國立交通大學

外國語文學系外國文學與語言學碩士班

碩士論文

漢語和英語中的數詞組對比分析

An Analysis of Numerical Expressions in Chinese and English

研 究 生: 葉怡君

指導教授: 劉辰生 博士

中華民國九十八年七月

漢語和英語中的數詞組對比分析

學生: 葉怡君　　　　　　　　　　　　　　指導教授: 劉辰生教授

國立交通大學外國語文學系外國文學與語言學碩士班

摘要

本篇論文的基本假設視數字爲總和的概念。數詞詞組之所以獨特，正是因爲其表達方式與真實語言中的文法不同，且數詞詞組的表達方式，主要經由算術上的加法所組成。在相關領域的研究中，最著名的爲語言學家 Hurford 所稱‘數字的文法’系統，他嘗試整合表面的文法和真實語言裡的句法學。藉此，這篇論文將會詳細解釋在 X-bar 理論結構下的數詞表達方式。爲此，數詞詞組首先必須被視爲與文章脈絡無相關且連續並排而成的字串。此外，數詞詞組在句法學上，必須被納入抽象名詞的範疇中，即使再小的數詞也是如此。這篇論文主張數詞詞組的表達方式，由累計的對等連接詞組(&P)，進而形成連接詞詞組(ConjP)。具體來說，數詞詞組表達方式之主張根據 Munn 所提倡的&P 分析，或稱爲 Boolean Phrase (BP)而來，而它所累計的對等連接詞組，則是從右邊節點加接進來 (right node adjunction)。此結構充分適用於英文的數詞詞組，因爲對等連接詞‘and’在 PF 上是不可或缺的；但對於中文的數詞詞組而言，&P 並不能完全符合且通用，其主要原因爲中文數詞表達中的對等連接詞‘又’是隱藏的。最後，這篇論文試圖證明&P 這樣的分析方式，將適用於中文和英文的數詞詞組表達方式，並強調就算在中文數詞詞組隱藏對等連接詞(&)的情況下也適用。

關鍵字: 抽象名詞，對等連接詞，X-bar 理論，對等連接詞結構，
附加結構

An Analysis of Numerical Expressions in Chinese and English

Student: Yi-Chun Yeh Advisor: Dr. Chen-Sheng Liu

Institute of Foreign Literatures and Linguistics

National Chiao Tung University

Abstract

The fundamental assumption of this thesis is that numbers are sums. This means that numerical expressions are primarily composed of notations for the arithmetical operation of addition that exists outside the ordinary syntax of language. Hurford (1975, 2003), the most eminent researcher of numerical expressions, calls this external system "the grammar of numbers," and he makes little attempt to integrate this external grammar with the ordinary syntax of language. This thesis, however, does attempt to account for numerical expressions within the framework of standard X-bar theory. To do this, it must be recognized that numerals exist as word strings that are free of context and that are arranged as paratactic concatenations. Moreover, it must be recognized that all numerals, even small lexical numerals, should be categorized syntactically as abstract nouns. Furthermore, when numerals are combined through addition they form nominal compounds. It follows, then, that co-ordination offers the best syntactical interpretation of numerical expressions. This thesis argues that numerical expressions can be configured as conjunction phrases (ConjP) of a specifically cumulative type called "and" phrases (&P). Specifically, it is argued that numerical expressions are best configured by following Munn's (1993) analysis of &P, or what he calls the Boolean Phrase (BP), as right node adjunction. This configuration works well for English numerical expressions, because the conjunction "and" is integral to such expressions at PF, but the &P configuration is problematic for Chinese numerical expressions, because the conjunction *you* that heads the phrase remains covert. In the end, this thesis suggests evidence that the &P analysis does work for both English and Chinese numerical expressions, despite the apparent problem of the covert & in Chinese numerals.

Keywords: abstract noun; co-ordination; X-bar theory; conjunction phrase (ConjP); adjunction

誌謝

能夠順利完成論文，首先，我要感謝我的論文指導教授劉辰生老師的悉心教導，他在語言學上有著豐富的學識，而且教導學生非常有耐心，對於句法的熱忱更讓我很佩服。平時他就很關心學生，在各方面也很會替學生著想，平常和老師討論問題時，他都會明快而敏銳的點出一些方向，讓我不至於像無頭蒼蠅一樣的摸索，這篇論文的發想與發展都因為老師的指點而更臻完善，非常感謝劉老師。

其次，我要感謝我的論文口試委員，林宗宏老師和蔡維天老師，感謝他們在百忙之中抽空擔任我的口試委員，對於論文的論點與分析都提出精闢的見解，更提供我修正的方向，在他們的斧正之下，我才能針對癥結所在逐一提出修正，讓自己的分析與立論更具說服力。

接著，要感謝陪我一路走過來的同學們：縉雯、雯靜、芳瑩和惠瑜。平時一起討論功課，也會忙裡偷閒的聚餐抒壓，能夠痛快的宣洩壓力，都是因為有這些朋友。特別感謝惠瑜在忙於寫作論文的同時，還會撥些出時間和我討論這篇論文的問題，並提供我一些想法，讓我在論文寫作的路上不再感到孤單徬徨，真的由衷感謝。還有，謦瑜也是身旁一位活潑可愛的學妹，她的率真與活力常帶給我許多歡樂，讓我在繁重的課業壓力之下能盡情歡笑，也謝謝她關心我的飲食、擔心我的健康，真的讓人感到窩心。另外，也要感謝文傑學長，佑慈學姐和靜玉學姐，謝謝他們不吝於提供我建議和想法，也樂於將他們自己的經驗和我分享。

再來，還要感謝我的室友們：小皮、秀涵和宜賢。我很高興自己很幸運地能和你們住在一起，謝謝你們帶給我的歡樂和支持鼓勵，還有謝謝你們在我心情不好時，總是能靜靜的聽我發洩情緒。我也會一直很懷念我們徹夜聊天都不想睡的時光。

當然，我也要感謝加拿大認識的老朋友們，因為你們的鼓勵和不斷的給予我信心，我才能一直在研究這條路上撐下去，也謝謝你們一直相信我能做得到。

最後，要感謝我最最親愛的爸爸、媽媽和我的家人。謝謝爸爸讓我在生活上無後顧之憂。還有媽媽，當初要不是您的堅持，我也無法順利完成碩士學業。這本論文謹代表我內心最深的感謝，我愛你們!

Table of Contents

Table of Contents

CHAPTER 1

INTRODUCTION

1.1 Basic Assumptions

The purpose of this thesis is to propose the idea that numbers are sums. This means that numbers are wholes composed of cumulative parts. Numbers embody a system of notation and operation that exists beyond the grammar of syntax, and yet numerical expressions consist of lexemes that also exhibit syntactical categorization. The matter is further complicated by the fact that the syntactical categorization of numerical expressions is extremely limited. There are neither verbs nor prepositions nor even sentences in numerical expressions. Because of this limitation numerical expressions cannot be said to possess arguments in the normal grammatical sense, nor can they be said to be subject to the case filter rule. Numbers can, however, be seen as nouns and noun phrases, and sometimes numbers seem to behave as if they were adjectives, but it is the position of this thesis that numbers are best categorized as nouns, no matter how they behave. There are simple lexical number nouns, such as "seven," and there are complex constructed number nouns, such as "three hundred and twenty-five," that are nominal compounds.

Because numbers are fundamentally notations of arithmetical – not grammatical – operations, efforts must be made to interface the two systems in any meaningful account of the composition of numerical expressions. The arithmetical operation involved in the organization of numbers is extremely simple. Numbers are sums. James Hurford, the most accomplished theorist of numerical expressions to date, expresses this empirical fact clearly and succinctly: "The value of a number is the sum of the values of its immediate constituents" (1975, p. 11). To translate this truth into grammatical terms we might say that numerical expressions are cumulative statements. Numbers are made by one basic operation: addition. Often there are thought to be two operations involved in the composition of numbers: addition and multiplication, but we should remember that multiplication is only a shorthand form of addition. The complex number 500 is actually composed by adding 100 *plus* 100 *plus* 100 *plus* 100 *plus*

100, but we usually say, for the sake of convenience, that 500 is composed by multiplying 100 *times* 5. In the end, numbers are sums.

