topicalizing the nominal constituent *gaa* “home” out of the verb-object sequence *ceotgaa* “to become a Buddhist monk” is not allowed, as shown in (19b). The ungrammaticality of the sentence can be attributed to the violation of the LIH. Specifically, the verb-object sequence *ceotgaa* should be a compound, not a phrase and the moved element *gaa* is part of the lexical item. Note that although *gaa* can be used as a word in certain contexts, the constituent *gaa* within the sequence
ceotgaa should be part of a word. In other words, there are two different gaa in fact. Examples in (21) and (22) below illustrate this point:

(21) a. nei go gaa
this CL home
"this family"
b. loeng tau gaa
two CL home
"two families"

(22) a. *ceot go gaa
go out CL home
b. *ceot tau gaa
go out CL home

In (21), gaa “home” is modified by its corresponding classifier go and tau. But in (22), one can see that the gaa of the verb-object sequence ceot-gaa “to become a Buddhist monk” cannot be modified by the classifier go or tau. Both *ceot go gaa and *ceot tau gaa are unacceptable. This shows that the gaa in ceot-gaa is not a word but is part of a word. The ill-formedness of (19b) can now be explained straightforwardly. The sentence in (19b) is ungrammatical, because it violates the LIH; it involves the movement (topicalization) of gaa which is part
Fourth, anaphoric rules may not refer to a subpart of a word (Postal 1969, Sproat 1988, Wang 1994, 1998). In other words, words are anaphoric islands. One cannot co-index an item with an element within a word. This is evidenced by the examples in (23) and (24) below:

(23) a. Drivers of trucks, fill them, up with diesel.
   b. *Truck, drivers fill them, up with diesel.

(24) a. Daaizoeng, ge lipjan deoi keoidei, hou caanjan gaa.
   elephant GE hunter to them very cruel PRT
   “Hunters of elephants are very cruel to them (elephants).”
   b. *Daaizoeng, lipjan deoi keoidei, hou caanjan gaa.
   elephant hunter to them very cruel PRT

In (23a) and (24a), trucks and daaizoeng “elephant” are co-indexed with the pronoun them and keoidei “them” respectively. The co-indexation is possible in both sentences, because trucks is part of the phrase Drivers of trucks while daaizoeng is part of the phrase Daaizoeng ge lipjan “hunters of elephants”. The two sentences do not violate the LIH in (14). But the situation is just the opposite in the (b)-numbered sentences. The co-indexation in both (23b) and (24b) involves a part of word. In (23b) the pronoun them is used to refer to Truck
which is a part of the lexical item *Truck drivers*. In (24b) the pronoun *keoidei*
“them” is co-indexed with *Daaizoeng* “elephants” which is a part of the
compound noun *Daaizoeng lipjan* “Elephant hunters”. The ungrammaticality of
(23b) and (24b) can therefore be attributed to the violation of the LIH.\textsuperscript{15}

The facts about the nature of word discussed in this section lend support to
the postulation in (14), the LIH. We have seen that different properties of words,
for instance, conjunction reduction and inseparability of word can be captured by
the LIH. In this study, we will assume the LIH as a primary criterion of
wordhood. To be more specific, the LIH will be taken as a defining criterion of
lexical structure.\textsuperscript{16} In the next section, we will go through several more issues
related to the process of compounding and then we will give our own definitions
of compounds.

\section*{2.4 Further Consideration on the Nature of Compound}

\subsection*{2.4.1 Compounding and Other Combinatory Processes}

In the literature, various studies attempt to explain compounding in terms of
other combinatory processes of language such as incorporation and affixation
(Baker 1988a, Oshita 1994). The present study assumes that compounding
cannot be reduced to other morphological or syntactic processes; it is distinct
from other combinatory processes. First, consider compounding, derivation and
inflection processes. The process of compounding differs from derivation and
inflection in that compounds do not consist of a stem or root and affixes but are formed from two or more stems or roots (Bybee 1985, Anderson 1992). In other words, the constituents in compounds are not a result of combining a lexical item with a grammatical item, but a combination of two or more lexical elements (Bybee 1985). This can be illustrated by the comparison in (25) below:

(25)

<table>
<thead>
<tr>
<th>Inflection</th>
<th>Derivation</th>
<th>Compounding</th>
</tr>
</thead>
<tbody>
<tr>
<td>sing + s</td>
<td>sing + er</td>
<td>singing + contest</td>
</tr>
<tr>
<td>read + s</td>
<td>read + er</td>
<td>reading + group</td>
</tr>
<tr>
<td>perform + s</td>
<td>perform + er</td>
<td>performing + art</td>
</tr>
</tbody>
</table>

The table in (25) shows that the expressions under the columns “Inflection” and “Derivation” consist of a lexical item and a grammatical element (precisely speaking, an affix). But the constituents of compounds under “Compounding” are not grammatical items. The compounds as shown in (25) are formed by a combination of lexical elements (a noun plus another noun in the given examples).

Now let us turn to compounding and incorporation. Some linguists, for instance, Baker (1988a), relate incorporation to the process of compounding, since they think that incorporation resembles compounding in that it also involves combination of two or more lexical elements. For instance, in noun incorporation, the two lexical items: the verb and the noun originally exist as independent words and they combine to form one single lexical element after the
noun is incorporated into the verb. However, Bybee (1985) points out that the semantic domain of the incorporated noun is usually restricted, but this sort of restriction is not found in compounding. Therefore, one cannot equate incorporation with compounding.\textsuperscript{18}

To sum up, compounding cannot be reduced to other morphological or syntactic derivation. It must be treated as an independent word formation process. In the next section, we will look closely into the nature of the constituents of compounds.

2.4.2 Status of the Constituents

The general conception on the nature of compound is that they are formed from words, i.e. free morphemes (Selkirk 1982, Matthews 1991, Spencer 1991, Tang 1991, Payne 1997, Fabb 1998, among others). However, as discussed in section 2.2.2 to 2.2.4, Chao (1968), Li and Thompson (1981) and Zhu (1982) suggest that compounds can contain bound morphemes. It is indeed the case one can observe with respect to Chinese type of languages including Cantonese. In Cantonese and Mandarin Chinese many compounds are not formed purely from words. This can be exemplified by the Cantonese compounds in (26) and (27) below:
(26) *fuk-mou* (where both morphemes are bound)

serve – duty

“to serve”

(27) *zing-gu* (where *gu* is a bound morpheme)

make – poison

“to fool somebody”

According to traditional analyses, the items in (26) and (27) would not be treated as compounds, because they contain bound morphemes. Instead, they will be regarded as derived (complex) words. But from our intuition, (26) and (27) should be compounds given their semantic noncompositionality and nonreferentiality of the nominal constituents. If a definition allows compounds to be formed from bound morphemes, then one will not have to reject the items in (26) and (27) as compounds. Researchers of other languages have also found that compounds can consist of bound morpheme. For instance, Fabb (1998:69) points out that some compounds “can be parsed into an independently attested word plus another morpheme which is not an independently attested word but also does not appear to be an affix”. Examples from English given by him include:
It is important to note that the assumption that compounding can involve bound morphemes will not cause difficulties in distinguishing the process of compounding from other affixation processes which are also concerned with a stem (or a root) and a bound morpheme (affix). The reason is that the productivity of those bound morphemes contained within compounds is much lower than that of real affixes. For instance, the bound morpheme gu “poison” in Cantonese is not productive in forming new words. It can only combine with verb zing “make” to form a verb-object compound zing-gu “to fool somebody” or with waak “trick” to become an adjective gu-waak “cunning”.

2.4.3 Degree of Separability

Chao (1968) notes that verb-object compounds in Mandarin Chinese demonstrate different degrees of expansion (commonly known as “separability” in the current literature). He analyzes the expandability of verb-object
compounds from five different perspectives: (i) solid verb-object compounds (compounds which are never ionized under any circumstances), (ii) compounds admitting of suffixes and complements to the verb, (iii) compounds admitting of modifiers to the object, (iv) compounds allowing inversions and (v) separation of constituents in questions and answers. We will briefly review his observations in order to have a deeper understanding of the constituents of compounds in general, and form a general background for our later discussion of properties of verb-object compounds in particular.

“Solid verb-object compounds” refer to those compounds which are not expandable under any circumstances. Examples include yan-shi “loathe-world = tired of the word” and fen-mian “separate-childbirth = to have parturition”. Many verb-object compounds, however, allow their constituents to be separated by different items. The examples below are taken from Chao (1968:426-429):

(29)

a. Suffix

xu-xian “continue-string = to remarry after the death of the first wife”

xu le xian “has remarried”

b. Modifier of the Object

sheng-shi “save-matter = to save labor”

sheng xudio shi “save a lot of labor”
c. Inversion

*da-pai* “hit-plate = to play mahjong”

*pai dou da wan le* “finished playing mahjong”

d. Question and Answer

*da-zi* “hit-character = to type”

*Da bu da zi* “Do typing?” *Da* “Yes.”

This section has shown that the constituents of some verb-object compounds in Mandarin Chinese can be separated under different situations. The observations that compounds demonstrate different degrees of separability lead us to rethink the nature of compounds. More concretely, we assume that compounds should be classified into two groups: separable and inseparable. The definitions of the two groups of compounds will be given in the next section.

2.4.4 Definitions of Compounds

In section 2.3, we have already gone through in detail how the LIH can capture different properties of words. The LIH as shown in (14) is now repeated below as (30):
Syntactic rules cannot make reference to any part of the internal structure of a lexical item.

There is no consensus on the notion of word. The LIH, however, has been shown to be a valid criterion of lexical structure, because it successfully distinguish the domain of syntax from that of morphology: the LIH is not relevant to a phrase but a lexical unit (words and compounds) is constrained by the LIH. Accordingly, the present study adopts the LIH as the defining criterion of wordhood and endorses the Thesis of Atomicity proposed by Di Sciullo and Williams (1987). The definition of word is given in (31) below:

(31) A word is atomic at the level of phrasal syntax and phrasal semantics.
    Its internal structure cannot be relevant in syntax.

Since compounds demonstrate a variety of lexical properties, they should observe the LIH also. In the rest of the study, the LIH will be taken as a general criterion of compounds. Concluding from the discussion in sections 2.4.1 to 2.4.3, this study defines compounds as follows:

(32) A compound is a morphological unit that observes the LIH and consists of two or more bound or free morphemes.
The definition in (32) essentially bases on the lexical integrity exhibited by compounds. In other words, (32) only covers those compounds that are strictly lexical. But from what we have seen in section 1.1.2 on phrasal properties of compounds, the discussion of the compound-phrase distinction in section 2.2.1 to 2.2.4, and the discussion of separability in section 2.4.3, we know that many compounds exhibit syntactic properties, for instance, separability. Consequently, the definition of compound given in (32) has to be modified in order to cover these compounds. In the present study, we propose that there are two types of compounds: one type is inseparable and the other separable. For the sake of clarity, we will call the inseparable type “lexical compounds” and the separable type “phrasal compounds”. The lexical compounds and the phrasal compounds are defined below as (33) and (34) respectively:

(33) Lexical Compound

A lexical compound is a morphological unit that consist of two or more bound or free morphemes and observes the LIH.

(34) Phrasal Compound

A phrasal compound is a morphological unit that consists of two or more bound or free morphemes. Its internal structure is visible to syntax.
Justification for such a distinction between lexical and phrasal compounds will be offered in the next chapter.

2.5 Summary

This chapter has been concerned primarily with the definitions of word and compound. It has been shown that there is no one universal criterion of word that can be applied to all languages. Nevertheless, it is argued that the LIH can be used as a primary criterion of wordhood.

A number of accounts on the distinction between compounds and phrases have been reviewed. Based on the analyses and criteria offered in these studies, we have proposed that compounds can be classified into two types: lexical and phrasal compounds. In the next chapter, we will examine the nature and behavior of Cantonese verb-object compounds, the focus of this study, along with an evaluation of the studies on the dual nature of this type of compound.
3.0 Introduction

In the previous chapter we have discussed the properties of different basic notions in linguistics: word, compound and phrase. Basing on these discussions, we now move on to the focus of this thesis: verb-object compounds (VOCs). This term has been used for a long time in the literature (e.g. Ding 1961, Chao 1968), but it is argued in a number of studies that the term is not precise enough. Some linguists argue that the term “verb-object compound” cannot adequately characterize the relationship of the nominal and the verbal constituents within the compound (Chi 1984, Wang 1994). They state that in a verb-object compound the nominal component is not necessarily a semantic object of the verbal component, so it is not appropriate to use the term “verb-object compounds”. This can be illustrated by the Mandarin Chinese examples in (1) below:

(1) a. kai-dao

open – knife

“to operate (on someone’s body or body parts)”
b. tiao-lou

jump – building

“to jump off a building”

The verb kai, meaning “open”, can take a door, a window, a can and so on as its semantic object. But a knife cannot be opened. So dao “knife” in (1a) should not be a semantic object of kai. In other words, dao is not a patient argument, but an instrument used to “open” literally somebody’s body or body parts. In (1b), the VOC consists of the verb tiao “jump” and the noun lou “building”, meaning that “a person jumps off a building”. So lou indicates the starting point of the action of the verb, i.e. source of the action, but is not a semantic object of the verb tiao “jump”. Similar examples can be found in Cantonese:

(2)  a. tiu-saan

jump – parachute

“to parachute”

b. cim-kat

sign – card

“to pay by credit card”

c. haang-gaai

walk – street

“to go shopping”
In (2a), *saan* "parachute" in the compound *tiu-saan* "to parachute" refers to the instrument needed for the action. It is not a semantic object of the verb *tiu* "jump". Consider (2b). When one *cim-kat* (literally "sign-card"), one is not actually signing on a card. The object *kat* "card" here indicates the instrument involved in the action designated by the compound. As for (2c), *gaai* "street" indicates the location in which the action *haang-gaai* "to go shopping" takes place. So the thematic role of *gaai* is location; it cannot be the object of the verb *haang* "walk".

In view of the existence of VOCs in which the nominal component is not a semantic object of the verbal component such as those in (1), some linguists think that other terms should be used instead of "verb-object compounds". For instance, Chi (1984) proposes a more general term "verb-noun compounds". Wang (1994:48) thinks that "verb-complement compounds" is a more precise term, because he argues that in verb-object compounds "the nominal is always an internal argument of some kind to the verb (either obligatory or optional) and it constitutes part of the complement structure of the predicate verb".  

Nevertheless, it is important to note that in some studies the authors provide a list of possible semantic roles of the nominal constituent when they discuss verb-object compounds (Hu 1992, Tsui 1999). For instance, Hu (1992) notes that the relationship between the constituents in a VOC is not simply a relationship between an action and a patient (theme). In terms of semantics, apart from being the object to which the action is acted on, the noun in a VOC can also be the
location of the action (e.g. *shang-an* "go up-coast = to go on land from sea") and instrument used by the action (e.g. *kai-dao* "open-knife = to operate on someone"). In the study by Tsui (1999), she notes that apart from being the affectee of the action denoted by the verb, objects in verb-object combinations can also be the location of the action and instrument, for instance, in verb-object compound *baai-saan* "worship-hill = visit grave on Qingming day", the “object” *saan* is the location of the action and in *daa-zam* "hit-needle = to inject drugs", *zam* is the instrument. Here we see that those linguists who use the term “verb-object compounds” are fully aware of the fact that the nominal component of the compound is not restricted to the direct object of the verbal component.

Given the wide usage of the term “verb-object compounds” in the literature (Ding 1961, Chao 1968, Li and Thompson 1981, Zhu 1982, Matthews and Yip 1994, Packard 2000), we will follow the tradition and use the term to refer to compounds which are formed from verb and noun in this study.

In this chapter we examine the behavior of verb-object compounds (VOCs) in Cantonese. We begin by reviewing the properties of Cantonese VOCs. Then we outline some attempts which aim to explain the nature of VOCs. Finally, we point out that the recognition of the existence of phrasal compounds is of great significance for a proper characterization of VOCs in Cantonese.
3.1 General Properties of Cantonese VOCs

Cantonese VOCs can be classified by parts of speech. The classification given in (3) divides VOCs into two general groups according to their grammatical categories:

(3) a. VOCs function as nouns

<table>
<thead>
<tr>
<th>忍</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>fei-dip</td>
<td>fly - plate</td>
</tr>
<tr>
<td>han-zing</td>
<td>carry out - affairs</td>
</tr>
<tr>
<td>wu-sat</td>
<td>protect - knee</td>
</tr>
<tr>
<td>tai-coeng</td>
<td>watch - place</td>
</tr>
<tr>
<td>daa-sau</td>
<td>hit - hand</td>
</tr>
<tr>
<td>ngo-dai</td>
<td>lay - bottom</td>
</tr>
<tr>
<td>laai-lin</td>
<td>pull - chain</td>
</tr>
<tr>
<td>bat-ho</td>
<td>pluck - river</td>
</tr>
<tr>
<td>tiu-zou</td>
<td>jump - flea</td>
</tr>
<tr>
<td>hon-gaang</td>
<td>look - shift</td>
</tr>
</tbody>
</table>

b. VOCs function as verbs

<table>
<thead>
<tr>
<th>忍</th>
<th>meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>soeng-din</td>
<td>load - electricity</td>
</tr>
<tr>
<td>gaap-haugung</td>
<td>match - evidence</td>
</tr>
<tr>
<td>jing-soeng</td>
<td>record - photo</td>
</tr>
<tr>
<td>luk-jam</td>
<td>record - sound</td>
</tr>
</tbody>
</table>

“UFO”
“administration”
“kneecap”
“guard”
“bodyguard”
“undercover”
“zipper or zip-fastener”
“tug-of-war”
“flea”
“guard”
“to take drugs”
“to lie together”
“to photo”
“to record”
sik-daufu  eat – tofu  “to tease flirtatiously”

fong-feigei  release – plane  “to stand somebody up”

zung-tauzoeng win – first prize  “to hurt one’s head”

ceot-hei  exhale – gas  “to vent one’s spleen”

zaat-paaau  tie – cannon  “to fast”

dok-seoi  measure – water  “to borrow money”

Owing to the limitation of space, the present study will focus only on those VOCs which function as verbs, i.e. the type of compounds as shown in (3b). In the following subsections, we will examine different properties exhibited by the verbal VOCs. We will show that although some VOCs behave like phrases in certain aspects, they are morphological in nature. The reason is that these compounds conform to the definitions of compound proposed in Chapter Two.

3.1.1 Meaning

3.1.1.1 Compositionality of Meaning

In Chapter Two we have discussed the validity of using the idiomaticity of meaning for the identification of compounds. Our conclusion is that one cannot simply reply on this semantic criterion to define compounds. Nevertheless, it is argued that semantic opacity is one of the prominent features of compounds. The
VOCs in (4) below have been classified according to the thematic roles of the nominal constituent:

(4) **Patient:**

- *daa-se* hit – snake “to punish non-residents in a hostel”
- *caau-ce* fry – car “to crash one’s car in an accident”
- *deng-bou* throw – pot “to break up”
- *sik-daan* eat – egg “to score nothing”

**Theme:**

- *bou-zuk* boil – congee “to talk on the phone for a long time”
- *duk-syu* read – book “to study”
- *zau-fan* run – powder “to deal in drug traffic”
- *daa-gaa* hit – price “to get quotations”

**Source:**

- *ceot-saan* go out – hill “to come out from retirement”
- *fong-gaam* release – jail “to be released from prison”
- *tiu-lau* jump – building “to jump off a building”
- *ceot-gaa* go out – home “to become a Buddhist monk”
Goal:

*ceot-hoi* leave – sea “to go sailing”

*maai-zaam* approach-stop “to stop at a station or stop”

*maai-wai* approach-seat “to get ready”

*jap-cong* enter – factory “to receive treatment in a hospital”

Instrument:

*luk-kat* rotate – card “to pay by credit card”

*daa-zam* hit – needle “to inject drugs”

*hoi-dou* open – knife “to have an operation”

*tiu-saan* jump – parachute “to parachute”

All the compounds in (4) are semantically opaque. Their meanings are not compositional; one cannot get the meanings of these compounds from their constituents.

Another important property related to the semantics of VOCs is the generic interpretation of the “object” in the compound. Consider the examples in (5) below:

(5) *duk-syu* read – book “to study”

*se-zi* write – character “to write”

*deng-waam* throw – curve “to make a turn”
\begin{tabular}{lll}
\textit{cau-gan} & draw – tendon & “to cramp” \\
\textit{cong-baan} & hit – board & “to make a mistake” \\
\textit{lok-daan} & drop – bill & “to order” \\
\textit{zaa-sau} & hold – hand & “to shake hands” \\
\textit{ceot-syu} & go out – book & “to publish books” \\
\textit{gwo-wu} & cross – account & “to transfer money” \\
\textit{sai-cin} & wash – money & “to spend money” \\
\end{tabular}

In the examples in (5), the objects in the compounds are obligatory but semantically bleached. They have generic interpretation rather than refer to definite objects.

From the illustrations above, one can notice three important properties regarding the semantic relations between the constituents of a VOC. First, different thematic roles can be identified with respect to the nominal constituent, for instance, patient, instrument, source and so on. Second, the meaning of the entire compound is not derivable from the verbal and the nominal components. For example, \textit{ceot-gaa} “leave – home” does not mean “leave one’s home”. Rather it has an idiomatic meaning, referring to someone becoming a Buddhist monk. Third, the nominal constituents (i.e. the “object”) of VOCs often have generic interpretation.
3.1.1.2 Anaphoric Reference

Anaphoric islandhood as a universal property of morphological structures is well-documented in the literature (Postal 1969, Sproat 1988, Wang 1994, 1998). This constraint assumes that morphological structures are island for anaphoric co-indexation. To be more specific, Wang (1994:57) states that “the non-head constituent in a morphological formation cannot serve as the antecedent for co-indexation.” This is evidenced by the sentences in (6) below:

(6) a. Ce-mun, waai zo, c; hoi m dou laa.

   Car-door broken ASP open not able PRT

   “The car-door is broken, it cannot be opened.”

b. *Ce,-mun waai zo, c; hoi m dou laa.

   Car-door broken ASP open not able PRT

In (6a), the co-indexation involves the whole compound ce-mun “car-door = door of the car” and the empty category. In (6b), it is ce “car” (part of the compound) that is co-indexed with the empty category. The co-indexation in (6a) is legitimate but the one in (6b) is not. We can account for the ill-formedness of (6b) if we take into consideration the effect of anaphoric islandhood on morphological formations. In (6b), the empty category is co-indexed with nominal constituent ce within the compound ce-mun, thus violating the anaphoric island constraint which forbids a proper subpart of a morphological
formation to be co-indexed. In contrast, the co-indexation in (6a) involves co-indexation of the whole compound ce-mun. (6a) does not violate the anaphoric island constraint; therefore, it is grammatical. This anaphoric islandhood is also observed in Cantonese VOCs.

3.1.2 Movement

3.1.2.1 Topicalization

Topicalization is a process whereby a constituent is “made into the topic of the sentence by being moved to a more prominent position at the front of the sentence” (Radford 1997:172). This process can be illustrated by the following sentences which involve the topicalization of the nominal constituent of VOCs in Cantonese:

(7) baai-zau “put-wine = to have a banquet”

\[
\text{Zau, ngo zau dim dou jiu baai ti ge!}
\]

wine I then how all need put PRT

“Having a banquet, to me, is a must!”

(8) gin-gung “see-job = to attend a job interview”

\[
\text{Gung, dou mei gin ti zau fong hei?}
\]

job all not yet see then release abandon

“What do you give up before attending the job interview?”
This phenomenon demonstrated in (7) to (9) leads some linguists to question the nature of VOCs (e.g. C.-T. Huang 1984, Jin 1991). Their puzzle is that if VOCs are morphological formations, their constituents should not be separable by syntactic rules. But as shown in the sentences in (7) to (9), the nominal constituents zau, gung and gaa of VOCs baai-zau, gin-gung and ceng-gaa do move to the front position, leaving the verbal constituents stranded. Consequently, some linguists have proposed different analyses to capture this property of VOCs. For instance, C.-T. Huang (1984) argues that those which are traditionally called “verb-object compounds’ should be analyzed as verb-object phrases stored in the lexicon, thereby providing an explanation for the separability demonstrated by the VOCs. Jin (1991), on the other hand, maintains the morphological status of VOCs but argues that there is a process which allows the constituents of VOCs to be separated. That process is ionization which is the formation of a compound from a phrase.

Note that some VOCs, however, do not allow their nominal constituent to be topicalized. This is evidenced by the contrast in (10) to (12) in which (a)-numbered sentences have canonical word order while (b)-numbered
sentences involve topicalization of the nominal constituent of VOCs.

(10) **hoi-sam** “open-heart = to be happy”

a. Gamjat ngo hou **hoisam** aa.
   today I very open-heart PRT
   “Today, I am very happy.

b. *Sam, gamjat ngo hou **hoi** t, aa.
   heart today I very open PRT

(11) **lau-ji** “stay-spirit = to pay attention to somebody or something”

a. Ngo **lauji** zo keoi hou noi gaalaa.
   I stay-spirit ASP he/she very long time PRT
   “I have paid attention to him/her for a long time.”

b. *Ji, ngo **lau** t, zo keoi hou noi gaalaa.
   spirit I stay ASP he/she very long time PRT

(12) **zyun-sam** “specialize-heart = to concentrate”

a. Nei jiu **zyunsam** zi teng dou syu gaa.
   you need specialize-heart then listen able book PRT
   “You need to concentrate in order to understand the lecture.”

b. *Sam, nei jiu **zyun** t, zi teng dou syu gaa.
   heart you need specialize then listen able book PRT
At this stage, we are not going to probe into the difference in the grammaticality
with respect to the topicalization of nominal constituents of Cantonese VOCs.
However, enough has been said to illustrate that not all VOCs behave exactly the
same. We cannot take it for granted that all VOCs form one homogeneous group.

3.1.2.2 Passivization

In Cantonese, a typical passive construction contains the word “bei”
(Matthews and Yip 1994:149). (13a) shows a Cantonese sentence with the
canonical word order: subject-verb-object while (13b) is a passive construction:

(13) a. Ngo sik zo go pingwo.
   I   eat ASP CL apple
   “I ate the apple.”

   b. Go pinggwo bei ngo sik zo.
   CL apple by me eat ASP
   “The apple was eaten by me.”

As discussed in section 3.1.1.1, the object of a VOC has a generic interpretation;
it does not refer to any specific entity in the real world. Therefore, one does not
expect that the nominal constituent (i.e. the “object”) of a VOC can be passivized.
This predication is borne out for Cantonese VOCs.⁶
(14) a. Ngo din zo faat aa, seng saambaak man gaa.
I electrify ASP hair PRT almost three hundred dollar PRT
“I curved my hair. This cost me almost three hundred dollars.”
b. *Faat, bei ngo din zo t, aa, seng saam baak man gaa.
hair by me electrify zo PRT almost three hundred dollar PRT

(15) a. Keoi sengjat tok-saucaang gaa.
he/she often support-elbow PRT
“He/She often refuses to help.”
b. *Saucaang sengjat bei keoi tok t, gaa.
elbow often by he/she support PRT

3.1.3 Separability

In Matthews and Yip’s grammar (1994), the separability demonstrated by VOCs is highlighted. Matthews and Yip provide a list of items that may intervene between the constituents of a VOC. These items include aspect markers, verbal particles, modifiers of the object, expressions of duration of frequency and personal pronouns. The study of the property of separability is particularly important for a proper understanding of VOCs because this property shows that a VOC, being a lexical item, behaves like a phrase in some respects.
In section 2.4.3, we have reviewed the separability of Chinese VOCs discussed by Chao (1968). The following discussions will describe the situations under which a Cantonese VOC can be separated.

3.1.3.1 Semantic Object

VOCs behave differently in terms of transitivity. Here, transitivity simply refers to the ability of a VOC to take a semantic object. Cantonese VOCs can be classified into two types: transitive VOCs and intransitive VOCs. A transitive VOC takes a semantic object while an intransitive VOC does not. Transitive VOCs can be further divided into three subtypes according to the position of the object that a VOC allows. The classification given in (16) to (19) below divides VOC constructions in terms of the transitivity of VOCs and the position that a semantic object can appear.

(16) Intransitive VOCs

a. *ceot-maau
   go out – cat “to cheat in an examination”
   Keoi sengjat ceotmaau gaa.
   he/she often go out-cat PRT
   “He/She often cheats in examinations.”

b. ceot-maau
   go out – cat “to cheat in an examination”
   *Keoi sengjat ceotmaau haausi gaa.
   he/she often go out-cat examination PRT
(17) Transitive VOCs

Type (I) – Postverbal Object

*waai-ji*
bear – doubt “to doubt somebody”

a. Nei mou leijau *waaiji* ngo gaa.
   you have not reason bear-doubt I PRT
   “There is no reason for you to doubt me.”

b. *Nei mou leijau waai ngo ji gaa.*
   you have not reason bear I doubt PRT

*zing-gu*
make – poison “to tease somebody”

c. Ngo sengjat zinggu ngo ge tunghok.
   I often make-poison my classmate
   “I often tease my classmates.”

d. *Ngo sengjat zing ngo ge tunghok. gu.*
   I often make my classmate poison

(18) Transitive VOCs

Type (II) – Intervening Object

*caau-jaujyu*
fry – squid “to fire an employee”

a. Nei m kanlik zou je zau caau nei jaujyu.
   you no diligent do thing then fry you squid
   “If you don’t work hard, I’ll fire you.”
b. *Nei m kanlik zou je zu caau jaufyu nei.
   you no diligent do thing then fry squid you

   **hoi-dou**
   open - knife “to operate (on somebody’s body or body parts)”

c. Aamaa kau jisaang mhou **hoi loudau dou**.
   mum beg doctor don’t open dad knife
   “Mum begged the doctor not to operate on dad.”

d. *Aamaa kau jisaang mhou **hoidou loudau**.
   mum beg doctor don’t open-knife dad

(19) Transitive VOCs

Type (III) – Object introduced by BONG/ZOENG

tai-beng
   see - disease “to diagnose”

a. Bong **bengjian taibeng** hai jisaang ge zyujiu gungzok.
   for patient see-disease be doctor’s major work
   “The major duty of a doctor is to diagnose his or her patients.”

b. *Taibeng **bengjian** hai jisaang ge zyujiu gungzok.
   see-disease patient be doctor’s major work

c. *Tai **bengjian beng** hai jisaang ge zyujiu gungzok.
   see patient disease be doctor’s major work
fan-leoi
separate — type “to classify or categorize something”

d. Ngodei jiu zoeng di syu fanleoi.
we need ZOENG these book separate-type
“We need to categorize these books.”
e. *Ngodei jiu fanleoi di syu.
we need separate-type these book
f. *Ngodei jiu fan di syu leoi.
we need separate these book type

Intransitive VOCs cannot take any semantic object. Transitive VOCs of Type (I) take post-verbal objects. Objects occur inside the compounds for transitive VOCs of Type (II). For transitive VOCs of Type (III), the objects are introduced by a preposition bong “for” or zoeng (similar to Mandarin Chinese ba). The classification of VOCs given in (16) to (19) above shows that not all transitive VOCs are separable by their objects. Only Transitive VOCs of Type II allow their constituents to be separated by their semantic objects.7

3.1.3.2 Aspect Markers

As compounds are assumed to have morphological status, one expects that a verbal Cantonese VOC will be treated as a single verb in syntax. A verbal aspect marker, for instance, gwo (experiential aspect marker, meaning “at least once before”) or zo (perfective aspect marker, indicating an event is completed)
should follow the VOC but not inside the VOC. This prediction is borne out with respect to the type of transitive VOC which takes postverbal object, i.e. Transitive VOCs of Type (I):

(20) a. Ngo mou waaiji gwo nei.
    I have not bear-doubt ASP you
    “I never doubt you.”

b. *Ngo mou waai gwo ji nei.
    I have not bear ASP doubt you

(21) a. Ngo fukmou zo ne gaan gungsi sap nin laa.
    I serve-duty ASP this CL company ten year PRT
    “I have served this company for ten years.”

b. *Ngo fuk zo mou ne gaan gungsi sap nin laa.
    I serve ASP duty this CL company ten year PRT

As one can see in (20) and (21), the aspect markers gwo and zo must follow the VOCs waai-ji “doubt” and fuk-mou “serve” respectively. However, if one takes Transitive VOCs of Types (II) and (III) into consideration, one finds that the aspect marker does not follow the VOC but occurs between the verbal constituent and the nominal constituent of the VOC. This can be shown by the examples in (22) and (23) below:
(22) Transitive VOCs of Type (II)

a. Neigo jisaang bong ngo hoi gwo dou.
   this doctor for I open ASP knife
   “This doctor has operated on me before.”

b. *Neigo jisaang bong ngo hoidou gwo.
   this doctor for I open-knife ASP

(23) Transitive VOCs of Type (III)

a. Ngodei jijing zoeng di syu fan zo leoi.
   we already ZOENG these book separate ASP type
   “We have categorized these books.”

b. *Ngodei jijing zoeng di syu fanleoi zo.
   we already ZOENG these book separate-type ASP

If VOCs are morphological formations, one will not expect anything other than infixes to enter into their structures. The appearance of the intervening aspect markers raises questions as to the morphological status of VOCs in Cantonese. We will return to this issue in detail when we discuss the formation of VOCs in Chapter Four.
3.1.3.3 Duration and Frequency Adverbials

In C.-T. Huang’s (1984) analysis, he makes use of a phrase structure condition on Chinese that he proposes in his previous study (1982) to explain the compatibility of compounds and adverbials indicating duration and frequency. The condition is given in (24) below:

(24) The Phrase Structure Condition (PSC)

Within a given sentence in Chinese, the head (the verb or verb phrase) may branch to the left only once, and only on the lowest level of expansion.

The PSC as stated in (24) requires that a verb in Chinese be followed by at most one constituent. So when a transitive verb is followed by its object, it cannot be followed by an adverbial, such as a resultative adverbial, a descriptive adverbial, a duration phrase or a frequency expression. The examples below are taken from C.-T. Huang’s (1984) analysis:

    he/she ride house DE very tired

   b. *Ta xie xin de hen kuai.
    he/she write letter DE very fast
c. *Ta kan shu le liang ge zhongtou.

he/she read book ASP two CL hour

d. *Ta xue Yingwen le liang ci.

he/she learn English ASP two time

In the examples above, the verb in each of the sentences is followed by two constituents—the object of the verb and an adverbial. In (25a) *qi “ride” is followed by *ma “horse” and *hen lei “very tired”. In (25b) *xie “write” is followed by *xin “letter” and *hen kuai “very fast”. The verb *kan “read” is followed by *shu “book” and *liang ge zhongtou “two hours” in (25c). In (25d) the verb *xue “learn” is followed by *Yingwen “English” and *liang ci “twice”. The ill-formedness of the sentences in (25) can therefore be attributed to the violation of (24), the PSC.

The Phrase Structure Condition proposed by C.-T. Huang (1984) has been challenged by Y.-H. Li (1985, 1990). The present study also has reservations about the applicability of the PSC to Cantonese. The PSC does rule out some ungrammatical Cantonese sentences. Consider the illustrations in (26) and (27):

(26) a. *Ngo hokji zo sap nin.

I learn-medicine ASP ten year

b. Ngo hokji hok zo sap nin.

I learn-medicine learn ASP ten year

“I have studied medicine for ten years.”
(27) a. *Ngodei ceoiseoi zo saam go zungtau.
    we blow-water ASP three CL hour
b. Ngodei ceoiseoi ceoi zo saam go zungtau.
    we blow-water blow ASP three CL hour

“We have chatted for three hours.”

Following the PSC in (24), one can argue that (26a) and (27a) are ill-formed, because in each sentence the verb is followed by two constituents. In (26a) hok “learn” is followed by the object ji “medicine” and the time adverbial sap nin “ten years”. In (27a) ceoi “blow” is followed by seoi “water” and saam go zungtau “three hours”. If the two sentences are rewritten so that the verb is followed by only one constituent (i.e. without violating the PSC), they will become grammatical as shown in (26b) and (27b). The PSC seems to be able to explain the grammaticality of sentences like (26) and (27). Nevertheless, we do find examples that argue against the PSC. In each of the following Cantonese examples, the verb is followed by two constituents:

(28) Ngo hok zo sap nin ji.
    I learn ASP ten year medicine

“I have studied medicine for ten years.”
If the PSC is operational in Cantonese, it should rule out the above examples in which a verb is followed by two constituents. In (28) the verb hok “learn” is followed by sap nin “ten years” and ji “medicine”. In (29) the verb ceoi “blow” is followed by saam go zungtau “three hours” and seoi “water”. However, both sentences are grammatical. This shows that the PSC in (24) may not be applicable to Cantonese.