Instead of following Hurford's (1975) analysis of numerical expressions as flat structures with ternary branching, I have turned to recent arguments for conjunction phrase (ConjP) interpretations of co-ordination. In particular, I have followed Munn's (1993) analysis of co-ordination as what he calls the Boolean phrase (BP) in which the first conjunct is linked with the merged constituent of the conjunction and the second conjunct through right node adjunction. I am convinced that this additive &P arrangement provides a satisfactory interpretation of numerical expressions within the framework of X-bar theory. At any rate, this analysis works well for English numerical expressions, because these use the overt conjunction "and" in their composition. The only problem now is that Chinese numerical expressions do not use an overt conjunction such as *you* (*yu*, *yehao*) in their composition. The challenge has been to maintain the &P analysis for Chinese numerical expressions, while at the same time justifying a covert conjunction as the head of &P.

1.2. A Preview of the Discussion

This thesis presents five interrelated proposals:

1. Numbers are sums composed of the arithmetical operation of addition (including multiplication, a shortened form of addition).
2. Numerical expressions are best interpreted syntactically as co-ordinate structures linking nouns to produce nominal compounds.
3. Numerical expressions exist as independent, context-free word strings that can be regarded as a singular form of paratactic and concatenated discourse.
4. Numerical expressions can be accounted for within the framework of X-bar theory by regarding them as additive or cumulative conjunction phrases (&P).
5. The &P analysis accounts for numerical expressions in both English and Chinese, incorporating both English "and" as a co-ordinator and Chinese *ling* ('zero') as a place holder.

I will first discuss the intrinsic qualities of numbers as arithmetical notations and syntactical nouns. This will be followed by a discussion and critique of Hurford's classical interpretation of numerical expressions, focusing mainly on his phrasal structure. I will then discuss the characteristics of English and Chinese numerical expressions, paying special attention to the use of "and" in English numerical expressions and *ling* in Chinese numerical expressions. This will be followed by a discussion of the properties of co-ordination as a syntactical category, suggesting that it is co-ordination that best matches and expresses the arithmetical operation of addition that is central to the composition of numbers. I will then present my argument for regarding numerical expressions as additive conjunction phrases (&P) formed, in the manner of Munn (1993). The discussion concludes with an argument for justifying the covert conjunction *you* as the head of &P in Chinese numerical expressions.

CHAPTER 2

THE NATURE OF NUMBERS

2.1 Numerals as Abstract Nouns

Hurford (1975, 1987, 2003) does not categorize numerals syntactically, and this creates a problem for any discussion of the phrasal construction of numerical expressions. Corbett (1978/2000) notes that number, as a categorical feature, is dominantly nominal, though verbs may also be marked for number. Moreover, linguists have often treated simple numbers as adjectives and multipliers and complex numbers as nouns. Nevertheless, Greenberg (1978) observes that numbers *as numbers* – in other words, abstract, context-free numbers – are always nouns. My thesis assumes that Greenberg is correct. All numerals, therefore, must be regarded as nouns. It follows, then, that when numerals combine to form a phrase, the result will be a nominal compound, a syntactic form that is common to both English and Chinese.

It is often taken for granted by linguists that low numbers – simple numbers, lexical numbers – especially those from 1 to 4 behave syntactically like adjectives, while higher numbers, 10 and all the complex numbers that follow, behave syntactically like nouns. But is this generalization true? Or does it represent an almost superstitious misunderstanding? I would like to propose that the grammatical nature of numbers needs to be examined more closely before we accept the common division of numbers into the categories of low adjectives and high nominals.

Greenberg (1978) states that numbers are either concrete or abstract, suggesting that it is possible to think of numbers in two distinct ways. Concrete numbers are derived from discourse. In other words, concrete numbers quantify things. They refer to nouns. They combine with nouns to form phrases. The contextuality of concrete numbers is especially evident in numeral classifier languages such as Chinese where the appropriate classifier must be used with each noun enumerated. Abstract numbers, on the other hand, are created mentally as intellectual concepts. They exist independently and they are not constrained by discourse, contextuality, or reference to nouns. Abstract numbers *are* nouns. Their principal uses are counting by recitation and performing mathematical computations. Sometimes a simple number may have different numeral expressions depending on whether it is concrete or abstract. In Chinese, for instance, 2 is expressed as *er* as the multiplier for 10 in the absolute system, but 2 is usually expressed as *liang* as the multiplier in the contextual system, especially for 100, 1000, and 10,000.

The difference between concrete numbers and abstract numbers has a long history of linguistic implications. The distinction between concrete numbers and abstract numbers is the same as the distinction between concrete nouns and abstract nouns. That is, concrete numbers are physical, because they are fused with nouns, but abstract numbers are not physical -- we cannot see, hear, taste, touch, or smell them -- because they are ideas and they are not fused with any nouns. There can be *three* books or *three* doors or *three* people. The number three is not limited to any one noun. To illustrate the difference between concrete nouns and abstract nouns Poncinie (1993) discusses Aristotle's "double use" of the common term *number*. In *Physics* IV, 11, Aristotle defines number as both something that is counted and something that can be used to count something else. What can be counted is the concrete number, and what can be used to count is the abstract number. This may seem to be a simple distinction, but the more we think about it, the more difficult it is to understand. What happens for example, when we count sheep? If we count three sheep, there is no doubt that the abstract number is *three.* But what is the term *three sheep*? Is it the quantity of the group? Or is it a numbered group? Is it *three*? Or is it *three sheep*? This is the concrete number, and it seems to suggest two interpretations.

Poncinie claims that the totality of the group, having contingent relations to other numbers, is an extensional use of the abstract number, while the numbered group is a concrete term akin to *aggregate* or *class.* What this means, then, is that there are really three kinds of numerical expressions. First, there is the abstract number. But the concrete number can be seen in two different ways: either as a totality referring to a group, or as a totaled group complete with existential features. To understand this split usage of the concrete number, let us return to the sheep. Can we apprehend and comprehend the difference between *three* sheep and *three sheep*? It is not an easy thing to do.

To complicate the matter, we probably intuit that three sheep is a different proposition than seventy-eight sheep. It is easier to regard three sheep as a numbered group of sheep than it is to regard seventy-eight sheep in the same manner, and it is easier to regard seventy-eight sheep as the totality of the group of sheep than it is to regard three sheep in that way. Why is this? To answer we must return to the common linguistic assumption that simple numbers are adjectival and complex numbers are nominal. Menninger (1969/1992) offers an explanation of how this belief probably began. Since only the first four numbers are usually considered to be adjectival, Menninger argues that these numbers correspond to the primitive habit of counting on one's fingers. In other words, low numbers are intimately connected to the physical reality of the entities they quantify. It is easier to see, touch, and imagine three sheep than it is to see, touch, and imagine seventy-eight sheep. This also means that it is easier to regard *three* as a word that describes sheep than it is to regard *seventy-eight* as a word that describes sheep. Thus *three* seems like an adjective, while *seventy-eight* seems like a noun.

Menninger provides a compelling reason why numbers – even low numbers – should not be regarded as adjectives. We might be tempted to claim that "three-ness" or even "seventy-eight-ness" are attributes of the counted sheep, the same as whiteness might be an attribute of these sheep. But there is a vast difference between numbers as attributes and colours as attributes. One sheep can be white, but one sheep cannot be three or seventy-eight. Numbers always refer to totalities, and as such they are always conceptual. In other words, numbers are always abstract, always detachable from the nouns they count. Contrary to popular linguistic belief, this is true of simple numbers as much as it is true of complex numbers. Menninger (p. 11) says a number word – a numerical expression – is "a special kind of word." He means it is an abstract noun that is, by its nature, somewhat mysterious.

Menninger illustrates the importance of abstract numbers by telling how Archimedes proposed measuring the universe by computing the number of grains of sand it would take to fill it. "The whole point of Archimedes' discussion is that even so inconceivably large a number as that of the grains of sand contained in the universe can not only be clearly understood and verbally expressed, it can even be easily exceeded: the limitless progression of the number sequence has finally been recognized!" (p. 141) One of humanity's greatest achievements, equaled only by the invention of writing and reading, was the discovery of the abstract nature of numbers. In grammatical terms, this means that numbers are actually nouns that name ideas or concepts that exist far beyond their connections to particular physical entities.