C.-T. Huang accounts for the distribution of duration and frequency adverbials with respect to compounds without assuming the PSC in his verb raising analysis (1992). We will review that analysis in the discussion of the separability demonstrated by Cantonese VOCs in Chapter Four.

3.1.4 Status of the Cantonese VOCs

Summarizing the description of the behavior of VOCs provided in sections 3.1.1 to 3.1.3, we can present the properties of VOCs as follows:
As indicated in the table above, Cantonese VOCs generally behave like morphological formations, except for the fact that some VOCs allow their nominal constituent to be moved, and some VOCs allow the elements like semantic object and aspect markers to intervene. As the VOCs in Cantonese show predominantly lexical properties, the present study assumes that all these VOCs should be morphological in nature.

Recall that in Chapter Two, two types of compounds have been identified; they are lexical compounds and phrasal compounds. Lexical compounds include those compounds which strictly follow the LIH. Those Cantonese VOCs which behave like morphological formations in all respects should belong to this type of compound. As for those VOCs which show certain phrasal properties, they should be classified as phrasal compounds which allow the constituents of the compounds to show a limited degree of separability.

The next section will discuss the existing accounts on VOCs. Basing on the analyses of these previous accounts, we will provide further justification for the dichotomy of lexical and phrasal VOCs proposed here.
3.2 Analyses of VOCs

In this section we will give a review on those studies which aim at explaining the nature of VOCs. Since there is no systematic study on Cantonese VOCs so far, the following discussions will be based on analyses of VOCs in Mandarin Chinese, a language that is closely related to Cantonese.

3.2.1 S.-F. Huang (1984)

S.-F. Huang's (1984) study attempts to come up with a structural principle that can capture the property of the syntax of Chinese VP. The Surface Structure Condition (SSC) proposed in his study is a descriptive generalization, as stated in (30) below:

(30) The SSC:

No surface structure is well-formed if it contains a sequence of the form:

Verb-C₁-C₂

where the verb is restricted to those verbs that are not ditransitive or cognate object verbs and C refers to any constituent.

The SSC relies on the linear order of the elements in a sentence. On the surface the SSC seems to be able to capture the (un)grammaticality of the VO constructions in Chinese:
In (31a), the verb mian “dismiss” is followed by one constituent zhi “duty”. The verb xi “wash” appears in (31c) twice; each time it is followed by only one constituent (zao “bath” and yi ge xiaoshi “an hour” respectively). Both (31a) and (31c) follow the SSC and so they are grammatical. In (31b) and (31d), on the other hand, the verbs are followed by two constituents. The verb mian “dismiss” in (31b) is followed by zhi “duty” and ta “he/she” while in (31d) the verb xi “wash” is followed by zao “bath” and yi ge xiaoshi “an hour”. Thus, the two sentences violate the SSC and so they are ungrammatical. Although the SSC can account for the grammaticality of constructions like those in (31), it is being called into question of its empirical inadequacy (e.g. C.-T. Huang 1984). Consider the sentences below:
Both (32a) and (32b) violate the SSC, because in these two sentences the verbs, which are not ditransitive or cognate object verbs, are followed by two constituents. In (32a), *dan* "carry" is followed by *xin* "heart" and *ta* "him/her" while in (32b), *chu* "leave" is followed by *ban* "edition" and *yi tao baikequanshu* "a set of encyclopedias". However, both (32a) and (32b) are grammatical sentences. This shows that the SSC is inadequate in describing the syntax of Chinese VOCs, since it wrongly predicts that (32a) and (32b) are ill-formed. The SSC proposed by S.-F. Huang (1984) is an inadequate condition because it fails to capture the difference between a VOC, e.g. *mian-zhi* "dismiss-duty = to fire" that is intransitive and a VOC, e.g. *dan-xin* "carry-heart = to worry" that is transitive and requires an obligatory semantic object.
The goal of C.-T. Huang’s study is to provide a unified account for the identification of compound and phrase in Chinese. In order to achieve this goal, two principles are used in his study. The first principle requires that the internal structure of a sentence or verb phrase in Chinese be largely head-final but only trivially head-initial, namely, the Phrase Structure Condition (PSC) which we have stated as (24) in section 3.1.3.3 and now repeated as (33) below. The second is a universal principle underlying much work in linguistic theory, which requires that the internal structure of a word or a lexical category be inaccessible to rules of syntax (Chomsky 1970, Lapointe 1979): the Lexical Integrity Hypothesis (LIH). The two principles are stated as follows:

(33) The PSC:

Within a given sentence in Chinese, the head (the verb or VP) may branch to the left only once, and only on the lowest level of expansion.

(34) The LIH:

No phrase-level rule may affect a proper subpart of a word.

Recall that the SSC proposed by S.-F. Huang (1984) wrongly predicts that a well-formed sentence such as Wo hen danxin ta. “I worry about him/her very
Sentences in (35) can be used to show that the LIH and the PSC are applicable to Chinese. Consider (35a) first. It is argued in C.-T. Huang’s study that dan-xin “to worry” in (35a) should be analyzed as a lexical item. For one thing, the item in question observes the LIH, because the two morphemes in dan-xin “to worry” are not separated. Furthermore, since dan-xin is followed by an object zhe jian
shi “this matter”, it should be treated as a single unit \( V_{[VO]} \). If \( dan-xin \) were analyzed as a phrase, then the verb \( dan \) “carry” would be followed by two objects \( xin \) “heart” and \( zhe jian shi \) “this matter”, violating the PSC.

Consider now the sentences in (35b), (35c) and (35d). C.-T. Huang points out that by the LIH, \( dan-xin \) would be a phrase, because in each of these examples, \( xin \) has been separated from \( dan \). Furthermore, since in each of these sentences, \( dan \) is not followed by another object, the PSC will not require the VO to be a word. Therefore, by both the LIH and the PSC, \( dan-xin \) is more appropriately regarded as a phrase, in particular an idiom phrase, given its semantic noncompositionality.

The existence of (i) VOCs like \( ti-yi \) “pick-discussion = to propose” which behaves as a single indivisible word and (ii) VOCs like \( dan-xin \) “carry-heart = to worry” which sometimes behaves like a separable phrase makes one wonder why some compounds (e.g. \( ti-yi \)) are not separable while some others (e.g. \( dan-xin \)) are separable in some limited ways. In view of this, C.-T. Huang hypothesizes that all VO combinations are in fact listed as phrases in the lexicon. Then, when these phrases are inserted into sentence-final position, nothing need take place. But if they are inserted into sentence-medial position, with an object following them, they would undergo a process of lexicalization and become words.

The merit of C.-T. Huang’s (1984) hypothesis is that it provides us with a solution to the controversial issue on whether VO sequences in Chinese are words or phrases. More importantly, with this hypothesis, linguistic theory need
not tolerate a contradictory notion of compounds which defines them in terms of LIH but speaks of the existence of some separable compounds. Note that within the generative framework, C.-T. Huang’s proposal of the lexicalization of VO phrases (VOPs) can be captured under the idea of reanalysis. In Larson’s (1988) study of double object construction, he proposes that optional V’-reanalysis can be applied at the level of syntax. C.-T. Huang argues that in Chinese a VOP is lexicalized when the phrase is followed by a structural object. This lexicalization process can be regarded as an instance of V’-reanalysis, because C.-T. Huang’s proposal regarding the separability of VO sequences is, in essence, a reanalysis of V-one-bar category into a V-zero category. With the availability of syntactic V’-reanalysis, the difference in separability exhibited by Chinese VO combinations is minimized. However, a theoretical problem ties up with the idea of the lexicalization of VOPs. Larson’s analysis allows optional reanalysis. But in C.-T. Huang’s analysis, the lexicalization of a VOP when it is followed by a structural object is obligatory. C.-T. Huang simply claims that there will be such a lexicalization process, without explaining why reanalysis is sensitive to particular lexical items in Chinese. This has weaken the theoretical solidness of his analysis and the explanatory power of the PSC, because his hypothesis on the lexicalization of VOPs is not well-grounded.
3.2.3 Paul (1988)

Based on the observation that many VO sequences are separable, Paul (1988) follows C.-T. Huang's (1984) analysis in that the ability to take a structural object serves as a criterion for the identification of VOCs. However, she does not discuss intransitive VOCs. As a matter of fact, her analysis does not make a distinction between transitive and intransitive VOCs. Those items which we normally regard as intransitive VO compounds such as *shang-feng* "hurt-wind = to catch a cold" and *chu-shi* "leave-world = to be born" are simply treated as intransitive VOPs in Paul's study without justification.

As for the distinction between VOCs and VOPs, Paul argues that VO sequences like *chu-ban* "go out-edition = to publish" should be treated as a VOC, since objects can immediately follow the sequence (just like an object can follow a simple verb). VO sequences like *kai-dao* "open-knife = to operate (on someone's body or body parts)" and *ba-zhi* "terminate-duty = to fire" should be VOPs, because they can be separated by aspect suffixes and modifiers of the object component. The distinction can be illustrated by the examples below:

(36) Zui jin Shangwu chu-ban le yi tao baikequanshu.

most near Commercial Press go out-edition ASP one CL encyclopedia

"The Commercial Press has published a set of encyclopedias recently."
At first glance, Paul’s criterion seems to be able to distinguish between VOCs and VOPs. According to Paul’s criterion, *chu-ban* “go out-edition = to publish” in (36) is followed by a structural object and so it should be a compound. For *ba-zhi* “terminate-duty = to fire” in (37), since it is separated by a perfective aspect suffix *le*, it is a phrase, not a compound. However, it can be seen that Paul’s criterion is in some cases problematic, since there are VO sequences which have dual status in that they function as words when followed by structural objects and as phrases elsewhere. Consider the examples below:

(38) a. Ni *dan*xin *ta* ma?
    you carry-heart he/she PRT
    “Do you worry about him/her?”

b. *Ni *dan* le *xin* *ta*.
    you carry ASP heart he/she

c. Ta *dan* le *ban* tian *xin*.
    he/she carry ASP half day heart
    “He/She worried for quite a while.”
(38a) shows that *dan-xin* "carry-heart = to worry" behaves as a transitive verb which takes an object *ta* "him/her" in the postverbal position. (38b) shows that insertion of an aspect suffix *le* into *dan-xin* results in an ungrammatical sentence. According to Paul’s analysis, *dan-xin* should be a compound. However, it can be shown that *dan-xin* also behaves like a phrase in some other situations. In (38c), an aspect marker *le* and an event quantifier *ban tian* are inserted between *dan* and *xin*. In (38d), the object component of the VO sequence, i.e. *xin* is modified by *shei de* “who”. Following Paul’s criterion, *dan-xin* in (38c) and (38d) will be regarded as a VOP, because the two components of the sequence can be separated. Here, one can see that Paul’s criterion of VOC fails to determine the status of the VO sequence in (38). In other words, compounds like *dan-xin* “to worry” which has dual status as illustrated in (38a) to (38d) serve as strong evidence to argue against Paul’s distinction between VOCs and VOPs.

3.2.4 Wang (1994)

Wang’s (1994) study systematically argues that Chinese verb-complement compounds (VCCs)\(^ {12} \) are formed by the process of noun incorporation. With
Reference to a modified version of Dowty’s (1979) proposal, the treatment of the definition of word by Di Sciullo and Williams (1987) and the process of noun incorporation argued by Baker (1988a), Wang identifies two sets of distinctions to classify VCCs in Chinese: (i) morphological and phrasal and (ii) lexical and syntactic. The opposition between morphological and phrasal formations is based on a difference in the relationship between the constituents of a VCC. In a morphological formation, the non-head constituent is structurally dependent on its head constituent; while in a phrasal construction, the non-head constituent is structurally free. The opposition between lexical and syntactic processes is based on a difference in the level of grammatical organization. Lexical processes are assumed to take place at the presyntactic level; while syntactic processes take place in syntax. These two sets of distinctions interact with each other, yielding four types of VCCs:

$$
\begin{array}{c|c|c}
\text{MORPHOLOGICAL} & \text{LEXICAL} & \text{SYNTACTIC} \\
\text{Morpho-LEXICAL} & \text{Morpho-SYNTACTIC} \\
\text{Phrasal-LEXICAL} & \text{Phrasal-SYNTACTIC} \\
\end{array}
$$
Examples of each type of VCCs are given in (40) below:

(40)

(i) Morpho-lexical VCCs

\(chu-jia\) leave – home “to become a Buddhist monk”

\(tao-nan\) escape – calamity “to seek refuge in flight”

(ii) Morpho-syntactic VCCs

\(qi-ma\) ride – horse “to go riding”

\(chou-yan\) inhale – smoke “to smoke”

(iii) Phrasal-lexical VCCs

\(qi\ liang-pi ma\) ride two-CL horse “ride two horses”

\(chou\ si-gen yan\) inhale four-CL smoke “smoke four cigarettes”

(iv) Phrasal-syntactic VCCs

\(peng\ yi-bizi hui\) bump one-nose dust “to be let down”

\(tong\ yi-ge mafengwo\) hit one-CL bee hive “to get into trouble”

Wang focuses on morpho-lexical and morpho-syntactic VCCs in his study. He postulates that these two types of VCCs are derived by the same process—noun incorporation, but at different levels of grammatical organization: the former takes place at presyntactic level in the lexicon, while the latter at S-structure in syntax.

Morpho-lexical compounds such as \(kai-xin\) “to be happy”, \(chu-jia\) “to become a Buddhist monk” and \(kai-dao\) “to operate (on someone’s body or body parts)” are fixed and unexpandable expressions with idiosyncratic meanings. In contrast, morpho-syntactic compounds such as \(kan-shu\) “read book”, \(kai-hua\) “to be in blossom” and \(kai-če\) “to drive” can be easily expanded into \(kan-xian-shu\)
“read-pleasure-book”, *kai-hong-hua* “open-red-flower” and *kai-qi-che* “drive-motor-car”. They represent a word formation process that is both syntactically productive and semantically compositional. The tree diagrams in (41) and (42) represent the underlying structure of *kai-xin* “to be happy” and *kai-hua* “open flower”, respectively:

(41)

```
VP
  /
 ... V'
  /
   V

V
 /
V  N
 kai  xin
```

(42)

```
VP
  /
 ... V'
  /
   V

V  NP
 /
V  N  N'
 kai  hua
```

The structure *kai-xin* in (41) is formed presyntactically and therefore cannot be freely expanded as a syntactic structure. On the other hand, as illustrated in (42) the expression *kai-hua* is not brought into a morphological structure until S-structure; thus, the postverbal constituent is free from the head verb at D-structure and can therefore be expanded syntactically.
Assuming that noun incorporation of these two types of VCCs takes place at different levels, Wang is able to provide a straightforward explanation for the differences in productivity between morpho-lexical and morpho-syntactic VCCs. Another merit of his study is that he shows the possibility for word formation process to take place at different levels of grammatical organization, thereby accounting for the differentiation of the syntactic behavior of VCCs in Chinese. Nevertheless, an important issue has been left unexplained in his study. That is the theoretical motivation for analyzing the compounding process as an instance of noun incorporation. Under the Minimalist paradigm (Chomsky 1993, 1995), movement, comparing to merging, is a costly operation. In consequence of the economy principle which bans “the use of superfluous constituents and operations” (Radford 1997:110), an operation like movement is used only when there is no other way of satisfying the grammatical requirement or explaining the language phenomenon. In Wang’s study, it is argued that the formation of Chinese VCCs is essentially a process of noun incorporation (i.e. a head-to-head movement). His analysis merely focuses on the technicality of the movement mechanism without providing the motivation for such a movement process. In light of the economy principle, one naturally wonders whether any alternatives, e.g. merging, can also explain the formation of VCCs. As a matter of fact, it is exactly the position taken by the present study—we assume that VOCs (i.e. VCCs in Wang’s analysis) are formed by a simple process of merging, but not movement. Our analysis will be presented in Chapter Four.
3.2.5 Her (1997)

Her’s (1997) study analyzes Chinese VO sequences from a diachronic perspective. Specifically, he interprets the variation of VO sequences, i.e. some behave like word and some behave like phrase, as the consequence of the interaction between two historical processes: lexicalization and ionization. The typology of VO sequences in Her’s study is shown in (43) below:

(43)

<table>
<thead>
<tr>
<th>Behave like WORD</th>
<th>Behave like IDIOM PHRASE</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>-</td>
<td>(1) Word</td>
</tr>
<tr>
<td>-</td>
<td>+</td>
<td>(2) Idiom phrase</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
<td>(3) Dual status</td>
</tr>
<tr>
<td>-</td>
<td>-</td>
<td>(4) Regular phrase</td>
</tr>
</tbody>
</table>

Her argues that VO sequences of Type (4) are completely governed by syntactic rules and have no place in the lexicon. Therefore, they behave neither as a word nor an idiom phrase. The other three types of VO sequences are obtained from the interaction of lexicalization (LEX)—formation of morphological items from phrases and ionization (ION)—formation of phrasal items from words or compounds. The formation processes of Type (1) to Type (3) VO sequences are shown in (44):
<table>
<thead>
<tr>
<th>Type 1: Word</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complete LEX</td>
<td>de-yi</td>
</tr>
<tr>
<td>([V][\ldots][O])</td>
<td>([VO])</td>
</tr>
<tr>
<td>Phrase</td>
<td>Word</td>
</tr>
<tr>
<td>“gain-spirit = be proud”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type 2: Phrase</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>No process involved</td>
<td>chi-cu</td>
</tr>
<tr>
<td>([V][\ldots][O])</td>
<td>([V][\ldots][O])</td>
</tr>
<tr>
<td>Phrase</td>
<td>Idiom phrase</td>
</tr>
<tr>
<td>“eat-vinegar = be jealous”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Type 3: Dual Status</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete LEX</td>
<td>dan-xin</td>
</tr>
<tr>
<td>([V][\ldots][O])</td>
<td>([VO])</td>
</tr>
<tr>
<td>Phrase</td>
<td>Word</td>
</tr>
<tr>
<td>(i.e. ([V][\ldots][O]) and ([VO]) coexist)</td>
<td></td>
</tr>
<tr>
<td>“carry-heart = to worry”</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Incomplete LEX</th>
<th>you-mo</th>
</tr>
</thead>
<tbody>
<tr>
<td>([VO])</td>
<td>([V][\ldots][O])</td>
</tr>
<tr>
<td>Word</td>
<td>Phrase</td>
</tr>
<tr>
<td>(i.e. ([VO]) and ([V][\ldots][O]) coexist)</td>
<td></td>
</tr>
<tr>
<td>“humour or to tease somebody”</td>
<td></td>
</tr>
</tbody>
</table>

Her holds that the original VOP often loses its phrasal status after lexicalization, yielding a VOC as illustrated in Type (1) above. VO sequences of Type (2) are phrasal in nature. Some of them have not undergone lexicalization nor ionization, but only have their meaning lexicalized (i.e. an idiom phrase) and some of them gain their phrasal status via ionization of words (which came from lexicalization).
With respect to Type (3), Her argues that they are the result of incomplete lexicalization or ionization. In sum, Her holds that the existence of a variety of VO sequences is the result of the interaction of two historical processes: lexicalization and ionization.