2.2 Numerals as Nominal Compounds

Ultimately, complex numeral expressions are nominal compounds. This means that in the phrasal structure of numerals two nouns – either simple or multiplier or even complex in form – are combined to form a larger complex noun. As Li and Thompson (1981) point out, both English and Chinese make frequent use of nominal compounds. Examples of English numerical nominal compounds are given in (1).

(1) a. 14: nine + five → fourteen

b. 78: seventy + eight → seventy-eight

c. 130: one hundred + thirty → one hundred and thirty

d. 60: six x ten → sixty

e. 1700: 17 x 100 → seventeen hundred (or, one thousand, seven hundred)

f. 1,000,000: one thousand x one thousand→ one million

When two numerals are combined in English, either through addition as in (1a - c) or through multiplication as in (1d - f), mathematics determines the resulting number, but the syntactical projection of nouns determines the resulting numeral expression. This is very obvious in (1b) where the two numerals are combined in such a way that the only difference between saying the operation of adding the two numerals and saying the resulting sum is the substitution of a hyphen for 'and'. In (1c) the meaning of the operational phrase 'plus' appears as 'and' in the numerical expression. In (1a) and (1d) transformations for *-teen* and *-ty* are involved. I will discuss these transformations presently. In (1f) the underlying operation for the product, is transformed by mathematics into a completely different lexical expression for the nominal compound. In (1e) the nominal compound can be pronounced in either of two ways. Hurford (1975) would prefer the second alternative because it conforms to his "packing strategy" concept whereby numerical expressions should incorporate the highest possible multiplier – in this case, one thousand, not one hundred.

Examples of Chinese numerical nominal compounds are given in (2).

(2) a. 16: *shi* + *liu* → *shi liu*

b. 20: *er* x *shi* → *er shi*

c. 35: (*san* x *shi*) + *wu* → *san shi wu*

d. 855: (*ba* x *bai)* + (*wu* x *shi*) + *wu* → *ba bai wu shi wu*

e. 1, 0370: (*yi* x *wan*) + *ling* + (*san* x *bai*) + (*qi* x *shi*) → *yi wan ling san bai qi shi*

f. 21, 3220 (*er* x *shi* x *yi* x *wan*) + (*san* x *qian*) + (*er* x *bai*) + (*er* x *shi*) →

er shi yi wan san qian er bai er shi

As in English, numeral expressions in Chinese are composed as nominal compounds through the mathematical operations of addition and multiplication. These two basic operations are illustrated clearly in (2a) and (2b). Moreover, the way the operations of multiplication and addition can be further combined is evident in (2c) and (2d). In (2e) and (2f) we notice two important features of Chinese nominal compounds. First, Chinese numerals are arranged in strings of four, unlike English numerals which are arranged in strings of three. For this reason 10, 000 or *wan* becomes an important Chinese multiplier. The second distinctive feature of Chinese numeral expressions is the appearance of *ling*, meaning 'zero', as a place-holder to mark the beginning of the last string of numbers if it begins with a 0. I will discuss the importance of *ling* in much detail as this thesis unfolds.

There are also other differences between Chinese and English numerals resulting from transformations. For instance, the number 16 is expressed with normal word order in Chinese as *shi liu* – literally, 10 + 6. In English, however, the numeral expression for 16 involves switching the normal word order and changing the morpheme 'ten' through inflection to 'teen', resulting in 'sixteen'. These transformations are true of all numeral expressions in English for the numbers 13 through 19, with distinct lexical morphemes for 11 and 12, while in Chinese all the numbers between 10 and 20 are expressed by maintaining both the lexical morpheme *shi* for 10 and the normal word order of 10 + 1 . . . 9. Moreover, in Chinese the numeral expression for 20 maintains both normal word order and morphological form, resulting in the nominal compound *er shi*, representing the multiplicative operation of 2 x 10. Once again, this pattern holds true for 30, 40, and so on, up to the next important lexical morpheme *bai* representing 100. In English, on the other hand, transformation morphologically changes 'ten' to 'ty' in 'twenty', 'thirty', and so on, up to 'ninety'. Finally, a number such as 35 is expressed in Chinese as *san shi wu*, literally 3-10-5, whereas in English the numeral expression is 'thirty-five', combining the 3 x 10 operation into 'thirty', resulting in two morphemes for 35, whereas in Chinese all three basic morphemes are preserved. This is to say that Chinese retains both the base multiplier 10 and normal word order, while English transforms the base multiplier and deletes it in expression.

It is significant to note here that the nominal compound form of numeral expressions occurs because they are constructed in the syntax. In other words, complex numeral expressions are unique. They do not, like simple numerals and multiplier numerals, exist in the lexicon, although, as I will soon argue, the complex numerals 11 – 99 behave, especially in English, as if they were independent lexemes. Is there a limit to the construction of complex numerals as nominal compounds? We have already seen that the ancient Greek philosopher Archimedes believed that it is possible to construct a numeral expression large enough to describe the universe. Given the abstract nature of numbers, this would seem to be entirely possible. Nevertheless, Hurford (1975) points out that numerical expressions need to be well-formed or sanctioned by natural language usage, and extremely large "theoretical" numbers may have no usable expression. Radzinski (1991) agrees and observes that it is the interface of arithmetical operations and grammatical operations that determine which numeral expressions are well-formed or even plausible.

2.3 Numerical Expressions as a Form of Discourse

Isomorphism between syntax and phonology has long been considered dubious, and yet consideration of numerical expressions raises the possibility that it might be necessary to search the interface of these two linguistic forms to fully comprehend how numerals are composed. We have already seen that numerals are abstract, context-free nouns. Verbs (and therefore adverbs) never exist in numerical expressions. Neither do prepositions. Conjunctions are the only other possible syntactic category to be found in numerical expressions. It should be noted, however, that the presence of a conjunction – always cumulative, represented by "&" – is regarded by Hurford (1975. p. 50) as an optional transformation. The conjunction "and" is certainly present in English numerical expressions, but a comparable conjunction, such as *you,* is not overtly present in Chinese numerical expressions. I shall argue, nonetheless, that the conjunction *you* covertly heads an &P in my binary-branching phrasal analysis of numerical expressions. In order to justify this analysis I shall appeal, to a large extent, to phonological data to support my syntactic interpretation.

Once again, Hurford's common sense approach to the construction of numerical expressions cannot be ignored. He says, in describing his analysis (*ibid.*): "I have not made an attempt to integrate the grammar of numerals into a grammar for the rest of the language." We have already seen that the grammar of numerals involves, more than anything else, the creation of sums through the operation of addition (supplemented by multiplication). We have also seen that the creation of sums results in numerals being, syntactically speaking, nominal compounds. But is now time to observe that numerical expressions may also be regarded as independent forms of discourse whose principal characteristic is that they are composed with an absolutely rigid word order arranged paratactically as unique concatenations. Although they are basically nouns or noun phrases, these word strings do not assign case internally, and they do not exhibit any movement whatsoever – making the Empty Category Principle irrelevant. Hurford also mentions that not all lexical items denote objects. In fact, some lexical items denote relationships. This is where & enters the grammar of numerals, as overt "and" in English numerical expressions and as covert *you* in Chinese numerical expressions. Either way, it is important to remember that there is always a distinction to be made between the grammar of numerals and grammar as a whole. As my argument progresses, I attempt to integrate the grammar of numerals into the general grammar, but in order to do this I will have to refer to the fact that the grammar of numerals represents a special form of discourse.

2.4 English Numerical Expressions

In English numerical expressions are divided into strings of three. For example, the number 357 contains a units' place, a tens' place, and a hundreds' place. This first string may be called the hundreds' level. This basic ordering of places is repeated in the next higher level or string of three digits, which may be called the thousands' level. This means that at the thousands' level there are units, tens, and hundreds (of thousands). Moreover, the same pattern holds true for the next higher level, that is, the millions' level. Once again, there may be units, tens, and hundreds (of millions). This system may be expanded to subsequently ever higher levels, practically *ad infinitum*. These facts are illustrated by the table in (3).

(3)	Level of Hundreds	Level of thousands	Level of millions
Units	0-9 of one thousand	0-9 thousands	0-9 millions
Tens	10-99 of one thousand	10-99 thousands	10-99 millions
Hundreds	100 -999 of one thousand	100-999 thousands	100-999 millions

This system may be illustrated, for example, by the number 716, 429, 357. The strings of three digits are conventionally marked by commas in arithmetical notation. At the hundreds' level we see 357; at the thousands' level, 429, and at the millions' level, 716. In the string of three digits at each of these levels there is a units' place, a tens' place, and a hundreds' place.

When these numbers are given numeral expression, each of the levels contains the conjunction "and." Thus, the number in our example is pronounced (without the emphasis suggested by the italics, which are provided for purely lexical demonstration) "seven hundred *and* sixteen million, four hundred *and* twenty nine thousand, three hundred *and* fifty-seven." Furthermore, something else is immediately noticeable from this numeral expression. Not only is "and" repeated at each level or in each string of three digits, but "hundred" is also repeated immediately preceding each "and." This means that each string of string of digits is divided into two parts: the hundreds' place on the left and a combination of the tens' place and the units' place on the right. The conjunction "and" stands in the middle of the division. This arrangement always occurs in English complex numbers, and it suggests a binary phrasal structure of two constituents combined by co-ordination at each level or in each string of three digits. These observations are summarized in the generalizations given in (4).

(4) a. English numerals are arranged in strings of three digits, consisting of a units' place, a tens' place, and a hundreds' place at each successively higher level of strings, beginning with the level of hundreds and proceeding to the level of thousands, the level of millions, and so on, virtually without limitation.

b. At each of the levels the morpheme "and" occurs following the morpheme "hundred." The effect of this is to bifurcate the string of three digits and suggest a phrasal structure with the hundreds' place in the X position and a combination of the tens' place and the units' place in the Y position. Moreover, this suggested binary structure appears to be co-ordinated by "and." Its function is additive.

Not all numbers, of course, contain all the possible places and strings. If (4) is true, what about a number such as 6,022? This number is pronounced "six thousand and twenty-two." In this numeral expression "and" occurs after the word "thousand," not the word "hundred." There is a simple explanation for this. The morpheme "and" occurs after the hundreds' place, but the hundreds' place is occupied by "zero," which is never spoken in English numeral expressions (except as 0 itself). The same might be said of the number 6,002, pronounced "six thousand and two." Here again, "and" occurs after the hundreds' place, and after the tens' place as well, but both these places are phonetically null, though they are still present in LF. Therefore, such examples as these do not invalidate the generalizations in (4). In fact, the higher the number containing "and" but not "hundred" is, the more awkward and possibly ill-formed its expression sounds. For example, 76,000,003 is pronounced "seventy-six million and three," but such an expression sounds somewhat strange.

Just as some numbers do not contain the morpheme "hundred" in their expression, other numbers do not contain the morpheme "and" in their expression. These numbers can be divided into two classes. The first of these classes contains the numbers 0-100, and the second of these classes contains numbers over 100 that are composed of multiples of 1-100 combined with various decimal multipliers, beginning with 100 and continuing with 1,000; 1,000,000; and so on. The first class, numbers from 0-100, needs no illustration beyond random selection, say 4, 19, 42, 75, and so on. The second class, however, needs further explanation and illustration. For a number over 100 not to have "and" in its expression it must contain "zero" in both the tens' place and the units' place in the first string or at the hundreds' level. An example would be 2,900 pronounced either "two thousand, nine hundred" or "twenty-nine hundred." Hurford (2003, pp. 42-43) would prefer the first expression, because of what he calls the "packing strategy," the idea that well-formed numerical expressions should contain the highest decimal multiplier, in this case 1000. He does admit, however, that an expression such as "twenty-nine hundred" also appears to be acceptable, though he cannot explain why. The important thing to note here is that the number ends with 0 in both the tens' place and the units' place at all levels The numeral could be any number from 1-100, and the multiplier could be any decimal beginning at 100. Further examples would be 53,500 ("fifty-three thousand, five hundred") and 49,000,000 ("forty-nine million").

The first class of numerals that does not contain "and" is especially interesting. Conventionally, the numbers 0-9 are considered to be simple numbers whose function it is to combine with other numbers, either through addition or multiplication, to form complex numbers (Hurford, 1975; Greenberg, 1978; and many others since). In other words, a complex number is either a sum or a product of two other numbers, one of which is usually a simple number and one of which is usually a decimal multiplier. Thus the number 37 is composed of multiplying the simple number 3 by the multiplier 10, attaining the product 30, and then adding the simple number 7, attaining the final sum 37. Nevertheless, in the light of this discussion there seems to be some justification for considering the numbers 0-99 as a distinct category of numbers in numerical expressions, mainly because they never contain "and." These numerals are not separate lexical items, but within numerical expressions they behave as if they were lexical because they frequently occur both before multipliers and after "and" at the end of numbers Examples would be "twenty-three million," "sixty-four thousand," and "five hundred and twenty-four.".

Thus we are led to another generalization, stated in (5).

(5) Within numerical expressions the numbers 1-99 often may be treated as separate lexical items. They often appear before higher multipliers beginning with 100, combining with them in phrasal structures. Their function here is multiplicative. The numbers 1-99 also appear after “and” at the ends of complex numbers. Their function in this case is additive.

Combining this generalization with the generalizations stated in (4) yields all the information needed to proceed with an interpretation of English numerical expressions within the framework of X-bar theory. To summarize: 1) English numerals are arranged in strings of three digits with a units’ place, a tens’ place, and a hundreds’ place; 2) the morpheme “and” appears after the hundreds’ place, suggesting an additive phrasal structure with two constituents joined by co-ordination; 3) numbers from 1-99 often appear before decimal multipliers and at the ends of complex numbers and can be treated as single lexical items in a multiplicative phrasal structure. At the core of the English system of numerical expression is the frequent appearance of “and” and its location after the hundreds’ place in strings of three digits at all levels.

2.5 Chinese Numerical Expressions

Chinese numerical expressions are similar to English numerical expressions in many ways. They are formed by a combination of multiplicative and additive arithmetical operations, and they are arranged in strings of digit places at successively higher levels. But Chinese numerical expressions are arranged in strings of four places instead of strings of three places, as in English. This means that 10,000, pronounced *wan*, becomes an important multiplier in Chinese, though it does not exist at all as a multiplier in English. Thus the English notated number 23, 417 would be notated 2, 3417 in Chinese, and instead of being pronounced *er san qian* (“twenty-three thousand”) at the beginning it is pronounced *er wan* (“two ten-thousand”) at the beginning. The remainder of the number is pronounced *san qian* (“three thousand”) *si bai* (“four hundred”) *shi qi* (“seventeen”). Each Chinese numeral contains four places: a units’ (*ge*) place, a tens’ (*shi*) place, a hundreds’ (*bai*) place, and a thousands’ (*qian*) place. These four places are repeated at successively higher levels. This illustrated in (6).

(6)	Level of *ge*	Level of *wan*	Level of *yi*
	1-10,000	10,000 – 10,000,000	100,000,000 – 100,000,000,000
ge place	0-9 of ten thousand	0-9 ten-thousands	0-9 ten-millions
shi place	10-99 of ten thousand	10-99 ten-thousands	10-99 ten-millions
bai place	100-999 of ten thousand	100-999 ten-thousands	100-999 ten-millions
qian place	1000-9,999 of ten thousand	1000-9999 ten-thousands	1000-9999 ten-millions

The most important difference between Chinese and English numerical expressions is that, unlike English, Chinese does not use any conjunction equivalent to "and" at any point in the composition of the expression. There is, however, diachronic evidence in Brainerd and Peng (1968) that in ancient times Chinese did use the morpheme *you* (or *yu*) equivalent to "and" in English as a co-ordinate conjunction in numerical expressions. According Liu and Peyraube (1994), grammaticalization first transformed the verb *you*, meaning "give," to a preposition and then later to a conjunction. Li and Thompson (1981) observe that *you* now occurs mostly in pairs, meaning "both . . and" Zhang (2006, p. 180), however, notes that the first *you* of the construction *you . . . you* is deletable, suggesting that if *you* did occur in Chinese numerals, it would do so in the middle of a string. It is also significant to note here that Yang (2005. p. 45) assumes that *you* is part of the internal logic of the composition of Chinese numerical expressions. This idea has merit, I believe, and I will develop it in Section 5.2.