Upon closer scrutiny, however, this historical approach does not have much explanatory value for an adequate description of VO structure in Chinese. In Her’s analysis, there is no constraint formulated to govern or condition the application of lexicalization or ionization. One is not able to know under what circumstances lexicalization or ionization will apply. Her has not explained the motivation for the process of lexicalization and ionization. So one may wonder why lexicalization or ionization happens in certain VO sequences but not others. Another related issue that Her’s study has not addressed is the criteria for becoming a candidate of a particular type of VO sequences. For instance, he has not explained why Type (3) of VO sequences will undergo an incomplete lexicalization or ionization but VO sequences of Type (1) and Type (2) do not.

In sections 3.2.1 to 3.2.5, we have reviewed briefly different accounts which aim at explaining the nature of VOCs and distinguishing VOCs from VOPs, pointing out their strengths as well as weaknesses. It is found that none of these analyses can offer a satisfactory explanation to the nature of VOCs.

In the next section, we will show that the very first step for a proper understanding of Cantonese VOCs is to classify VOCs into two types—one type being lexical and the other type phrasal.
3.3 Existence of Two Types of VOCs in Cantonese

3.3.1 Distinction between Lexical and Phrasal VOCs

The studies of compounds are of significance because many compounds possess the properties of both words and phrases. In other words, compounds blur the boundary between morphology and syntax. Some linguists accept that some morphological items can behave like syntactic phrases in limited ways (e.g. Chao 1968, Borer 1988, 1991, 1997, Wang 1994). But some linguists deny the existence of separable compounds (e.g. C.-T. Huang 1984, Packard 2000). In the previous sections, different analyses on the nature of VOCs have been reviewed, the question of the dual nature of VOCs (i.e. showing both morphological and syntactic properties) remains unanswered.

In section 3.1.4, it is suggested that we can make use of two types of compounds—lexical compounds and phrasal compounds proposed in Chapter Two for the analysis of Cantonese VOCs. The definitions of these two types of compounds are repeated as (45) and (46) below:

(45) Lexical Compound

A lexical compound is a morphological unit that consist of two or more bound or free morphemes and observes the LIH.

(46) Phrasal Compound

A phrasal compound is a morphological unit that consists of two or more bound or free morphemes. Its internal structure is visible to syntax.
Lexical compounds are those behaving like words. Phrasal compounds are those "problematic" compounds which show both morphological and syntactic properties. Traditional analyses hold that compounds should behave like words in every aspect. However, the present study argues that acknowledging the existence of phrasal compounds allows us to (i) have a better understanding of the nature of compounds and (ii) make a clear distinction between compounds and phrases. In Chapter Two we argued that the LIH can be used as the criterion for the identification of compounds. Those compounds that observe the LIH are inseparable and are termed as "lexical compounds". And those compounds whose internal structure is visible to syntax are called "phrasal compounds".

Our assumption regarding the dichotomy of lexical and phrasal VOCs gains support from the discussion of existing analyses in previous sections which have shown that it is inadequate to treat all VOCs as one homogeneous group. The present analysis, therefore, assumes that the classification of lexical and phrasal compounds proposed in Chapter Two can be applied in the characterization of VOCs. Specifically, this study proposes that there are two types of Cantonese VOCs which function as verbs. The first type is "lexical VOCs" and the other type "phrasal VOCs". Their definitions are given as (47) and (48) respectively:
Lexical Verb-Object Compound

A lexical verb-object compound is a morphological formation which consists of a verbal constituent and a nominal constituent and observes the LIH.

Phrasal Verb-Object Compound

A phrasal verb-object compound is a morphological formation which consists of a verbal constituent and a nominal constituent. Its internal structure is visible to syntax.

There are two advantages for this classification of Cantonese VOCs. First, we can explain why VOCs show both morphological and syntactic properties. Precisely speaking, it is phrasal VOCs but not lexical VOCs that have dual status. Second, we are able to differentiate VOCs from VOPs. Specifically, lexical VOCs (e.g. gwaan-sam “close-heart = to care about somebody or something”) are constrained by the LIH but VOPs (e.g. sik pingwo “eat an apple”) are not. Those VO sequences which are not completely constrained by the LIH but have idiomatic meanings are phrasal VOCs, for instance, hoi-dou “open-knife = to operate (on someone’s body or body parts)”. The constituents of this phrasal VOC can be separated as in hoi keoi dou “open-him/her-knife = to operate on him/her” and hoi gwo loeng ci dou “open-ASP-twice-knife = have been operated twice”. But we will not treat hoi-dou on a par with ordinary phrases, because
unlike normal phrases, the meaning of *hoi-dou* is not compositional. Furthermore, the object *dou* “knife” in expressions such as *hoi keoi dou* and *hoi gwo loeng ci dou* are not referential; it does not refer to an actual knife in the real world. The data in (49) and (50) are examples of the two types of Cantonese VOCs proposed in the present study:

(49) Lexical VOCs

| a. caap-sau | insert — hand | “to interfere something” |
| b. ceot-bing | go out — soldier | “to attack a country” |
| c. zing-gu | make — poison | “to tease somebody” |
| d. zing-gwai | make — ghost | “to tease somebody” |
| e. hei-cou | uplift — grass | “to draft a document” |
| f. gwaan-sam | close — heart | “to care about somebody” |
| g. zaat-gan | tie — root | “to settle in a particular area” |
| h. tau-san | throw — body | “to enter a particular field” |
| i. ceot-baan | go out — edition | “to publish something” |
| j. lau-sam | stay — heart | “to pay attention to something” |

(50) Phrasal VOCs

| a. saan-seoi | spread — water | “to leave” |
| b. dok-kei | measure — period | “to arrange one’s schedule” |
| c. jing-tau | admit — head | “to claim responsibility” |
| d. zeoi-tau | gather — head | “to meet” |
| e. laai-caai | pull — firewood | “to die” |
| f. zong-daangung | install — spring | “to set somebody up” |
| g. haap-cou | swallow — vinegar | “to be jealous” |
| h. daa-hobaau | hit — wallet | “to steal a wallet” |
| i. bok-zeoi | connect — mouth | “to argue with somebody” |
| j. dan-dunggu | boil — mushroom | “to dismiss somebody” |
The technicality of (i) the nature of the verbal and nominals constituents of the compounds and (ii) the formation processes of both types of VOCs will be offered in Chapter Four.

3.3.2 Diagnostic Tests for the Lexical-Phrasal VOC Distinction

Li and Thompson (1981) note that reduplication is a morphological process in which a morpheme is repeated so that a new word is formed by the original morpheme plus the repeated morpheme. Applied to verbs, this reduplicated form signals delimitative aspect. Specifically, Li and Thompson (1981:29) state that “the reduplication of an action verb has the semantic effect of signaling that the actor is doing something ‘a little bit’” (see also Li 1998 and Zhu 1998 for relevant discussions).

Li and Thompson point out that only volitional verbs are allowed to undergo delimitative aspect reduplication. They further note that if the volitional verb is a VOC whose components are separable, then only the verbal constituent of that VOC is reduplicated. For instance, in shui-jiao “sleep-sleep = to sleep”, only the verbal constituent of the compound, i.e. shui is reduplicated to form shui-shui-jiao to signal delimitative aspect. If a VOC whose components are inseparable is to signal delimitative aspect, then the reduplication involves the entire compound. The inseparable compound liu-yi “stay-spirit = to pay attention
to something”, for instance, will be reduplicated to form *liu-yi-liu-yi but not *liu-liu-yi in order to be in the delimitative aspect.

In light of the analysis given by Li and Thompson, we believe that the two types of Cantonese VOCs should employ different means to express delimitative aspect. Consequently, the difference in the behavior of the two types of VOCs in signaling delimitative aspect should be able to function as a diagnostic test for the distinction of lexical and phrasal VOCs. Our prediction is borne out when we consider the position of haa, a marker used to signal delimitative aspect in Cantonese, with respect to the two types of VOCs. Lexical and phrasal VOCs locate the marker haa in different positions. The examples in (51) to (54) below illustrate how the two types of VOCs signal delimitative aspect:

(51) Lexical VOC gwaan-sam
    a. gwaan-sam “close-heart = to care”
    b. gwaan sam haa “to care a little”
    c. *gwaan haa sam

(52) Lexical VOC waai-ji
    a. waai-ji “bear-doubt = to doubt”
    b. waai ji haa “to doubt a little”
    c. *waai haa ji
(53) Phrasal VOC fan-gaau
   a. fan-gaau “sleep-sleep = to sleep”
   b. fan haa gaau “to sleep a little”
   c. *fan gaau haa

(54) Phrasal VOC ceoi-seoi
   a. ceoi-seoi “blow-water = to chat”
   b. ceoi haa seoi “to chat a little”
   c. *ceoi seoi haa

In (51) and (52), the lexical VOCs gwaan-sam “close-heart = to care” and waai-ji “bear-doubt = to doubt” can signal delimitative aspect by placing haa after the compound as in (51b) and (52b). Insertion of haa into the compounds results in ungrammatical expressions as indicated in (51c) and (52c). In contrast, the phrasal compounds in (53) and (54), i.e. fan-gaau “sleep-sleep = to sleep” and ceoi-seoi “blow-water = to chat” can only signal delimitative aspect by inserting haa between the verbal and the nominal constituents, as shown in (53b) and (54b). Placing haa after the compounds, as in (53c) and (54c), are ungrammatical.

From the examples in (51) to (54), one can clearly see that the lexical and phrasal VOCs in Cantonese differ systematically with respect to the position of
the delimitative aspect marker *haa*. This difference can therefore be used as a diagnostic test for distinguishing lexical VOCs from phrasal VOCs.

### 3.4 Summary

In this chapter the focus has been the behavior of VOCs in Cantonese. A general survey of the prominent properties of these compounds has been provided. Then various analyses of VOCs have been reviewed. It is shown that these accounts cannot adequately explain the phrasal properties of VOCs in Cantonese and they fail to make a clear distinction between VOCs and VOPs. It is further shown that Cantonese VOCs should not be treated as a homogeneous group. They should be divided into two types: lexical VOCs and phrasal VOCs. In the next chapter, we are going to present an analysis of Cantonese VOCs that will make use of this division of VOCs and will accommodate those unique language phenomena observed in these compounds.
4.0 Introduction

The central theme which has run through this thesis is that a compound has a dual nature—behaving like both words and phrases. One of the difficult questions concerning the study of compounds is that it is hard for us to know whether we are looking at morphology or syntax or both. Borer (1998:151) says, “The interaction between syntax and word formation has always been a battle ground, on which many important linguistic wars have been fought”. The issue of compounding, to a large extent, is a war of this type. In the literature we can see radically different opinions as to whether a particular phenomenon, for instance, compounding, should be regarded as a morphological or a syntactic process (or a process belonging to some other domain).

Many of the questions about compounding hinge on the more fundamental question of the autonomy of the morphological module: Is morphology an independent component, subject to its own restrictions, or should it be subsumed under the syntactic component, observing syntactic restrictions? As Borer (1998:152) remarks in an article discussing the relation between morphology and syntax, this is not an easy question to answer:
Proponents of an independent word-formation component must show that such a component includes operations and constraints which cannot be reduced to independently motivated syntactic conditions. They must further show that an independent range of phenomena that cannot otherwise be accounted for. Proponents of an exclusively syntactic word formation, on the other hand, must do the opposite: they must provide a way of accounting for the richness of WF phenomena, without appealing to any syntactic processes which are not otherwise motivated.

Issues related to the distinction between morphology and syntax have not been completely resolved yet. However, it does not imply that investigation of phenomena lying on the boundary between these two modules is impossible. On the contrary, research on morphosyntactic issues such as compounding, cliticization and incorporation will definitely enhance our understanding of morphology and the relation between morphology and syntax.

In the previous chapter, we have seen the efforts of different analyses to provide a unified account of compounds in the hope of forming new theories regarding the process of compounding. We have seen that these existing analyses have left different issues to be tackled with respect of the nature of VOCs, and we have further shown that the problems of these analyses are caused by, among other things, treating all compounds as one single group. In the present study, it is argued that there are in fact two types of VOCs in Cantonese: lexical VOCs and phrasal VOCs. In this chapter we attempt to integrate the insights gained from the previous analyses and then propose a mechanism by which these two types of VOCs are formed.
4.1 Theoretical Assumptions

4.1.1 Parallel Morphology

In Chapter One the major approaches to the relation between morphology and syntax have been briefly reviewed. Central to the present morphosyntactic study of Cantonese VOCs is the concept of Parallel Morphology as formulated in Borer (1988) (see also Borer 1991 on the inchoative-causative, and Spencer 1991 for a similar proposal\(^1\). The core claim of Parallel Morphology proposed by Borer (1988, 1991, 1997) is that words can be formed either before syntax, at syntax or after syntax and this difference in the levels of formation has no effect on the lexical status of the derived words, but only on their syntactic behavior. In a study of Hebrew compounds and constructs, Borer argues that both compounds and construct state nominals in Hebrew are words and their formation process is similar, but the place of the formation is different, resulting in slightly different syntactic properties. To explain the derivation, Borer proposes that the Word Formation component (i.e. morphology) is organized in such a way that its operation can apply at any levels of the grammatical system, given that no well-formedness conditions are violated. A diagrammatic representation of her idea (1988:3) is given in (1) below:
The arrows in (1) indicate that the insertion of lexical items is possible at every step of the derivation: prior to D-structure, at syntax and at PF. In other words, lexical items can be formed at different levels. We will have a general review of the arguments in Borer’s (1988) study for the purpose of a thorough understanding of the idea of Parallel Morphology. We will also consider an analysis on Chinese compounds and an analysis on Japanese compounds which show that word formation process can indeed take place in various levels of the grammar.

The analysis on Hebrew compounds and construct state nominals presented by Borer (1988) is an attempt to account for the behavior of two lexical formations with similar structure. More importantly, it aims to show that the model presented in (1) is valid. First, consider the common properties of these two types of formations. When a compound in Hebrew is made definite, the non-head must be modified instead of prefixing the definite article ha- to the head.²
From (2a) to (2c), one can see that only the non-head can be modified if a compound is to be made definite. The head of a compound cannot be modified directly. Construct state nominals behave like compounds in this regard. Their head cannot be modified:

(3) a. ᵇdashif ha-yalda

scarf the-girl

the girl’s scarf

b. *ھdashif (ha-)yafe ha-yalda

scarf the pretty the-girl

Intended meaning: the girl’s pretty scarf
modified. Another similar property between compounds and construct state nominals is phonological:

(4) a. baýt
    “house”

b. beyt xolím
    house sicks
    “hospital”

c. ha-caíf shel ha-yalda
    the-scarf of the-girl
    “the girl’s scarf”

d. cǝ́íf ha-yaldá
    scarf the-girl
    “the girl’s scarf”

From (4), one can see that the compounds in (4b) and the construct in (4d) have word-like phonology, having only single primary stress.

Although compounds and constructs in Hebrew have similar properties in that they behave like words, they differ in certain aspects. First, construct state nominals are productive and semantically transparent but compounds are unproductive and semantically opaque. Second, construct state nominals allow modification of their constituents but compounds do not. Thirdly, conjunction of
the non-head with ve-/u- “and” is permitted in constructs as in (5a) but it is ungrammatical in compounds as in (5b):

(5) a. shomer batim u-mexoniyot
   guard houses and-cars
   “a guard of houses and cars”

   b. *gan yeladim ve-xayot
   garden children and-animals
   Intended meaning: kindergarten and zoo

Another difference between compounds and constructs in Hebrew lies in the referentiality:

(6) a. hu bana li shney batey əec ve-’exad mi-plastik
   he built for-me two houses wood and-one from-plastic

   b.*hu bana lanu shney batey xolim ve-’exad le-zkenim
   he built for-us two houses sicks and-one for-old(s)

The head of a construct can be referred to by a pronominal element as shown in (6a) but this kind of anaphoric reference is not possible in compounds as in (6b).

As we have seen earlier, both compounds and constructs in Hebrew behave like words and so they should be formed by the word formation component, the
morphological component. Since construct state nominals exhibit syntactic properties but compounds do not, it is sensible to assume that the formation processes of the two types of morphological structures take place at different levels. Specifically, compounds should be derived before syntax while constructs should be derived at syntax. Borer regards the contrast between the syntactic behavior of compounds and that of construct state nominals in Hebrew as strong evidence to show that word formation operations can be applied at different levels of grammatical organization.