The other major difference between Chinese numerical expressions and English numerical expressions is that Chinese sometimes incorporates the morpheme *ling*, meaning "zero," whereas the digit 0 is always phonetically null in English, except when it stands alone. In Chinese numerals *ling* always appears in the place of a multiplier when the value of that multiplier is 0, but ling only appears once in any string of four digits. For example, the number 56, 0000, 0025 is pronounced *wu shi liu yi ling er shi wu*. There are six zeros in this numeral, but only one *ling* is pronounced. According to Brainerd and Peng (1968), the *ling* that is pronounced is the one nearest to the end of the number. It would seem, therefore that the principal function of *ling* is to hold the place(s) of the absent multipliers. In this example it would be difficult to process and redundant to pronounce the fact that all the multipliers between *yi* and *er shi wu* have been omitted. In English the number 56,000,025 is pronounced "fifty-six billion and twenty-five." Here too mention of the multipliers between "million" and "twenty-five" has been omitted. And yet there is an intervening morpheme, namely the conjunction "and." It is tempting, in this example, to equate the English use of "and," with the Chinese use of *ling*, because both these morphemes seem to hold the places of intermediate multipliers and to signal that something has been omitted. Nevertheless, we have already seen that the use of "and" in English numeral expressions appears to be predictably related to the function of co-ordination, and it therefore seems to exhibit phrasal qualities within the framework of X-bar theory. This is not immediately apparent in the case of *ling*, which is essentially a place holder. I will have more to say about the similarities and differences between "and" and *ling* when I present my analysis in Section 5.

In the meantime, the phonological rules in (7) may be said to govern the use of *ling*:

(7) a. Although *ling* appears medially in Chinese numerals, it is always a number, not a conjunction equivalent to "and," which appears medially in English numerals.

b. When *ling* appears in the coefficient position before any multiplier, the multiplier is phonologically null at PF. For example, 305 is pronounced *san bai ling wu,* not *san bai ling shi wu*.

c. When we have more than one consecutive ling, only one of them is pronounced because of phonological haplology. For example, the number 56, 0000, 0025 is pronounced *wu shi liu yi ling er shi wu*. There are six zeros in this numeral, but only one *ling* is pronounced.

d. If *ling* appears at the end of a numeral, it is phonologically null at PF. For example, 1500 is pronounced as *yi qian wu bai* or *yi qian wu*, not *yi qian wu bai ling ling*.

e. If a number ends with a multiplier, pronunciation of the multiplier is optional. Thus, 3500 is pronounced either as *san qian wu bai* or *san qian wu*.

CHAPTER 3

HURFORD'S STANDARD ANALYSIS AND MY FIRST RESPONSE

3.1 Hurford's (1975, 2003) Analysis

The phrase structure rules given by Hurford identify three constituents of numerical expressions: Phrase, Number, and M(ultiplier). The value of a Phrase is the product of its constituents, and the value of a Number is the sum of its constituents. The value of an M is always 10 or a multiple of 10. Hurford's structure rules state that both Number and M are recursive. His basic phrase structure rules are given in (8).

(8)

$$\text{NUM} \rightarrow \begin{Bmatrix} \text{DIGIT} \\ \text{NUMPHRASE (NUM)} \end{Bmatrix}$$

$$\text{NUMPHRASE} \rightarrow \quad \text{NUM} \quad \text{M}$$

In Hurford's own words (2001, p. 10758): "Here, `NUM' represents the category Numeral itself, the set of possible numeral expressions in a language; `DIGIT' represents any single numeral word up to the value of the base number (e.g., English *one*, *two*, . . ., *nine*); and `M' represents a category of mainly noun-like numeral forms used as multiplicational bases (e.g., English -*ty*, *thousand*, and *billion*). The curly brackets in the rules enclose alternatives; thus a numeral may be either a DIGIT (e.g. *eight*) or a so-called NUMPHRASE (numeral phrase) followed optionally by another numeral (e.g., *eight hundred* or *eight hundred and eight*). If a numeral has two immediate constituents (i.e., is not just a single word) the value of the whole is calculated by adding the values of the constituents; thus *sixty four* means 60-4. If a numeral phrase (as distinct from a numeral) has two immediate constituents the value of the whole is calculated by multiplying the values of the constituents; thus *two hundred* means 2 – 100."

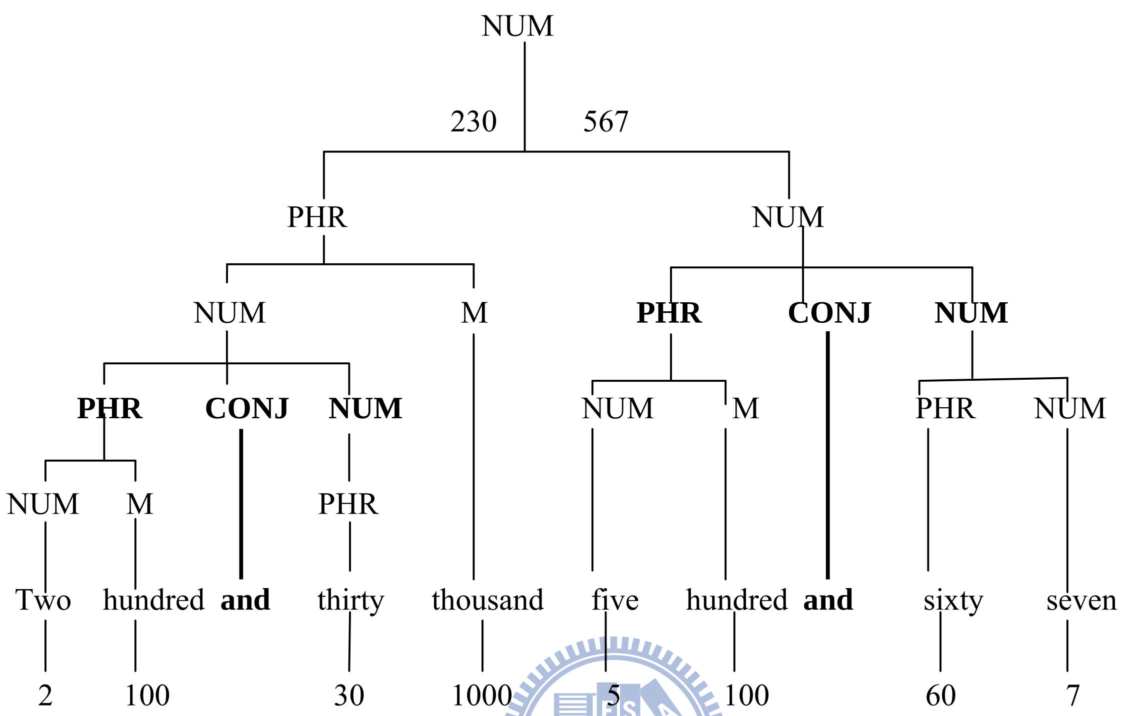
There are three problems associated with Hurford's structure:

1. The first problem is the simple fact that his categories of Phrase, Number, and M are not related to the standard syntactical categories of Noun, Verb, Adjective, and Preposition. Hurford claims (1975, p. 19) that Phrase, Number, and M are syntactical categories, but all that he specifically mentions in regard to syntax is that Multipliers are always nouns.

2. The second problem is that his phrasal structure incorporates ternary branching to include the conjunction "and." This violates a major principle of X-bar theory, which states that phrasal structures must be binary branching (Kayne, 1984; Pollard, 1984; Kornai & Pullum, 1990). This flaw is illustrated in (9) below.

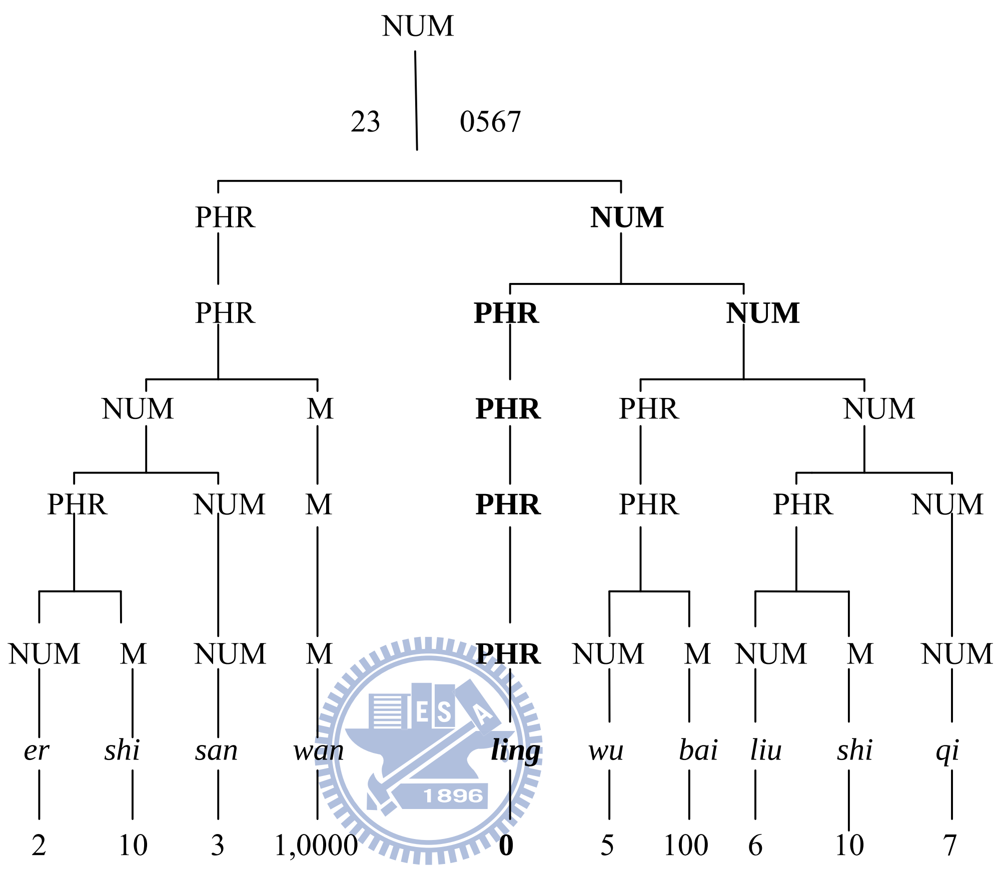
3. The third problem involves the placing of *ling*, meaning "zero," when the structure is applied to Chinese numerals. This morpheme is placed in the Phrase category when logically it should appear in the Number category. This problem is illustrated in (10) below.

(9) English numeral 230, 567



The problem of ternary branching is obvious here. This seems to occur because of the attempt to combine nouns and a conjunction syncategorematically through his conjunction insertion rule (p. 50), but this violates the principles of X-bar theory. It is also interesting to note that Hurford makes no attempt to incorporate in his structure the multiplicative composition of two complex Phrases, 30 and 60. Apparently this operation has been left out because it is necessary to pronounce “thirty” and “sixty” as integral parts of the overall numerical expression. In other words, the operations of multiplying the Numbers 3 and 6 by the Multiplier 10 to form the Phrases 30 and 60 are phonologically null. Hurford has omitted these operations in his structure because he is only interested in representing the actual pronunciation of the numerical expression.

(10) Chinese numeral 23, 0567



.

The only possible justification for Hurford's structure here is to say that *ling* combines with the deleted M *qian* at LF to form a Phrase and that is what is projected. There seems to be some plausibility for this when we note that in Hurford's structure of the English numerical expression in (8) he projects both "thirty" and "sixty" as phrases, presumably because the numbers 3 and 6 have already combined with the M 10 and these operations are phonetically null.

3.2 My First Response to Hurford

The phrasal structure I first considered is compatible with Hurford's generally accepted analysis. At the same time, my proposal is more precise than Hurford's in its terminology, and it therefore more accurately describes the ways that numeral expressions are actually generated.

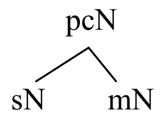
In my proposed analysis all the constituents are nouns, so in the beginning I believed it to be important to distinguish the unique feature of the different kinds of nouns and nominal compounds used in the proposed phrase structure. This was done in the following ways.

1. A noun that serves as multiplier (10, 100, 1000 and so on) is designated [mN].
2. A simple number (0-9) is designated [sN].
3. A complex number – that is, a number made by multiplication or addition – is designated [cN]. However, all complex numbers are categorized, as in (4) and (5).
4. A complex number that is a product of multiplication is designated [pcN].
5. A complex number that is the sum of addition is designated [scN].
6. A complex number that is combined by [scN] and [mN] is [mP].

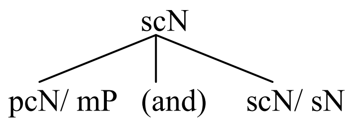
Labeling all the numbers and numerical expressions as nouns indicated their major phrasal category, and sub-labeling them according to their functional characteristics indicated the exact and complete ways that they operate in the phrase structure. Hurford's terminology for nominal phrase structure rules may be more parsimonious; but my terminology was more extensive, precise, and meaningful.

The fundamental phrase structures for the proposed nominal compounds of numerical expressions are given in (11).