In the literature there are other word formation studies which can be treated as evidence supporting the model presented in (1). For instance, Wang’s study (1994) on Chinese verb-complement compounds (VCCs). He argues that the two types of VCCs he identified: morpholexical and morphosyntactic VCCs are formed at different levels but via the same process: noun incorporation. Wang proposes that morpholexical VCCs are formed at a presyntactic level, namely lexical-semantic level, while morphosyntactic VCCs are formed at S-structure. The tree diagrams in (7) and (8) represent the D-structure of a morpholexical VCC *kai-xin* “open-heart = to be happy” and D-structure of a morphological syntactic VCC *kai-hua* “open-flower = to blossom” respectively (taken from Wang 1994:79-80):
According to Wang’s analysis, both VCCs in (7) and (8) are formed via the process of noun incorporation, but they behave differently in terms of syntax. He argues that the difference in the level of formation of the two types of VCCs explains the difference in their syntactic properties. Specifically, morpholexical VCCs like *kai-xin* “open-heart = to be happy” is formed presyntactically, its postverbal nominal constituent is not an independent NP at D-structure; therefore, this compound cannot be expanded syntactically. But morphosyntactic VCCs like *kai-hua* “open-flower = to blossom” are formed at S-structure, so its nominal constituent is syntactically visible; therefore, this compound can be expanded to
form kai-hong-hua “open-red-flower”, for instance. The analysis proposed by Wang here lends support to the model of Parallel Morphology given in (1). It demonstrates that word formation can take place either before syntax or at the syntactic level.

Another piece of evidence supporting the model presented in (1) is concerned with Japanese compounding. Shibatani and Kageyama (1988) argue that compounds in Japanese can be formed either in the lexicon or after rules of syntax and phrasal phonology have applied. What is particularly interesting about their analysis is the compounds formed by the latter process, i.e. what they refer to as postsyntactic compounds which consist of a noun phrase and a verb, as exemplified in (9) below:

(9) a. Amerika-hoomon
    America-visit
    “a visit to America”

b. Yooroppa-ryokoo
    Europe-travel
    “an European tour”

c. Syeikuspia-kenkyuu
    Shakespeare-study
    “study of Shakespeare”
Consider the major characteristics that make Japanese postsyntactic compounds resemble words. First, there are no case particles inside the compounds. Furthermore, their verbal constituents do not inflect for tense. It is also observed that the compounds do not allow modifiers to intervene, exhibiting lexical integrity, as evidenced by the contrast in (10) below:

(10) a. Yooroppa o nonbiri ryokoo-tyuu ni
   Europe ACC\(^4\) leisurely travel-middle in
   “in the middle of traveling Europe leisurely”

b. *[Yooroppa-nonbiri-ryokoo]-tyuu ni
   Europe-leisurely-travel middle in

Examples in (10) clearly indicate that postsyntactic compounds in Japanese resist syntactic interruption. In short, there is enough evidence to show that postsyntactic compounds should be morphological. Nevertheless, these compounds also demonstrate properties that are absent in lexical compounds. For instance, anaphoric islandhood, i.e. the constraint forbidding anaphoric relation to make reference to the non-head of a lexical item, does not apply to postsyntactic compounds. Another property distinguishing the postsyntactic compounds from lexical compounds is that they have the accent pattern which correspond to the syntactic phrases but not the tonal patterns associated with lexical items. Spencer (1991) elaborates on this property of Japanese compounds.
when he discusses the notion of parallel morphology. Consider the illustration given in his study (1991:443-444):

(11) a. aMERIKA
   b. aMERIKA-Hoomon
   c. *aMERiKA
   d. *amerika
   e. aMERIKA o hoOMON no ori
   f. aMERIKA-hoOMON no ori

   (High-toned syllables are shown in capitals)

In Japanese accent system, a lexical item bears exactly one unbroken stretch of high tone, i.e. one sequence of one or more high-toned syllables. This is illustrated in (11a) which is a word and (11b) which is a lexical compound. (11c) and (11d) are examples showing impossible patterns. The word aMEriKA in (11c) has two stretches of high tone separated by a low tone while america in (11d) has no high tone. Now consider (11e) which is a phrase. In this example, the accentuation is different from that of lexical item in that the string has two stretches of high tone. The postsyntactic compound in (11f) has the same accentuation as the corresponding phrase in (11e); it also contains two stretches of high tone.

Given the observations that the items in (9) show a variety of lexical
properties, Shibatani and Kageyama argue that they should be compounds. The syntactic properties exhibited by these compounds only suggest that they are not formed in the lexicon but at a later stage of derivation. Shibatani and Kageyama argue that the accentuation of compounds illustrated in (11) provides strong support for analyzing them as formed postsyntactically. Specifically, they suggest that since these compounds have essentially the same accentuation as the corresponding phrase, these compounds should be formed after the phrasal phonological rules concerning pitch assignment have applied. In other words, the compounding process is in effect turning syntactic phrases into compounds; therefore, it should be postsyntactic. In order to capture this particular compounding phenomenon in Japanese, Shibatani and Kageyama propose the model in (12) below:

(12)

One can see that Shibatani and Kageyama’s model in (10) is similar to the one proposed by Borer (1988) as shown in (1). To sum up, Borer’s analysis of compounds and construct state nominals in Hebrew offers a case study of her
Parallel Morphology model operating in the syntactic level, while Shibatani and Kageyama's detailed study on postsyntactic compounding in Japanese can be regarded as an instance of this model operating at the level of PF.

4.1.2 Lexical-Semantic and Lexical-Syntactic Representations

Levin and Rappaport Hovav's (1995) argue in their study on unaccusativity that each verb is associated with two lexical representations: a lexical-semantic representation and a lexical-syntactic representation. The nature of these representations are given as follows (Levin and Rappaport Hovav 1995:20-21):

(13)

The lexical semantic representation, sometimes called a lexical conceptual structure (Hale and Keyser 1986, 1987, Jackendoff 1990) or simply a conceptual structure (Jackendoff 1983), encodes the syntactically relevant aspects of verb meaning, whereas the lexical syntactic representation—typically called an argument structure—encodes the syntactically relevant argument-taking properties of a verb.

These two lexical representations are used to explain the formation and nature of two related verb types: causative and unaccusative verbs. Sentences in (14) and (15) are examples of unaccusative and causative verbs respectively.
(14) a. The ship *sank* slowly.
   
   b. The door *opened*.
   
   c. The sky *cleared*.
   
   d. The ball *rolled* down the hill.
   
   e. The window *broke*.

(15) a. The bomber *sank* the ship.
   
   b. I *opened* the door.
   
   c. The wind *cleared* the sky.
   
   d. John *rolled* the ball down the hill.
   
   e. Kate *broke* the window.

The phenomenon that a verb can be used in both unaccusative and causative constructions is known as the causative alternation. Some more examples that allow causative alternation include *dry, melt, rotate* and *spin*. Levin and Rappaport Hovav argue that causative verbs are the basic forms whereas unaccusative verbs are derived from them. According to their analysis, unaccusative verbs are formed by a decausativization process in which the external theta role of a causative verb is suppressed; it is a process in which a two-place predicate is changed to a one-place predicate. Take the formation of the unaccusative verb *break* as an example. The external theta role of this verb is suppressed with respect to its lexical-semantic representation. So no external
theta role enters into the subsequent level to become an external argument. In other words, the verb *break* has no external argument to be projected from its lexical-syntactic representation (argument structure) to the level of syntax. This explains why we can only see the internal argument of the unaccusative *break* in syntax such as (14e). As for the causative counterpart of *break*, Levin and Rappaport Hovav assume that, unlike the formation of unaccusative verb, there is no suppression of external theta role of this basic form. So the lexical-semantic representation of the verb contains both external and internal theta roles. These theta roles become the external and internal arguments of the verb respectively and subsequently project to the syntactic level, obtaining a construction such as (15e). The idea of lexical representations prior to syntax advocated by Levin and Rappaport Hovav convincingly explains the formation process of unaccusative and causative verbs in English.

The assumption of the two lexical representations gains further support from the analysis of passive constructions. Consider the English passive sentences in (16) below:

(16) a. Tony was killed.

b. Elsa was brought to the farewell party.

Grimshaw (1990) argues that passive forms of verbs are formed at the level of argument structure. Passivization can be understood as a process in which the
external argument of a verb is suppressed; this external argument therefore cannot project to the level of syntax. However, the suppressed external argument can be "brought" back to the adjunct position of the sentence with a by-phrase. This explains why phrases related to the external argument (e.g. manner of action) can appear in passive constructions, as exemplified in sentences (17) below:

(17) a. Tony was cruelly murdered by the killers.
    b. Elsa was purposely brought to the farewell party by Peter.

In contrast, unaccusative verbs are not compatible with phrases indicating the purpose or the manner of the action. Consider the illustrations in (18) below:

(18) a. The enemy sank the ship deliberately. [active]
    b. The ship was sunk (by the enemy) deliberately. [passive]
    c. The ship sank. [unaccusative]
    d. *The ship sank to collect insurance. [unaccusative with purpose]
    e. *The ship sank deliberately. [unaccusative with manner of action]

The analysis of Levin and Rappaport Hovav provides a straightforward explanation to the grammaticality of sentences in (18). Passive verbs are formed at the level of argument structure. At this level the external theta role has already been projected into the external argument, it is the external argument (but not the
external theta role) that is suppressed during the process of passivization. Therefore, the external argument can still be recovered, as in (18b). In contrast, the formation of unaccusative verb takes place before the verb enters the argument structure. More concretely, external theta role has already been suppressed in the derivation process, no external argument will be projected to the level of argument structure. Consequently, no effect of the external argument with respect to unaccusative constructions can be seen in the level of syntax. Thus, insertion of phrases related to external argument into unaccusative constructions is ungrammatical, as shown in (18d) and (18e).

The analysis presented by Levin and Rappaport Hovav (1995) supplies strong evidence for (i) the existence of grammatical representations between the lexicon and syntax (see Gu 1995 for an analysis of Chinese causatives in the same spirit), and (ii) the idea that morphology can operate at different levels of the grammatical organization (as argued in Borer 1988, 1991, 1997 and Spencer 1991, among others). The idea of Levin and Rappaport Hovav is represented diagrammatically in (19) below:
4.1.3 Lexical Syntax

With the development of more abstract and powerful syntactic theories in recent years, many researchers attempt to explain morphological processes in terms of syntax and argue that morphological and syntactic structures are derivable from essentially the same mechanism (Hale and Keyser 1993, Ackema 1999). The analysis on Dutch compounds presented by Ackema (1999) is a manifestation of a unified treatment of syntactic and morphological formations. He proposes an X-bar schema which is able to deal with the derivation of both morphological and syntactic structures. The X-bar schema proposed in his study is given in (20) below:

\[
(20) \begin{align*}
X^n & \rightarrow X^{n-1}, Y^n \\
X^{n-1} & \rightarrow X^{n-2}, Z^n
\end{align*}
\]

where \( n = 0 \) or 2

According to Ackema (1999), within one rule of the schema in (20), “\( n \)” must have the same value. If one is dealing with morphological formation, then the value of “\( n \)” should be zero. If, on the other hand, one is using the schema to deal with phrasal formation, then the value of “\( n \)” should be two. A structural representation of the schema in (20) is given in (21):
The diagram in (21) is what Ackema calls the “Syntax below Zero” model. One can see that the formation of $X^0$ category employs exactly the same mechanism as that of $X^2$ (i.e. XP).

Earlier attempts (Selkirk 1982 and Lieber 1992) to unify morphological and syntactic principles have a common problem; they need to alter the content of the standard X-bar principles when they apply the principles to morphology. In other words, these attempts cannot be regarded as true unification, because under these analyses morphology and syntax are not regulated by the same set of principles. The inadequacy of these earlier attempts reflects the merit of Ackema’s proposal. His analysis is able to present morphological and syntactic structures with identical structural representations. This is indeed the most desirable result of a morphosyntactic theory, because it represents a true unification of the representations of the morphological and syntactic components.
4.1.4 VP Shell

Under the current generative approach to language, the sentences in (22) in which the verb has two complements will be analyzed as having a double-VP structure:

(22) a. John filled the pool with water.
    b. Peter rolled the ball down the hill.
    c. Paul broke the window into pieces.
    d. Mary loaded the truck with hay.
    e. David took everything from his mother.
    f. Alice got a present from her teacher.

Following Larson (1988, 1990), Hale and Keyser (1993) and Chomsky (1995), sentences in (22) are argued to have a complex structure, comprising an inner VP shell and an outer vp shell. Take (22a) as an example. The verb filled is first merged with its PP complement with water to form the V-bar filled with water. This is then merged with the DP the pool to form a VP as shown below:

(23) 

```
(23)          VP
             /  \
            /    \
           DP     V'
             /     \
            /       \
          the pool V       PP
             /         \       
          filled     with water
```
The VP structure above is then merged with an abstract causative light verb $\varnothing$. This light verb is affixal in nature and consequently the lexical verb *filled* raises to adjourn to it. The resulting structure is then merged with *John*, the subject of the sentence, to form the complex vp as illustrated in (24) below:

(24) \[
\text{vP} \\
/ \\
/ \\
D \\
/ \\
/ \\
\text{John} \\
/ \\
/ \\
v \\
/ \\
/ \\
v' \\
/ \\
/ \\
\text{filled} \\
/ \\
/ \\
\varnothing \\
/ \\
/ \\
\text{DP} \\
/ \\
/ \\
v' \\
/ \\
/ \\
\text{the pool} \\
/ \\
/ \\
v \\
/ \\
/ \\
\text{PP} \\
/ \\
/ \\
t_i \text{ with water}
\]

The subject *John* then moves to surface subject position due to independent principles of grammar—the Case Filter (Chomsky 1981) as generally assumed in the generative literature. More concretely, the complex vp in (24) merges with an abstract INFL to form I-bar, and the subject *John* raises into SPEC of IP to check its nominative case. Radford (1997:202) notes that the agentive light verb $\varnothing$ is transitive and is assumed to check the accusative carried by the DP, i.e. *the pool* in sentence (22a). A structural representation of the sentence (22a) is given in (25) below:
From (25), one can see that the structure of sentence (22a) comprises an outer vp shell with an inner VP shell embedded within it. This way of analyzing the structure of a sentence is known as VP Shell analysis.

4.2 Levels of Formation of Cantonese VOCs

In Chapter Three, a distinction between lexical and phrasal Cantonese VOCs has been made. The opposition between these types of compounds is based on the LIH. Lexical VOCs have been observed to strictly follow the LIH. They behave like words. Phrasal VOCs, however, do not follow the LIH completely. They have the general properties of morphological formations, for instance, idiomatic meaning and referential opacity. But they also show certain
syntactic properties. As shown in Chapter Three, they allow a limited degree of separability of their constituents.

It is our proposal that the reason for the difference between the behavior of the two types of VOCs lies in the difference of the level of formation. Specifically, the present study proposes that the two types of compounds are formed in morphology but at different levels. As such, we seek to argue against lexicalist analyses, which can be treated as linear models, and which order the morphology as a completely isolated module prior to the syntactic component. The view that morphology is completely isolated from other components is commonly referred to as the Strong Lexicalist Hypothesis which holds that no interaction between the syntactic and the morphological components is allowed. The output of morphology simply serves as the input of syntax. However, it should be noted that the idea that word formation can take place at different levels is not novel in the literature in fact (Borer 1988, 1991, 1997, Shibatani and Kageyama 1988, Spencer 1991, Wang 1994). The present study takes the view that morphology is an independent module in the grammar but, at the same time, it interacts with other components, thereby giving rise to representations which exhibit properties of different modules.

From the discussion of Parallel Morphology in section 4.1.1, we have seen different analyses on different languages, to wit, Borer (1988), Wang (1994) and Shibatani and Kageyama (1988), arguing for an autonomous morphological component which is responsible for the word formation process (including
compounding). Using evidence from Hebrew, Chinese and Japanese compounds respectively, these analyses convincingly argue that an adequate account of the behavior of these compounds requires adopting the idea that word formation processes can take place at different levels (interfaces) of the grammatical system. Assuming that this understanding of the nature of morphology is correct, we have a natural explanation for the different behavior between lexical and phrasal VOCs in Cantonese. We may assume that both lexical and phrasal VOCs are morphological items but they differ in their level of formation. What we regard as lexical VOCs should be formed at a presyntactic level, because that will explain why their internal structure is completely inaccessible by syntax. As for those phrasal VOCs, they should be formed in syntax or a level which is accessible by syntax. The various syntactic properties exhibited by this type of VOCs can therefore be viewed as the manifestation of the accessibility of their internal structure by the syntactic component. This treatment of Cantonese VOCs offers a unified account with respect to their formation. If our assumption is on the right track, then we will have one outstanding question: the place that the derivation processes take place.

In the literature the idea of morphological derivation taking place at different levels has been discussed by many linguists, in particular, Zubizarreta (1985), Grimshaw (1990) and Levin and Rappaport Hovav (1995). In section 4.1.2, we have reviewed the study on unaccusative and causative constructions by Levin and Rappaport Hovav (1995). If their proposal of lexical
representations is adopted, our hypothesis that the formations of lexical and phrasal Cantonese VOCs take place at different levels will receive theoretical support. Specifically, we can assume that there are two interface levels in the morphological component that correspond to the two lexical representations: the *lexical-semantic representation* and the *lexical-syntactic representation* discussed in Levin and Rappaport Hovav's study (Chin 2001). These two morphological levels are *lexical-semantic interface* and *lexical-syntactic interface* at which lexical VOCs and phrasal VOCs are assumed to be derived respectively. The lexical-semantic interface, comparing to the lexical-syntactic interface, is a level which is closer to the lexicon. Any operation at this level will be taken as purely lexical, meaning that no effect from the syntax proper will be possible. This salient property of the lexical-semantic interface gives us a straightforward explanation for the behavior of the lexical VOCs if they are assumed to be derived at this interface level. More concretely, the formation of a lexical VOC involves a verb and a noun; both are selected from the lexicon. The derivation of the compound takes place at the lexical-semantic interface by means of merging. Once this process is finished, the compound will be treated as a single lexical unit thereafter. In other words, its internal structure is not visible in the subsequent levels: lexical-syntactic interface and syntax. Therefore, lexical VOCs in Cantonese strictly follow the LIH and behave like words in every aspect. In contrast, phrasal VOCs exhibit various syntactic properties. We assume this has to do with the level where they are formed, that is, the
lexical-syntactic interface. Comparing to the lexical-semantic interface, this level is closer to syntax. It can be assumed that the lexical items formed at the lexical-syntactic interface will have their internal structure visible to syntax. A corollary of this line of assumption is that compounds formed at this level will have a dual nature. This prediction is borne out for phrasal VOCs. This type of compound exhibits morphological properties on the one hand and demonstrates syntactic properties on the other hand. This shows that the assumption that phrasal VOCs in Cantonese are formed at lexical-syntactic interface is valid.