(11) a. The Multiplicative Structure



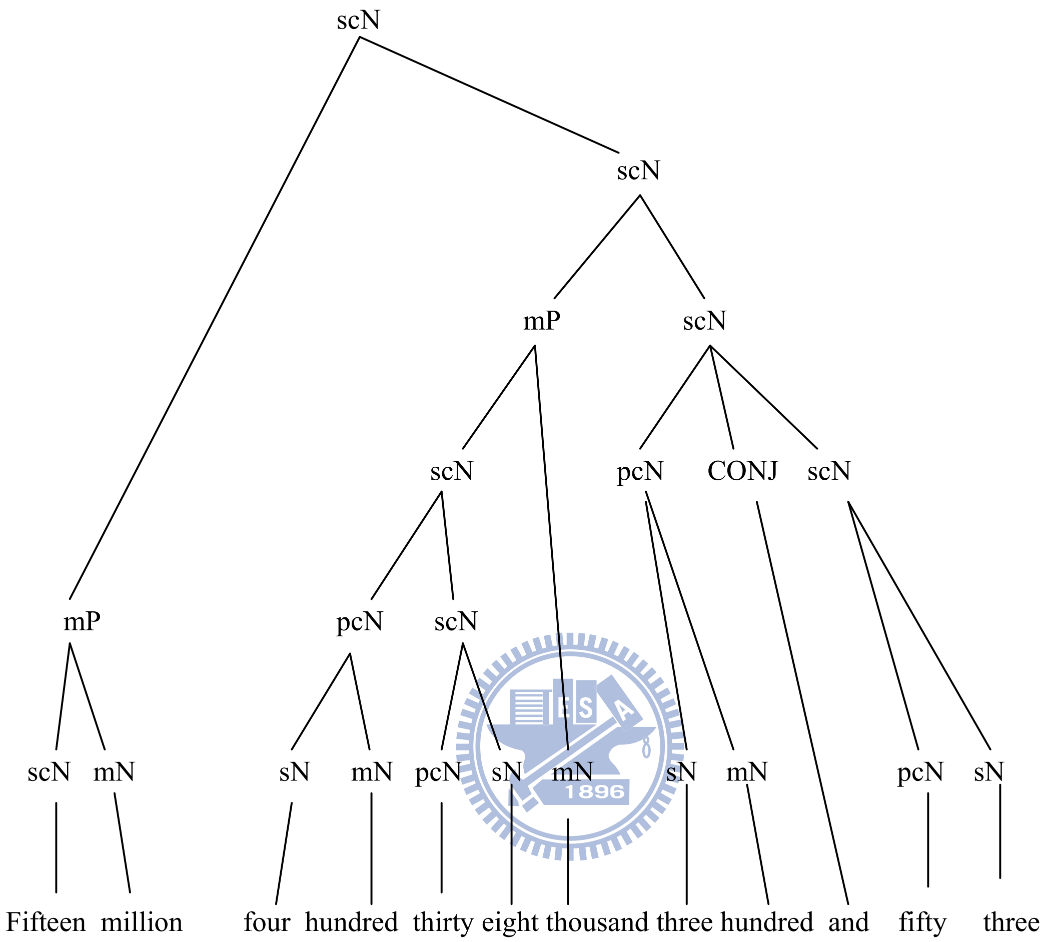
b. The Additive Structure



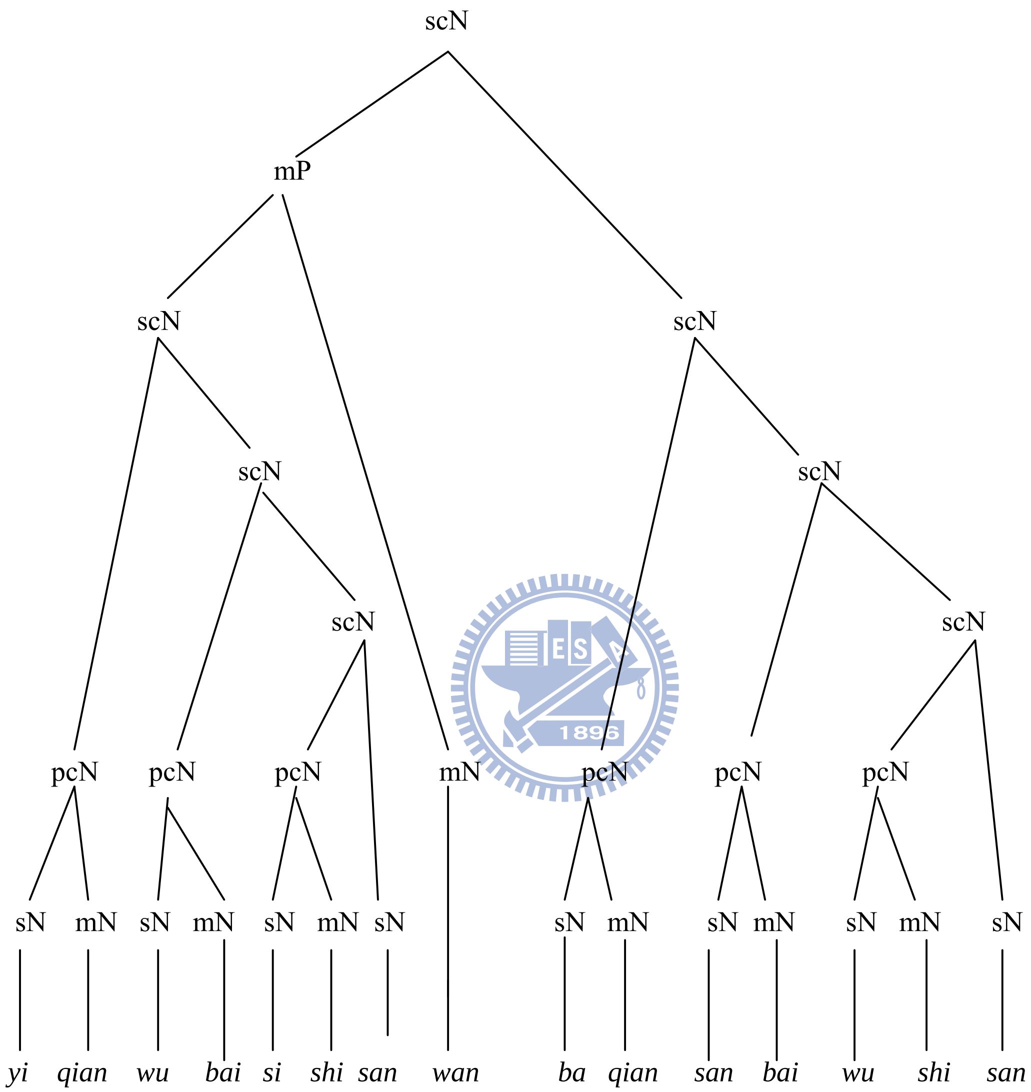
At first I was willing to accept Hurford's flat and ternary branching additive structure with the insertion of the conjunction "and", even though it is difficult to combine this part of the structure with the multiplicative part of the structure. Just the same, I suspected that ternary branching is somehow inappropriate for analyzing the structure of numerical expressions.

The full phrasal structures I first considered for the construction of numerical expressions as nominal compounds in English and Chinese are presented in (12) and (13).

(12) English numeral 15,438,353



(13) Chinese numeral 1543,8353



I think that my first revision of Hurford's phrasal analysis of numerical expressions does have a certain amount of merit. Not only does my revision clarify Hurford's terminology and make it more exact, but my revision also challenges a basic assumption on which Hurford builds his system. From the outset I have assumed that all numbers are abstract nouns, and therefore they should be regarded as absolutely context free. This assumption represents an important departure from Hurford's somewhat tentative assumption that small numerals from 0 to 9 behave as adjectives, while larger numbers behave as nouns. For example, in a note on Corbett's (1978) statement that the higher numbers are, the "nounier" they become, Hurford (1980, p. 247) says, "I believe that he is right." And yet Hurford does not incorporate the syntactical category of adjective into his proposed phrasal structure for numerical expressions. The most that he says in regard to syntactic categories is that his Phrases are nouns. I am convinced, however, that all numbers are abstract nouns, and that my precise – though, admittedly, complicated – system for labeling all numerical nouns according to their arithmetical functions is, at the very least, an improvement on Hurford's nomenclature.

My first response to Hurford's analysis was that he is not willing to make the attempt to fully integrate the grammar of numerals with the grammar of language. Because of this apparent reluctance Hurford is not able to analyze numerical expressions according to X-bar theory, and, since I am personally committed to acceptance of this theory, I found Hurford's analysis to be unsatisfactory. I felt that there must be a way to construct a phrasal structure for numerical expressions that did not violate the rule of binary branching simply because it had to incorporate the transformation of conjunction insertion in English. Besides, Chinese numerical expressions do not even exhibit conjunction insertion at PF. The only solution appeared to be the construction of a binary branching phrase that incorporates conjunctions as part of the phrase. Therefore, I propose that numerical expressions should be analyzed as additive or cumulative Conjunction Phrases (&P). Throughout the remainder of this thesis I will advance my argument for this interpretation of numerical expressions, providing what I believe is a plausible revision of Hurford's phrasal analysis.

CHAPTER 4

NUMERICAL EXPRESSIONS AS CONJUNCTION PHRASE (&P) ADJUNCTS

4.1 On Co-ordination

We have seen so far that, in both English and Chinese, in the most fundamental manner possible, numbers are always sums. Therefore, numerical expressions encode, first and foremost, the arithmetical operation of addition. Intuitively, then, we must expect that the most appropriate syntactical operation for the construction of numerical expressions is co-ordination. In my first response to Hurford's classic analysis I attempted to improve upon the nomenclature of Hurford's phrase structure by identifying the various noun phrases involved in the arithmetical operations of numeral expressions, giving these noun phrases specific functional names. But that was not sufficient to really expand or revise Hurford's analysis. In order to do that we need to interpret numerical expressions according to X-bar theory, but before we proceed in that direction, we need to examine certain matters related to co-ordination in general.

According to Carston and Blakemore (2005), the central issue in the current discourse on co-ordination is the matter of symmetry versus asymmetry. There is an intuitive sense whereby co-ordinated elements possess both similar semantic values and similar syntactical status. Moreover, the meaning of "and" appears to suggest that in terms of truth-conditional propositions, *P & Q* is equivalent to *Q & P*, though pragmatics suggests that in some cases *P & Q* is actually *P & then Q.* For these reasons, early interpretations of co-ordination, such as that of Jackendorff (1977), tended to be represented by flat structures, either non-headed or multi-headed, with the conjunction mediating between or among symmetrical syntactic elements. This symmetrical interpretation of co-ordination is still favoured by some authors. Even early X-bar theorists sometimes claimed that co-ordination is an exception to the conventions of the X-bar schema. Nevertheless, with the development of Chomskeyan Principles and Parameters theory and Minimalism in the 1990s, most authors have come to accept the phrasal structure of ConjP whereby the two conjuncts of a co-ordinated structure are not symmetrical in that XP is connected to a constituent formed by the conjunction and YP.

Kubo (2007) points out that it is impossible to ignore the problematic nature of co-ordination in relation to the ambivalence of its symmetrical and asymmetrical features. Of particular interest to Kubo is the fact that co-ordinate constructions are paratactically construed, suggesting that they possess a fundamental symmetry. But the suggestion of symmetry is not the only significant feature of paratactical arrangement. For the purposes of my overall argument in this thesis it is the centrality of the conjunction, particularly "and," in parataxis that I would like to emphasize. If numerical expressions are co-ordinate structures, as I assume they are, their paratactical construction, in their context-free status as a miniature independent form of discourse, implies that adjunctive "and" is required in their formation. While I contend that the paratactical construction of co-ordinate structures reveals the centrality and necessity of "and" in numerical expressions, I do not, however, argue that numerical expressions – or, indeed, co-ordinate structures in general – are to be interpreted as phrasally symmetrical. On the contrary, I assume that numerical expressions, like all co-ordinate structures, are best accounted for as asymmetrical phrases following the binary branching and strong endocentricity principles of X-bar theory. In fact, Kubo also points out two additional characteristics that display the asymmetry of co-ordinate structures: they exhibit c-command relations between the first and second conjuncts, and they exhibit co-ordination internal consistency, as in Ross (1967). I assume that these two features are also evident in numerical expressions.