To sum up, our hypothesis on the levels of formation of Cantonese VOCs can be represented diagrammatically below:

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Formation of Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lexical-Syntactic Interface</td>
<td>Formation of Phrasal VOCs</td>
</tr>
<tr>
<td>Lexical-Semantic Interface</td>
<td>Formation of Lexical VOCs</td>
</tr>
<tr>
<td>Lexicon</td>
<td></td>
</tr>
</tbody>
</table>

4.3 Mechanism of the Formation Processes

4.3.1 Issue of Transitivity

In the previous section, we have presented our proposal with respect to the levels of formation of VOCs in Cantonese. We now move on to the actual mechanism of the derivation process. Examining the verbal constituents of the VOCs discussed in this study, we find that the formation processes of both types
of VOCs target on the same type of verb in terms of transitivity. Specifically, the input of the derivation processes of lexical and phrasal VOCs must be a transitive verb. This can be shown by the examples below:

<table>
<thead>
<tr>
<th>Lexical VOCs</th>
<th>Phrasal VOCs</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. fu-zaak</td>
<td>a. tai-seoi</td>
</tr>
<tr>
<td>b. zi-lik</td>
<td>b. pou-coupei</td>
</tr>
<tr>
<td>c. gaa-gung</td>
<td>c. wui-hei</td>
</tr>
<tr>
<td>d. gan-zung</td>
<td>d. zou-saigaai</td>
</tr>
<tr>
<td>e. dak-zeoi</td>
<td>e. sau-zai</td>
</tr>
<tr>
<td>f. cung-si</td>
<td>f. zeoi-goek</td>
</tr>
<tr>
<td>g. jap-hau</td>
<td>g. bou-wok</td>
</tr>
<tr>
<td>h. ceot-hau</td>
<td>h. jap-biu</td>
</tr>
<tr>
<td>i. zip-sau</td>
<td>i. sau-fung</td>
</tr>
<tr>
<td>j. tai-jii</td>
<td>j. caat-haai</td>
</tr>
</tbody>
</table>

(26)

Lexical VOCs

a. fu-zaak  bear – duty  "to be responsible for"
b. zi-lik    deliver – power "to focus on something"
c. gaa-gung  add – work  "to process something"
d. gan-zung  follow – trace "to tail after somebody"
e. dak-zeoi  obtain – guilt "to offend somebody"
f. cung-si   follow – matter "to work on something"
g. jap-hau   enter – mouth "to import something"
h. ceot-hau  go out – mouth "to export something"
i. zip-sau   receive – hand "to take over something"
j. tai-jii   carry – proposal "to propose something"

Phrasal VOCs

a. tai-seoi  watch – water "to be on the watch"
b. pou-coupei pave – lawn  "to lose in horse racing"
c. wui-hei   return – gas  "to rest"
d. zou-saigaai do – world "to rob somebody"
e. sau-zai   collect – button "to brake"
f. zeoi-goek gather – foot "to meet somebody"
g. bou-wok   repair – cauldron "to rectify something"
h. jap-biu   enter – ticket "to cash a cheque"
i. sau-fung  collect – wind  "to collect information"
j. caat-haai polish – shoe  "to flatter somebody"
In view of the fact that the verbal constituents of all compounds, lexical or phrasal, are transitive, the present study assumes that the compounding process of VOCs in Cantonese is sensitive to transitivity—only transitive verbs can enter the formation processes.

Another important observation regarding the compounding process is the transitivity of the output. From the data collected in this study, all lexical VOCs are transitive while phrasal VOCs are intransitive. No exception has been found in either group. This study assumes that this difference in transitivity is connected with the levels of formation of the VOCs. As we discussed in previous section, lexical and phrasal VOCs are formed at different interface levels, to wit, lexical VOCs are derived at the lexical-semantic interface and phrasal VOCs are derived at the lexical-syntactic interface. The present analysis argues that compounding processes taken place at different interfaces generate output of different nature. We propose that the compounding process at lexical-semantic interface is a lexicalization process which results in the formation of a completely new lexical item which is transitive in nature. In contrast, the compounding process at lexical-syntactic interface is a simple concatenation process which does not entail formation of new words. Technically, the lexical specification of the lexical and phrasal VOCs are different. Lexical VOCs are the output of the compounding process taken place at the lexical-semantic interface. So the lexical information encoded, in particular, the argument structure of a lexical VOC is different from that of its verbal constituent. Phrasal VOCs, on the
other hand, are derived by the compounding process at the lexical-syntactic interface. The lexical specification of this type of compound depends solely on the information encoded in the verbal constituent. Under this analysis of the properties of the two interfaces, the systematic difference between the transitivity of lexical VOCs and that of phrasal VOCs can be attributed to the difference in the levels of formation of the two types of VOCs.

After discussing the relation between the properties of the two interfaces and the transitivity of the lexical and phrasal VOCs, we now turn to the structural representations of the formation of the two types of compounds. In the literature there are different formal approaches to characterize morphological processes. One of the most rigorous attempts is to argue that morphological structures are essentially derived by X-bar principles (Selkirk 1982, Lieber 1992, Ackema 1999, Packard 2000). In this section, we are going to consider Ackema’s (1999) analysis which is a manifestation of a unified treatment of syntactic and morphological formations. The insights in his study should be able to theorize the present proposal of the formation of lexical and phrasal VOCs. Consider the X-bar schema proposed in his study:

\[ (27) \ X^n \rightarrow X^{n-1}, Y^n \]
\[ X^{n-1} \rightarrow X^{n-2}, Z^n \]
where \( n = 0 \) or 2
This mechanism is essentially the X-bar syntax in current literature. A structural representation of the schema in (27) is given in (28) below:

The diagram in (28) shows that one can have identical structural representation for both syntax (i.e. above $X^0$) and morphology (i.e. below $X^0$). On both levels one can identify structurally distinct head, specifier and complement positions.

This study will make use of Ackema’s idea for the analysis of Cantonese VOCs.

In Chapter Three two types of VOCs in Cantonese have been proposed; they are lexical VOCs and phrasal VOCs. In order to distinguish these two types of VOCs structurally, we are going to argue that lexical and phrasal VOCs have their respective structures. Specifically, we assume the representations in (29) and (30) represent the structure of lexical and phrasal VOCs respectively:
Consider the representation for lexical VOCs in (29) first. The V^-2 is transitive and hence selects N^0 as its complement. They will merge to form a V^-1 and finally project to become V^0. As for phrasal VOCs, unlike lexical VOCs, the V^-2 does not merge with N^-2 immediately after it is selected from the lexicon. The resultant V^0 (i.e. the compound) is formed with the N^0 being adjoined to the V^0.

This proposal of the structural representations of the Cantonese VOCs gives us a concrete mechanism by which the compounds are derived. Furthermore, it further theorizes the notion of X^-2 proposed in Ackema’s analysis. Our analysis argues that V^-2 are selected from the lexicon for the formation of VOCs. In doing so, we are, indeed, suggesting that lexical items can be of X^-2 status in the lexicon.

Before closing this section, we will briefly explain the relevance between
Ackema’s schema adopted here and the approach to morphological processes assumed in this analysis. The conception of morphology that the present study adopts is set out in 4.1.1. In particular, we assume that word formation process can take place at different levels of the grammatical system. This is essentially the claim of Parallel Morphology. Following much work on this parallel approach to morphology (Borer 1988, 1991, 1997, Shibatani and Kageyama 1988, Spencer 1991, Wang 1994, among others), we then assumed that the study by Levin and Rappaport Hovav (1995) can explain the formation of lexical and phrasal VOCs and the differentiation of syntactic behavior observed in VOCs. The relevance of Levin and Rappaport Hovav’s analysis to a parallel approach to morphology is that it specifies the linguistic levels where parallel morphology is operative. In this section we have introduced the model of “Syntax below Zero” proposed by Ackema (1999) and used it to represent the formation of Cantonese VOCs. The representations in (29) and (30) can therefore be taken as the structural realizations of the notion of parallel morphology.

We have discussed the theoretical significance of the structural representations in (29) and (30). In order to establish the correctness of these structures, we will show in sections 4.3.2 and 4.3.3 that the proposed structures can account for the difference of the behavior of the two types of compounds.
4.3.2 Formation of Lexical VOCs

At the beginning of this thesis we assume that the lexicon is an inventory in which lexical entries of a language are stored without specifying what sort of information is contained in this inventory. With respect to lexical items, it is generally agreed that the lexicon contains information about their morphological, semantic, syntactic and phonological properties. Focusing on the syntactic properties, we assume that the following information is contained in lexical entries: (i) categorial features, (ii) subcatgeorial features, (iii) selectional properties and (iv) thematic information. Consider the syntactic information encoded in the verb *oi* “love” in Cantonese, for example:

\[(31)\]  
**oi** “love”:

\[
[+V, -N] \quad \text{(Categorial features)}
\]
\[
[ _ \text{NP}] \quad \text{(Subcatgeorial features)}
\]
\[
<\text{Human}> <\text{Human}> \quad \text{(Selectional properties)}
\]
\[
<\text{Agent} \text{ or } \text{Experiencer}, \text{Theme}> \quad \text{(Thematic information)}
\]

This type of characterization of the syntactic information encoded in a lexical item will assist us in explaining the behavior of the lexical VOCs.

Let’s look at the formation of Cantonese lexical VOCs. As set out in previous section, the formation of a lexical VOC is structurally represented as in (29) in which a $V^2$ takes a complement, $N^0$ and then they merge to form $V^1$
which will subsequently project to become $V^0$. Two important properties about the $V^0$ should be noted. First, in terms of syntax this $V^0$ does not have an internal structure, because the $V^2$ and the $N^0$ will combine to become one single unit right after they are selected from the lexicon. Second, since the VOC is formed at the lexical-semantic interface, the $V^0$ formed by compounding is not merely a concatenation of its verbal and nominal constituents, but a formation of a completely new lexical item which is encoded with unique lexical information. In other words, the lexical information contained in the constituent $V^2$ is different from that of the resultant $V^0$. Take *zing-gwai* “make-ghost = to tease” as an example. The syntactic information encoded in the verb *zing* “make” is outlined in (32) below:

(32) *zing* “make”:

\[ [+V, -N] \text{ (Categorial features)} \]
\[ [_ NP] \text{ (Subcategorial features)} \]
\[ <\text{Agent}, \text{Theme}> \text{ (Thematic information)} \]

From (32) we know that this verb will assign a “theme” role to its internal argument. The present study assumes that during the process of compounding, the theta-role assignment proceeds as follows (details aside):
The “theme” role is saturated by the N^0, i.e. gwai “ghost”. After the saturation of
the theta-role, a completely new V^0 zing-gwai is formed with its own unique
lexical information. The lexical entry zing-gwai should contain the following
information, other details aside:

(34) zing-gwai “make-ghost = to tease somebody”:

[+V, -N] (Categorial features)

[ _ NP] (Subcategorial features)

<Agent, Patient> (Thematic information)

Subsequently this new verb will project to syntax to undergo the normal merge
and move processes to form larger structures in which it behaves like a transitive
verb. To illustrate how the present proposal can adequately explain the nature of
lexical VOCs, let’s consider the following puzzling phenomenon concerned with
transitive VOCs, namely the fact that a non-double-object verb like zing “make”
takes two objects. This is exemplified in (35) below:
In (35), one can see that the verb zing “make” is followed by two nouns gwai “ghost” and ngo “I”. From the argument structure of zing presented in (32), it is impossible for zing to take two objects, because the verb has only one theta role to assign to its complement. Under our analysis, however, one has a straightforward explanation for the grammaticality of constructions like (35). Sentence (35) is grammatical, because it is not that case that the verb zing “make” is complemented by two nouns. Rather, it is the compound zing-gwai “make-ghost = to tease somebody” that takes ngo “I” as its direct object (the internal argument). To be more specific, the verb zing and the noun gwai undergo compounding, obtaining an entirely new lexical item zing-gwai which requires a subject and an object, thereby forming a sentence like (35). This analysis is in conformity with the argument structure of the transitive compound zing-gwai as shown in (34) which requires an argument to take up the “patient” role and an argument to take up the “agent” role.

The analysis of the formation of lexical VOCs outlined here also provides an interesting solution to the problem of the difference in the position of the aspect markers between the two types of VOCs in Cantonese. We will come to
that later in our discussion. In this section, enough has been shown with respect to the validity of the tree structure in (29) in explaining the formation of lexical VOCs. Now we turn to the formation of phrasal VOCs in Cantonese.

4.3.3 Formation of Phrasal VOCs

To account for the separation phenomena observed in compounds, various analyses have been proposed (C.-T. Huang 1984, Baker 1988a, Booij 1990, Wang 1994, Ackema 1999, among many others). One of the prominent positions in the literature is that a separable compound should not be considered as a single lexical item. The reasoning is that a morphological structure should not allow syntactic rules to operate on its internal constituents. Since the structures (i.e. separable compounds) in question exhibit phrasal properties, it goes logically that these elements cannot be morphological in nature. The analysis proposed in this study will depart from the mainstream of studies, arguing that phrasal VOCs in Cantonese should be considered as morphological items.

Before defending our position, we will first take a quick look at Booij’s (1990) treatment of Dutch separable complex verbs (SCVs), a phrasal approach to separable compounds. The reason for reviewing Booij’s study is that the theoretical problems in her analysis are typically found in other phrasal analyses. A brief examination of her analysis will serve as a background for our morphological analysis to be presented shortly. In Booij’s (1990) study on SCVs
in Dutch, it is argued that SCVs should not be analyzed as words, but rather as phrases that are formed within the lexicon. To that end Booij has proposed a V* category, a “minimal projection of V”. This projection, according to Booij, is dominated by the VP node and the dominating node for SCVs. The following structure is assigned to SCVs:

\[(36) [P V]_{v*} \] where P stands for particle

Booij (1990:60) claims that (36) is a “specific kind of phrasal construct” and considers it as “a level in between V and the projection of V that dominates V and its objects”. Apparently, Booij’s proposal of the V* level can capture the separability demonstrated by complex verbs in Dutch. However, the proposal of such a transitional V* level necessarily entails a serious theoretical problem to current syntactic theories, because under generative approach to syntax, an intermediate level between XP and X⁰ already exists—that is X’ (X-bar). The incorporation of the notion of V* into standard theory of syntax inevitably leads to a superfluousness. In Booij’s analysis, there is no justification on the theoretical motivation for proposing this intermediate unit, V*. In fact, Booij herself has admitted that further investigation is required to incorporate V*, the notion of “minimal projection of V”, into the theory of projection levels.

The present analysis does not need to assume the existence of items like V* category which is an ad hoc notion to be incorporated into the standard theory of
syntax. What we are arguing is that phrasal compounds are formed at a level which is close to syntax, i.e. the lexical-syntactic interface. This allows the syntactic component to access the internal structure, thereby giving rise to formations which show both morphological and syntactic properties. To be more specific, let us look closely at the structural representation for phrasal compounds proposed in (30). For the ease of explanation, the VOC baau-tou “explore-belly = to improvise” will be taken for illustration. Under our analysis, the verb baau and the noun tou are first selected from the lexicon in which they are stored. They will not combine until they reach the lexical-syntactic interface which is a level prior to the syntactic level. The combination process will be as follows:

The dominating node of the structure in (37) is a $V^0$, i.e. the “starting point” of syntactic projection in the level of syntax. Since the derivation process of this phrasal compound takes place at a level prior to syntax—the lexical-syntactic interface, under the present analysis the complex (i.e. verb-object) nature of the $V^0$ is assumed to be visible in syntax. Consequently, syntactic rules can apply on the constituents of the compound, for instance, raising of the verbal constituent
and topicalization of the nominal constituents. This explains why this type of compound exhibits different syntactic properties.

4.4 Phenomena in Relation to Cantonese VOCs

In the previous sections it is suggested that an adequate distinction between lexical VOCs and phrasal VOCs is only possible if we argue that while both of them are formed prior to syntax—the former is formed at a level closer to the lexicon and the latter is formed at a level closer to the syntax proper. This section provides empirical support to give a full justification for the present analysis. In particular, it will explain two puzzling phenomena concerning VOCs in Cantonese: (i) the separation of phrasal Cantonese VOCs and (ii) the position of aspect markers in lexical and phrasal VOCs.

4.4.1 Separation of Constituents of Phrasal VOCs

At the beginning of this chapter, an analysis on the formation of both lexical and phrasal Cantonese VOCs is presented. It argues that the difference in the behavior of the two types of compounds is due to the difference in their levels of formation. In this section, the separation of the constituents of the compounds will be investigated in detail. Recall that in Chapter Three, the items that can appear between the verbal and the nominal constituents of the Cantonese VOCs have been reviewed. These items include semantic object, aspect markers and
duration and frequency adverbials. This separation phenomenon causes a serious problem to the notion of compound, because compounds are generally argued to be lexical in nature and should not allow anything except infix to enter their structures. The analysis proposed in this study is able to solve the problem nicely. Consider Cantonese VOCs involving event quantification first:

(38) Ngo duk zo sap nin syu.
I read ASP ten year book
"I studied for ten years."

(39) Ngodei ceoi zo saam go zungtau seoi.
we blow ASP three CL hour water
"We chatted for three hours."

In (38) and (39), the time adverbial sap nin "ten years" and saam go zungtau "three hours" are intervening between the VOCs duk-syu "read-book = study" and ceoi-seoi "blow-water = to chat" respectively. Under our analysis, this phenomenon is predictable, because the compounds in the examples are phrasal VOCs which by definition allow syntax to have access to the internal structure. For the technical details about the derivation of sentences like (38) and (39), the present study assumes that one can make use of the insights in C.-T. Huang’s (1992) verb raising analysis.
To explain how the sentences involving VOCs and event quantification are derived, C.-T. Huang (1992) proposes an analysis in which a sentence like (40) is analyzed on a par with English gerundive constructions like (41):

(40) Chinese sentence containing quantity expression

Ta kan le san tian shu.

he/she read ASP three day book

“He/She read (books) for three days.”

Underlying structure (partial):

[IP Ta [r[i -le][vp e san tian kan shu]]] where e is a transitive verb which is phonetically empty and semantically bleached, meaning “do”

(41) English Gerundive Construction

John is angry at Bill’s carelessly dismissing his argument.

Underlying structure (partial):

[IP Bill’s [r[i -ing][vp carelessly dismiss his argument]]]

The motivation for treating Chinese sentences like (40) (and sentences (38) and (39) also) as involving a structure of gerundive nominalization is that they demonstrate a feature that is commonly observed in English gerundive constructions like (41). That feature is syntax-semantics mismatch. In sentences (38) and (39), for instance, one can see that verbal measure phrases like sap nin “ten years” and saam go zungtau “three hours” occur in a syntactic position to modify the object, but they in fact quantify over the actions semantically, given the fact that the objects syu “book” and seoi “water” in (38) and (39) respectively.
are not events and therefore cannot be quantified in terms of frequency and duration. In English gerundive constructions, a similar kind of syntax-semantics mismatch is also observed in the gerundive phrase. More concretely, a gerundive phrase demonstrates both verbal and nominal properties. In (41), the verb *dimiss* within the gerundive phrase takes a direct object *his argument* and is modified by an adverbial *carelessly*. These are typical features of the head of a VP. But considering the syntactic position of the phrase, one can see that the phrase occupies an NP position.

In view of the similarity between the Chinese sentences like (40) and English gerundive constructions like (41), C.-T. Huang proposes that sentences such as (40) involve a structure of gerundive nominalization in which a VP is embedded within a nominal clause IP, as a complement to the nominal I^0. This IP is in turn embedded as the complement of an empty transitive verb which is “phonetically empty and semantically bleached, meaning ‘do’” (C.-T. Huang 1992:486). So for the Cantonese sentence in (38), the derivation will be as follows:
From the tree structure above, one can see that the sentence contains two VPs. The verb *duk* is the head of the embedded VP while the higher VP is headed by an empty transitive verb. The gist of C.-T. Huang’s analysis presented in (42) is that there is a VP embedded under another VP with a phonetically empty head which will trigger verb movement from the embedded VP. More concretely, the empty verb lacks the ability to assign structural case; therefore, it requires a lexical verb to support in order to assign case to the entire gerundive construction. According to the analysis proposed in the present study, a phrasal compound like
duk-syu is formed at the lexical-syntactic interface and so its VO internal structure is still visible to syntax. In other words, after the verb duk combines with syu "book" to form the compound, the verbal constituent is still allowed to separate from the nominal constituent. So as required by the Case Filter (Chomsky 1981), the verb duk in the embedded VP is triggered to move out of its original position to fill the empty transitive verb, enabling this phonetically empty transitive verb to assign case to the entire gerundive construction and derive the surface structure as in (38).

Recent development in generative syntax allows us to keep the merit of C.-T. Huang’s analysis (1992) and, at the same time, reduces its structural complexity. His idea of verb-raising in Chinese can be captured in the light of the current syntactic analysis of light verb. The empty transitive verb proposed in his analysis can be treated as light verb. More specifically, the present study argues that the VP Shell analysis discussed in section 4.1.4 provides a minimally complex and straightforward explanation to the separation phenomena demonstrated by Cantonese phrasal VOCs as in (38) and (39). The tree representation under C.-T. Huang’s approach in (42), for instance, can be revised as the one in (43) which makes use of the VP shell (as discussed in Larson 1988, 1990 and Hale and Keyser 1993):
From the VP shell analysis above, one can see that the verb *duk* in the inner VP moves up to adjourn to the empty head in the upper VP. The motivation for the verb movement is to satisfy the Case Filter (Chomsky 1981). The light verb *∅* is affixal in nature and has to be supported by a lexical verb.

This analysis lends support to our assumption that phrasal VOCs in Cantonese are morphological in nature. Specifically, it is not the case that the internal structure of the phrasal VOC is being intervened by other constituents. Rather, it is the verbal constituent that undergoes verb-raising, resulting in a surface word order in that the VOC appears to be separated.
4.4.2 Positions of Aspect Markers

This section is concerned with the different position of aspect markers with respect to lexical VOCs and phrasal VOCs. Consider the case of phrasal VOCs first. Aspect markers appear between the verbal and the nominal constituents of a phrasal VOC, as exemplified in (44) below:

(44) a. Ngo mou hoi gwo dou.
I have not open ASP knife
“I’ve never had an operation.” or “I’ve never operated on others.”
b. Jisaang tung ngo hoi zo dou.
doctor for I open ASP knife
“The doctor has operated on me.”
c. Jisaang hoi gan dou.
doctor open ASP knife
“The doctor is doing an operation.”
d. *Ngo mou hoidou gwo.
I have not open-knife ASP

e. *Jisaang tung ngo hoidou zo.
doctor for I open-knife ASP

f. *Jisaang hoidou gan.
doctor open-knife ASP
From the examples above, it is found that the constituents of phrasal VOCs are separated by aspect markers in signaling different aspects. Placing the aspect markers after phrasal VOCs are not acceptable. The case of lexical VOCs, however, is just the opposite. Aspect markers cannot enter into the compounds. They must appear after the compounds. Consider the examples in (45) below:

(45) a. Keoi zinggu gwo naampangjau houdo ci.
    she make-poison ASP boyfriend many times
    “She has teased her boyfriend many times.”

b. Gamjat ngo zinggu zo houdo jan.
    today I make-poison ASP many person
    “I have teased many people today.”

c. Go lousi zinggu gan baan hoksaang.
    CL teacher make-poison ASP CL student
    “The teacher is teasing the students.”

d. * Keoi zing gwo gu naampangjau houdo ci.
    she make ASP poison boyfriend many times

e. * Gamjat ngo zing zo gu houdo jan.
    today I make ASP poison many person

f. * Go lousi zing gan gu baan hoksaang.
    CL teacher make ASP poison CL student
The difference in the grammaticality of sentences in (45) indicate that aspect markers can only appear after but not inside lexical VOCs. From the illustrations in (44) and (45), one can see that the positions of aspect markers with respect to lexical and phrasal VOCs are completely different. The present study argues that this difference regarding aspectual marking of the two types of VOCs can be accounted for if we assume that aspect markers in Cantonese are not lexical items but are intrinsic verb features.

Following Gu's (1995) study on Chinese aspectual marking, this study suggests that aspect markers in Cantonese are just like Chinese aspect markers which are base-generated as verb suffixes. Gu's proposal is that the licensing of aspectual features can be best explained neither by overt verb movement nor by lowering of aspect marker plus LF movement of the verb complex, but by LF feature checking. She argues that aspect markers are not projected as lexical items but exist as intrinsic features of verbs in the lexicon. In the lexicon, a verb is stored with it a bunch of intrinsic V-features. These features must be checked off when the verb is projected. In her analysis, Gu (1995) makes use of the Checking Theory, the Procrastinate principle and the Full Interpretation principle to explain the licensing conditions for the aspectual features. Specifically, when a verb is projected from the lexicon, the aspectual features which are part of the verb must be licensed for Full Interpretation. Otherwise, the derivation concerned will be crashed. As the aspectual feature in Chinese are weak features, the checking process can be "delayed" (i.e. Procrastinate) until LF. Finally, these
aspectual features will be checked off at LF and appear morphologically with the verb as V-zhe, V-le and V-guo.

On the basis of the proposal made by Gu (1995), the present study argues that aspectual licensing in Cantonese involves the same procedures. As a result, in the sentences (44a) to (44c) the aspect markers will not be analyzed as items intervening the internal structure of the phrasal VOC hoidou. Instead, these aspect markers are originally intrinsic verb features which are part of the verb in the lexicon. When the verb is projected to LF, these intrinsic verb features are checked off for Full Interpretation. Eventually, they are realized morphologically as in (44a) to (44c).

What is left unexplained is the aspectual marking of the lexical VOCs. Assuming aspectual features are intrinsic features of verbs, one expects that aspect markers will also be found within the internal structure of the lexical VOCs, following the verbal constituent. But the observation is just the opposite—aspect markers appear after the lexical VOCs, following the nominal constituent as shown in sentences (45a) to (45c). This seems to contradict the analysis proposed by Gu (1995). However, we believe that the realization of aspect markers in the rightmost boundary of the lexical VOCs is, in fact, the expected result under the analysis of Cantonese lexical VOCs proposed in the present study. In section 4.3.2 where the formation of lexical VOCs in Cantonese is discussed, it has shown that after the $V^2$ is combined with the $N^0$, a completely new lexical verb is formed. This new verb has nothing to do with its
constituent verb; it has its own lexical specification with its unique lexical information encoded. In other words, this new verb is an entirely new morphological item with a unique set of intrinsic features, including the aspectual features we are now concerned with. The internal structure of the compound is not relevant to the level of syntax and the levels thereafter. Consequently, when the aspectual features of a lexical VOC are checked off at LF, LF will simply treat the compound as a lexical verb without making reference to its internal structure. Therefore, the aspect markers must appear at the rightmost boundary of the compounds. This is exactly what we have observed with respect to the aspectual marking of the lexical VOC zinggu “make-poison = to tease somebody” in (45). Aspect markers appear after the compounds, obtaining expressions like zinggu-gwo, zinggu-zo and zinggu-gan as shown in (45a) to (45c) respectively.

In sum, the observations concerning aspectual marking of VOCs confirm the analysis we proposed in the present study—the compounding process of lexical and phrasal VOCs take place at different interface levels.

4.5 Summary and Theoretical Consequences

The main task of this chapter has been to provide an account for the different syntactic behavior of lexical and phrasal VOCs in Cantonese. Our analysis argues that the formation of lexical VOCs and phrasal VOCs is better
explained by an interactive approach such as Parallel Morphology than by a purely syntactic model (e.g. Baker 1988a) or a purely morphological one (e.g. Bisetto and Scalise 1999). Specifically, our study shows that although both types of VOCs are morphological in nature, they are formed at different levels of the grammatical system: lexical VOCs are formed at the lexical-semantic interface while phrasal VOCs are formed at the lexical-syntactic interface. The present analysis attributes the difference of the syntactic properties of the two types of VOCs to this difference in the levels of formation. Lexical VOCs are formed at a level closer to the lexicon and thus, their constituents are not affected by syntactic operations. In contrast, phrasal VOCs are formed at a level which is closer to the syntactic level and so their internal structures are visible to syntax. As such, our analysis is contrary to what has generally been assumed in lexicalist theories which hold that internal structure of morphological formations are irrelevant to syntax and requires such theories to be modified so as to permit the internal structure of a morphological item to be accessible outside the domain of word itself.

The theoretical consequences of the present analysis can be viewed from two perspectives: (i) the relevance between the findings of the present analysis and the question of the autonomic status of morphology and (ii) the contribution of the present study to the understanding of the relation between morphology and syntax. Consider the issue on the autonomy of the morphological component first. One of the important assumptions in this study is the dichotomy of lexical
and phrasal VOCs and the way they are derived. This aspect of the analysis is based on a particular approach proposed by Levin and Rappaport Hovav (1995). Given that two lexical representations are assumed to be associated with each verb, their analysis successfully accounts for the different morphosyntactic phenomena such as causative alternation in English. Our analysis on Cantonese VOCs has shown that the two interface levels corresponding to the two lexical representations (i.e. the lexical-semantic representation and the lexical-syntactic representation) can account for a number of otherwise unexpected language facts concerning Cantonese VOCs. This can be served as strong evidence showing that morphology is an independent module, because the two interface levels are presyntactic and are presumably included in the morphology.

The autonomic status of morphology has been established. But it does not imply that there is no connection between the morphological and syntactic components. The VOCs in Cantonese studied in this thesis illustrates in a very clear fashion the problem that syntactically separable compounds pose for current theories of morphology—the constituents of a morphological formation can be syntactically separated, but the item itself behave like a word in general. Our analysis argues that there exists a type of phrasal compound which shows a limited degree of separability. The crucial idea about the nature of these compounds is the level at which they are derived. We argue that they are formed at a level where syntax has accessibility. In other words, we claim that syntax can "penetrate" into morphology. So for the question of the relation between
morphology and syntax, our analysis lends support to the assumption that the two components exist as independent module but can interact with each other.
Chapter Five  Concluding Remarks

The present study, which has drawn on data mainly from Cantonese, is an attempt to answer the questions of the nature and the behavior of a particular type of compound, namely the verb-object compounds (VOCs). In conducting this study, we are seeking a theoretical analysis to account for a range of empirical phenomena in relation to Cantonese VOCs. Specifically, we attempt to answer the following research questions (as set out in Chapter One):

1/ What are the characteristics of VOCs in Cantonese?

2/ Why do Cantonese VOCs exhibit both lexical and phrasal properties?

3/ What is the status of morphology in the grammatical system?

4/ What is the relation between morphology and syntax?

We have first discussed the unique properties a compound has that sets it apart from either a purely morphological structure (a word) or a purely syntactic structure (a phrase). Then we have shown that although no single criterion can be used to define what a compound is, we have found that the Lexical Integrity Hypothesis can be used as a primary criterion for defining compounds. We then move on to the focus of this
study: Cantonese VOCs and review existing analyses on this particular type of compounds. We have evaluated the merits and demerits of different approaches to the investigation of the nature and the behavior of VOCs. From our point of view, none of these analyses can adequately account for the properties of the two types of VOCs we proposed in the present study, namely the *lexical VOCs* and the *phrasal VOCs* as defined below:

**Lexical VOC**

A lexical verb-object compound is a morphological formation which consists of a verbal constituent and a nominal constituent and observes the LIH.

**Phrasal VOC**

A phrasal verb-object compound is a morphological formation which consists of a verbal constituent and a nominal constituent. Its internal structure is visible to syntax.

Finally, we have shown that by acknowledging the existence of the two types of VOCs, we are able to explain the nature and the formation of VOCs and account for the difference in the behavior between the two types of VOCs in Cantonese.

The analysis presented in this study relies on (i) the identification of two types of Cantonese VOCs: *lexical VOCs* and *phrasal VOCs*, and (ii) the assumption that word formation process can take place in different levels of the grammatical organization.
Two types of VOCs in Cantonese are shown to be derived via essentially the same mechanism. The difference in their formation processes is that they are formed at different levels, as illustrated in the following schema:

<table>
<thead>
<tr>
<th>Syntax</th>
<th>Formation of Phrases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Lexical-Syntactic Interface</strong></td>
<td>Formation of Phrasal VOCs</td>
</tr>
<tr>
<td><strong>Lexical-Semantic Interface</strong></td>
<td>Formation of Lexical VOCs</td>
</tr>
<tr>
<td><strong>Lexicon</strong></td>
<td></td>
</tr>
</tbody>
</table>

From the schema above, it can be seen that both lexical VOCs and phrasal VOCs are formed at a level prior to syntax with phrasal VOCs formed at a level closer to syntax. We claim that it is this difference in the level of formation that leads to the differences in the behavior of the two types of Cantonese VOCs. To be more specific, a lexical Cantonese VOC like *gwaan-sam* “close-heart = to care somebody or something” is formed at the lexical-semantic interface, so when it enters syntax it will be treated as one single unit, a lexical item $X^0$. The LIH requires that no syntactic rules can apply to any subpart of a lexical item. That explains why lexical VOCs do not exhibit syntactic properties. In contrast to lexical VOCs, phrasal VOCs in Cantonese show various syntactic properties. As shown in the scheme above, phrasal VOCs such as *baau-gaak* “explore-blank = to break into somebody’s house” are formed at the lexical-syntactic interface which is higher than the lexical-semantic interface at which lexical VOCs are formed. This assumption is of vital importance in accounting for the syntactic behavior
of phrasal VOCs. Specifically, phrasal compounds are formed at a level which is closer to the domain of syntax; therefore, their internal structures are assumed to be visible to syntax and subjected to different syntactic operations.

After establishing the existence of two types of Cantonese VOCs and accounting for the difference in their syntactic behavior, we are able to explain the puzzle regarding the “dual status” of compounds. This study argues against the claim that a VOC has the status of a lexical item and a phrase at the same time. More concretely, it has provided structural evidence to show that it is the phrasal VOCs (but not lexical VOCs) that behave like a phrase under certain circumstances. The very nature of these phrasal VOC, however, is morphological for they are formed by morphology.

Questions concerning the autonomy of morphology and the relationship between different linguistic modules in the Language Faculty have always been the focuses of study in the field of morphosyntax. Arguments on these issues tend to fall on two extremes. Take the core of our study—morphology—as an example. It has been argued by some linguists that the morphological component is reducible to syntax or phonology or lexicon. However, morphology has also been argued by strong lexicalists to be a completely isolated component which does not have any interaction with other components. The present investigation reveals that the process of compounding cannot be explained by rules or processes of other linguistic components such as syntax. In Chapter Four, it has been shown that an independent morphological component is responsible for this word formation process. This component has its own unique characteristics which cannot be reduced to properties of other modules. The present
study has further shown that the morphological module can also interact with modules which are designated for other linguistic processes.

Our analysis deals primarily with morphosyntactic issues relating to the process of compounding in Cantonese. Due to the limitation of space, the present study cannot probe into the formation of all types of compounds in Cantonese. Nevertheless, it is hoped that the work presented here can serve as a first step towards a comprehensive theoretical account of the nature of Cantonese compounding. Further investigation should be carried out (i) on different types of compounds such as verb-verb compounds and synthetic compounds and (ii) on a wider variety of issues related to the process of compounding, for instance, morphophonological properties of compounds in order to have a deeper understanding of the nature of Cantonese compounding and the role that morphology plays in the grammatical system.
Chapter One

1. Among the proponents of a modular approach to the organization of the grammatical system (the Language Faculty), two different views on the relation between different modules exist: (i) modules are independent and are completely isolated from each other and (ii) modules are independent but are interactive with other modules.

2. This thesis is written within the generative framework. Due to the scope of this study, other approaches such as Lexical-Functional Grammar (LFG) and Head-Driven Phrase Structure Grammar (HPSG) will not be examined.


4. In the present study, when we use the term “independent” or “autonomous” to describe a module or component of the grammar, we mean that the module has its own set of rules and principles and is not reducible to other components. The use of the term does not have any implication on the interaction between the modules concerned.

5. The term “verb-object compound” is used to include those verb-noun combinations in which the noun is not perceived as the object of the verb. A discussion of the choice of this term for the present study will be provided in Chapter Three.

6. The Cantonese data considered in this thesis are based on Cantonese spoken in Hong Kong.

7. There are two difficulties concerning the definition of compound. First, different studies employ different terminology to refer to the constituents in a compound. Many studies use the term “word” (Selkirk 1982, Spencer 1991, Fabb 1998). But Matthews (1991) uses “lexeme” instead. In the English grammar written by Quirk et al (1986:1567), a compound is defined as “a lexical unit consisting of more than one base and functioning both grammatically and semantically as a single word”. Both
Tang (1991) and Matthews and Yip (1994) state that a compound consists of two or more independent morphemes to form a single new word. Since technical terms like “word”, “lexeme” and “morpheme” are different (at least, slightly) in meaning, one cannot be certain whether these linguists are referring to the same thing or not. Another difficulty ties with the definition of word itself. Linguists have not agreed on the definition of word. So when one says that a compound is a word or consists of words, one is basing the claim on a concept which has not reached a consensus in the field yet. That in turn complicates the task of defining a compound. We will return to these issues in Chapter Two where we will have a detailed discussion on the notions of word, compound and phrase.

8. The romanization system for Cantonese adopted in the present study is Jyutping, a system developed by the Linguistic Society of Hong Kong. Tones are omitted.


10. Gu Yang (personal communication) points out that in Mandarin Chinese it is possible to say zou ni houmen “enter your backdoor = to attain one’s goal by improper ways or personal connection”. I have checked the corresponding usage with native speakers of Cantonese. They say that neither zou nei haumun “enter your backdoor” nor zou nei ge haumun “enter your backdoor” (intended meaning of both expressions: to attain one’s goal by improper ways or personal connection) is acceptable in Cantonese.

11. This representation is borrowed from Fabb (1998). XY%Z is to be interpreted as: [XY] is a compound of word class Z.

12. Selkirk (1982) proposes that the generation of English compounds can be captured by the following set of rewriting rules:

\[
\begin{align*}
N & \rightarrow N/A/V/P \quad N \\
A & \rightarrow N/A/P \quad A \\
V & \rightarrow P \quad V
\end{align*}
\]

In the schema above, the labels N, A, V and P stand for noun, adjective, verb and preposition respectively.

13. They are also called coordinate or dvandva (a Sanskrit term meaning “pair”) compounds.

14. In this section, we review the major approaches regarding the study of the relation between morphology and syntax. There are, of course, other approaches in the literature, for instance, semantic approach such as Fanselow’s (1988) study.

15. This section cannot include all the varieties of lexicalism reported in the literature.

16. Radford (1997:6) relates the idea of minimalism and the learnability as follows:

[Li]inguistic theory should provide grammars which make use of the minimal theoretical apparatus required to provide a descriptively adequate characterization of linguistic phenomena...to minimize the acquisition burden placed on the child, and thereby maximize the learnability of natural language grammars.

17. In the literature “noun incorporation” has been used to refer to (i) a structure like *Yede a-seuan-mū-ban.* “You saw that man.” in Southern Tiwa and (ii) an operation deriving the structure given in (i).

18. Sadock (1985, 1987, 1988, 1991) proposes a theory which he calls Autolexical Syntax to capture the fact that certain structures, for instance, noun incorporation, need to observe both morphological and syntactic principles. According to his analysis, constructions can be simultaneously assigned a morphological representation and a syntactic representation, and these two representations do not need to match. Sadock’s theory deals with the morphological and syntactic components at the same time. Therefore, it can be regarded as a parallel treatment to the study of the relation between morphology and syntax. However, his theory does not assume that morphology can apply at several or all levels of representation (Spencer 1993). This is different from the central idea of parallel morphology which argues that word formation can take place at various levels of the grammatical system; therefore, Sadock’s theory will not be included in the present discussion of parallel morphology.

19. The classification of compounds in (17) does not exhaust all types of compounds discussed in the literature. (17) aims at offering a general picture with respect to the productiveness of the process of compounding in Cantonese.

20. Since the authors of some references cited in this study have the same last name, we will include the initials of these authors when their works are referred in order to avoid confusion. For instance, the present study has cited an article written by Huang Cheng-teh James in 1984, and an article written by Huang Shuan-fan also in 1984. For the sake of clarity, in the main text of this thesis, we will cite their names in the following way: C.-T. Huang (1984) and S.-F. Huang (1984).


Chapter Two

1. The relation between the concepts of character, morpheme and word is very complex in Chinese and Cantonese (Killingley 1993:12). Although the term 字 in Chinese and Cantonese has the same meaning as the term “word” in English, for example, 字數 means “word count”, the term 詞 (but not 字) is used to refer to the concept of word in linguistic studies (Chao 1947, Zhu 1982, Tang 1991, Hu 1992). The term 字 simply refers to “character”, i.e. every monosyllabic unit (Chao 1947). The table below indicates the Chinese equivalents of several closely related linguistic terms used in morphosyntactic studies:

<table>
<thead>
<tr>
<th>ENGLISH</th>
<th>CHINESE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Character</td>
<td>字</td>
</tr>
<tr>
<td>Morpheme</td>
<td>語素</td>
</tr>
<tr>
<td>Word</td>
<td>詞</td>
</tr>
<tr>
<td>Compound</td>
<td>複合詞</td>
</tr>
<tr>
<td>Phrase</td>
<td>詞組/短語</td>
</tr>
</tbody>
</table>


2. Different linguists have tried to define the notion of word from different perspectives. As Packard (2000) notes, in the literature we have orthographic word, semantic word, phonological word, morphological word, syntactic word, psycholinguistic word and so on. This reflects the multidimensional properties of word and, at the same time, the difficulty in giving a precise definition for this notion.

3. For a detailed analysis of Mohawk, see Baker (1988a).

4. An exception of this generalization is aspectual marking. Aspect suffixes in Chinese (and Cantonese) must attach to V^0; therefore, they can help determine the boundary of verbs.

5. Blackbird is a particular species of bird. A female blackbird is brown in color.

6. See endnote (4) in Chapter Two.

7. The contrast between the stress pattern of a phrase and a compound is well-recorded in the literature. In the SPE model (Chomsky and Halle 1968), for instance, the Nuclear Stress Rule places main stress on the rightmost constituent of a phrase, and
the Compound Stress Rule places the stress on the leftmost element of a compound.

8. Cook and Newson (1996:135) hold that "phrases must be 'endocentric': a phrase always contains at least a head as well as other possible constituents." In other words, a structure which lacks a head cannot be a phrase.

9. We focus on C.-T. Huang's (1984) work here, because his analysis provides a very comprehensive discussion on the criteria of compoundhood proposed in Chao's (1968) renowned grammar of Chinese.

10. The romanization system for Chinese adopted in the present study is Pinyin, the most widely used transcription system for Chinese data in the literature.

11. Li and Thompson (1981:80) point out that for both idiomaticity of meaning and separability of constituents, the behavior of verb-object compounds form a continuum:

   [(I)d]iomaticity and separability of the verb-object compound cannot be predicted on a regular basis. Some verb-object compounds are highly idiomatic; some, less idiomatic; some, not very idiomatic. Similarly, some verb-object compounds are completely separable; some are separable to a certain degree; others are almost like a regular verb-plus-object phrase in terms of separability.


13. For instance, in her study of the syntax of words, Selkirk (1982:70) proposes the Word Structure Autonomy Condition below:

   No deletion or movement transformation may involve categories of both W-structure and S-structure.

   Di Sciullo and Williams (1987: 65) formulate this notion of lexical integrity as the Thesis of Atomicity:

   Words are "atomic" at the level of phrasal syntax and phrasal semantics. The words have "features", or properties, but these features have no structure, and the relation of these features to the internal composition of the word cannot be relevant in syntax.

   In this section, we will propose our version of the Lexical Integrity Hypothesis.

14. Tang Wai-lan Gladys (personal communication) provides an alternative analysis of the ungrammaticality of (22a) and (22b). She points out that the observations that classifiers are not compatible with the gaa in ceotgaa may be due to semantic constraints. More concretely, ceotgaa is a compound, so the nominal constituent (i.e. the "object") is nonreferential. In (22a) and (22b), the classifiers go and tau pick out the referent of the entity in question; therefore, both examples are ungrammatical.

15. Wang (1994) proposes a hypothesis, namely the Anaphoric Island Effect of Word Formation to account for the ungrammaticality of sentences like (23b) and (24b).
Interested readers are referred to the original work and his later study (1998) which has reformulated the hypothesis to a universal constraint on grammar.

16. We only aim to give a general characterization of the concept: word, i.e. lexical structure. Arguments related to the fine distinction between morphological and syntactic words will not be included in the present discussion. Interested readers are referred to Di Sciullo and Williams’s (1987) work which is one of the most serious treatments of the theory of wordhood in the literature.

17. See also the discussion in 2.4.2.

18. Bybee uses Pawnee as an illustration. In Pawnee, nouns which are regularly incorporated into verbs include those referring to body parts, natural phenomena, foods and cultural products. Personal names of individuals or tribes, kinship terms, names of animals and names of particular species are not usually incorporated. Compounding is not subject to restrictions of this kind.

19. Fabb (1998:70) notes that the bound morpheme in a compound should not be treated as a real affix, because the unattested part in compounds “fail to resemble affixes morphologically (they are relatively unproductive compared to most affixes), and there is no good evidence or phonological grounds for considering them to be affixes. They are also unlike affixes, semantically; judging by their contribution to the word’s meaning, they have lexical rather than grammatical meanings.”

20. Lexical compounds, but not phrasal compounds, are constrained by the LIH.

21. The distinction between the separable and inseparable compounds fits well with different languages including Chinese, Cantonese and Dutch. See Wang (1994, 1998) for Chinese compounds; Matthews and Yip (1994) for Cantonese compounds and Ackema (1999) for Dutch compounds.

22. Note that in the literature the term “phrasal compounds” is used to refer to compounds that contain phrasal expressions, for instance, in Roeper’s (1988) study on head-to-head movement and Lieber’s (1992) Deconstructing Morphology. This definition of the term is different from the one adopted in this thesis. In the present study, the term “phrasal compounds” reflects the behavior but not the structure of the compounds concerned. It refers to those compounds which exhibit phrasal properties, in contrast to lexical compounds which show only lexical properties.

Chapter Three

1. Wang (1994:47) has provided a list of Chinese data to support his idea that the
“object” of a verb-object compound may not necessarily be a direct object or patient. Some of his examples are presented here for illustration:

**Patient:**
qi-ma  ride-horse  “to do horse-riding”
xiong-che  repair-car  “to do car-repairing”

**Theme:**
diao-yu  hook-fish  “to go fishing”
zuo-meng  make-dream  “to dream a dream”

**Location:**
zhu-yuan  stay-(in)-hospital  “to be hospitalized”
guang-jie  walk-(in)-street  “to go window-shopping”

**Source:**
chu-jia  leave-(from)-home  “to become a priest”
chu-tu  leave-(from)-earth  “to be unearthed”

**End-point:**
deng-lu  step-(onto)-land  “to be engaged in”
tiao-he  jump-(into)-river  “to jump into a river”

**Goal:**
kao-benzi  test-(for)-licensure  “to have a driving test”
tao-sheng  escape-(for)-life  “to escape for survival”

**Instrument:**
kai-dao  cut-(with)-knife  “to have an operation”
tiao-san  jump-(with)-parachute  “to parachute”

**Cause:**
yang-bing  nurture-illness  “to be on a sick-leave”
tao-huang  flee-famine  “to run away from famine”

From the data above, one can see that a wide range of semantic relations is present between the verbal and the nominal constituents. Wang, therefore, argues that it is problematic to use the term “verb-object compounds” to cover all these verb-noun structures.

2. See endnote (22) in Chapter Two.
3. The present study focuses on the Cantonese VOCs which function as verbs. We will occasionally refer them as verbal VOCs. But notice that we are not making any reference to “verbal compounds” which is a term commonly used to refer to synthetic compounds whose “head is derived by affixation from a verb, such as truck driver, in which truck appears to be an argument of the (stem) verb drive.” (Spencer 1991:309).
4. Yip Choy-yin Virginia (personal communication) brings this property of verb-object compounds to my attention.
5. Wang (1994) relates the grammaticality of constructions like (6) in Chinese with specificity. See original work for relevant discussion.
6. Matthews and Yip (1994:151) note that verb-object compounds may be passivized in the form of indirect passive. For example:

---

187
Keoidei loeng go sengjat bei jan *gong-haang-waa.*
they two CL often by people say-idle-talk
"Those two have people gossiping about them all the time."

Keoi geng bei jan *caau-jaujyu.*
he/she fear by people fry-squid
"He/She is afraid of getting the sack."

In the sentences above, the nominal constituents of the VOCs are not moved. Since this section concerns the movement of the "object" of verb-object compounds, these passive constructions will not be discussed.

7. It should be noted that for some VOCs, the semantic object of the compound can appear in more than one position. Consider the examples in (i) and (ii) below:

(i) Variations of Transitive VOCs - Type (I)
   a. Ngo sengjat *zinggu* ngo baan tunghok.
      I often make-poison my class classmate
      "I often tease my classmates."
   b. Ngo baan tunghok sengjat bei ngo *zinggu.*
      my class classmate often by I make-poison
      "My classmates are often teased by me."

(ii) Variations of Transitive VOCs - Type (II)
   a. Ngo jiu *zin* keoi *pei* caak keoi gwat
      I must fry he/she skin demolish he/she bone
      "I must kill him/her!"
   b. Ngo jiu zoeng keoi *zin* pei caak gwail
      I must ZOENG he/she fry skin demolish bone
      "I must kill him/her!"

In (ia), the object *ngo baan tunghok* "my classmates" follows the VOC *zing-gu* "tease". This object moves to a pre-verbal position in indirect passive construction as in (ib). In (iia), we find that one can accommodate the object *keoi* inside the compounds *zin-pei* and *caak-gwat*. But one can also introduce the object by *zoeng* as in (iib). These variations will not be discussed in the present study.

8. For the analysis of ditransitive verbs, readers are referred to the original work, C.-T. Huang’s (1982) study.

9. As a matter of fact, the Chinese equivalents of (28) and (29) are also grammatical. This casts further doubt to the validity of the PSC.

10. Note that in Packard’s (2000) work on Chinese morphology, the discussion on the nature of VOCs relies essentially on C.-T. Huang’s (1984) analysis. Therefore, Packard’s (2000) study on VOCs will not be reviewed.

11. In fact, with C.-T. Huang’s (1992) analysis on verb movement, the PSC can be abandoned.

12. Wang (1994) uses the term “verb-complement compounds” (VCCs) to refer to verb-object compounds.

13. Li and Thompson (1981:232) define the delimitative aspect as “doing an action ‘a
"little bit", or for a short period of time."

14. Li and Thompson (1981:29) claim that non-volitional verbs do not undergo delimitative aspect reduplication, for example, \textit{wang} “forget”.

15. There are, of course, other means to express delimitative aspect in Cantonese. The reason for us to choose \textit{haa} as the focus of the discussion is that its position with respect to the two types of VOCs can be used as a diagnostic test for the lexical-phrasal VOC distinction.

\textbf{Chapter Four}

1. Spencer (1991) proposes a model for the morphological module, arguing that it is autonomous of other levels of representation and has its own set of elements and principles of combination. But this module runs parallel with other grammatical components and can interact with them. A diagrammatic representation of the model is given below (taken from Spencer 1991:455):

\begin{table}[ht]
\centering
\begin{tabular}{|l|l|}
\hline
\textbf{Morphology module} & \textbf{Lexicon} \\
\hline
Derivational morphology & List of lexemes, idioms etc. \\
Paradigmatic word formation & Idiosyncratic word forms \\
Lexical compounding & \\
Anticausatives etc. & \\
Monoclausal causatives & \\
Chukchee NI & \\
Adjectival passives & \\
Synthetic compounds? & \\
Regular inflection etc. & \\
Hebrew construct state nominals & \\
Synthetic compounds? & \\
Bicausal causatives & \\
Verbal passives & \\
Mohawk NI & \\
Pronominal clitics & \\
Phrasal affixation (including English POSS aux. clitics, a/an) & \\
Kwakwala, Polish clitics & \\
Post-syntactic compounding & \\
\hline
\end{tabular}
\end{table}

\textbf{Syntax}

\begin{center}
\{ \text{D-structure} \\
\text{S-structure} \\
\text{(LF)} \}
\end{center}

ECP

Binding theory etc.

\textbf{Phonology}

\begin{center}
Prosodic domains \\
Phonotactic constraints \\
Phonological rules
\end{center}
2. All the Hebrew examples cited in this section are taken from Borer (1988).
3. See endnote (12) in Chapter Three.
4. “ACC” represents accusative case.
5. Note that many of those who advocate a modular approach to linguistic studies in general, and the autonomic status of morphology in particular, would not accept the degree of radicalism maintained by “strong lexicalists”. Proponents of an independent morphological module (Di Sciullo and Williams 1987, Spencer 1991, Borer 1997, among others) notes that morphology can interact with other components, such as syntax. See also endnote (1) in Chapter One.
6. Spencer (1991:344) notes that the syntactic effects observed in compounds “are the result of essentially morphological rules or principles permitting argument structure to be accessible outside the domain of the word itself, a limited ‘leak’ in the Strong Lexicalist Hypothesis.”
References


Cheung, Hung-nin. 1972. Xianggang Yueyu Yufa de Yanjiu [香港粵語語法的研究 Cantonese as Spoken in Hong Kong]. Hong Kong: The Chinese University of Hong Kong.


Cambridge, MA: MIT Press.


Liang, Tao. 1999. *Hoenggong Gaujyu* [香江舊語 Hong Kong Old Sayings]. Hong Kong: Subculture Co.


Ng, Soeng-gwong. 1996. *Gwongzauwaa Jat Jat Jat Tai* [廣州話一日一題 Cantonese 365]. Hong Kong: The Commercial Press.


Wu, Hao. 1994b. *Zukmanfaa Jyujin II* [俗文化語言 II *Language of Vulgar Culture II*]. Hong Kong: Subculture Co.