As Carston and Blakemore (2005) observe, most linguists today accept the asymmetry of co-ordinate structures. This is true even of authors working outside the paradigm of X-bar theory, Principles and Parameters Theory, and Minimalism. Hudson (2003), for example, a proponent of Word Grammar, assumes that in English the conjunction and the second conjunct in a co-ordinate structure form a constituent that is combined with the first conjunct asymmetrically. Zhang (2006) calls the first conjunct the external conjunct and the combination of the co-ordinator and the second conjunct the internal conjunct. I shall use this convenient terminology throughout the remainder of this thesis. Ross (1967) first proposed the constituency of the conjunction and the second conjunct on phonological grounds, arguing that an intonational pause is possible between the first conjunct and the conjunction, but not between the conjunction and the second conjunct, as in (14):

(14) a. John left, and he didn't even say good-bye.

b. John left. And he didn't even say good-bye.

c. *John left and. He didn't even say good-bye.

Another solid argument for asymmetry in co-ordinate structures comes from pragmatic processing whereby the logic of co-ordinate truth statements is not always (P & Q) / (Q & P), but sometimes (P & then Q), as in Hudson's (2003) example, given in (15):

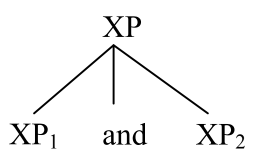
(15) a. She gave him the key, and he unlocked the door.

b. *He unlocked the door, and she gave him the key.

Not only do the conjunction and the second conjunct form a constituent in each of these sentences, but the semantics and pragmatics of the co-ordinated statement indicate that word order cannot be altered without changing the meaning. (P & Q) is not symmetrical with (Q & P). This simple fact has important implications for the asymmetrical and rigidly ordered structure of numeral expressions, as I will illustrate in Section 5.

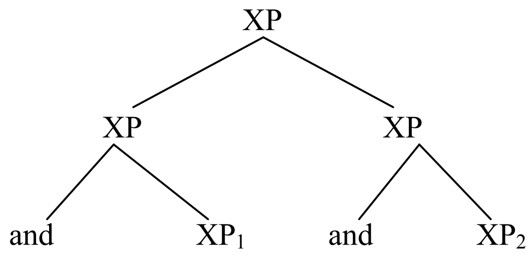
Cormack and Smith (2005, p. 395) sum up the complexity of co-ordination precisely: "Coordination appears to be symmetric, but the grammar is only capable of providing asymmetric structures. In a standard Principles and Parameters version of projection, two phrasal categories can be related in either of two ways. They may be linked (asymmetrically) to a particular head as specifier or complement of that head, or they may be linked (again asymmetrically) as adjunct and host." A convenient overview of the possibilities of phrasal tree structures for co-ordination is given in (16).

(16) a.



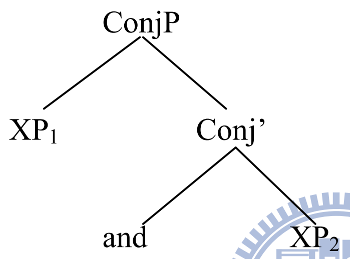
The traditional flat structure, as in Jackendorff (1977).

b.



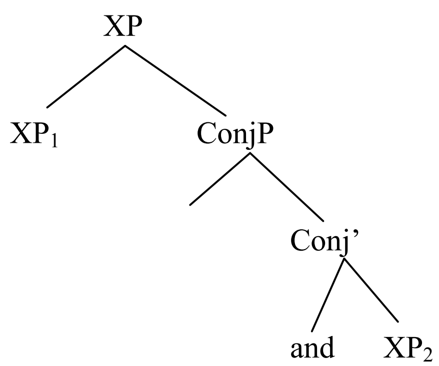
The flat structure with adjoined conjunctions, as in Sag *et al*. (1985).

c.



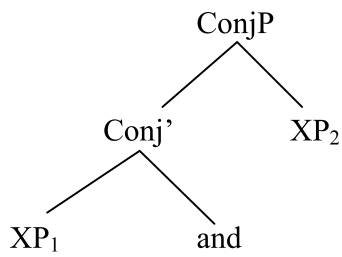
The conjunction phrase with specifier-complement relation, as in Zoerner (1995), Johannensen (1998), and Zhang (2006).

d.



The conjunction phrase with right node adjunction, as in Munn (1993).

e.



The conjunction phrase with left node adjunction, as in Kayne (1994).

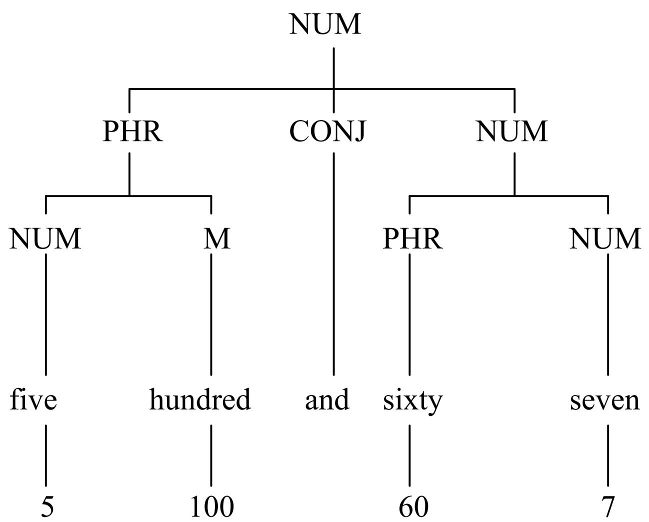
Kubo (2007) points out that all such phrasal interpretations share the assumption that a conjunction like "and" is merged in the narrow syntax. But Kubo also observes that not all natural languages use an overt conjunction for syntactical co-ordination. Drawing on Haspelmath (2005), Kubo identifies languages that use conjunctions as having syndetic co-ordination and languages that do not use conjunctions as having asyndetic co-ordination. Most European language, such as English, use syndetic co-ordination, while many other natural languages, particularly ones that do not have a long traditional of writing, use asyndetic co-ordination. Such languages rely strongly on intonational pauses to indicate co-ordination at PF. It is interesting to note here that, according to Ross (1967), intonational pauses before the conjunction are an important reason for believing that the conjunction and the second conjunct form a constituent. It seems reasonable to assume that in some languages the conjunction has been deleted after the pause because it seems to be redundant. I shall discuss the possibility of isomorphism between syntax and phonology in some detail in Section 5. In fact, Kubo's empirical observations about asyndetic co-ordination are crucially important to my thesis. I assume that numerical expressions are co-ordinate structures that can be analyzed as conjunction phrases. English numerical expressions use the conjunction "and," but Chinese numerical expressions do not use an equivalent conjunction, usually transcribed as *you*. I assume, therefore, that the "&" slot in Chinese numerical phrase structure in covert. According to Kubo, the conjunction phrase analysis cannot account for co-ordination in languages that use only asyndetic co-ordinate structures. In such languages ". . . the whole meaning of co-ordinate structures cannot be determined by a non-existent co-ordinate conjunction" (p. 8). I will dispute this assertion in Section 5.

4.2 Numerical Expressions as Adjunction

To reiterate Hurford (1975, p. 11), "The value of a number is the sum of the values of its immediate constituents." This simple but important fact needs to be kept in mind whenever we are considering the syntactical composition of numerical expressions. Essentially, numbers are sums. They express addition much more than multiplication. Recall that multiplication might be seen as a kind of addition. To say "10 *times* 3" is actually to say "10 *plus* 10 *plus* 10." Sums require at least two components, called summands: one number *and* another number. Thus, it appears that the syntactical operation of co-ordination should be especially appropriate for the composition of numerical expressions. This in turn suggests that conjunctions such as "and" should be an integral part of numerical expressions. Though desirable, this is not an easy thing to accomplish for those who wish to work within the parameters of X-bar theory.

The problem is illustrated very well by Hurford himself (2003). The phrasal structure in (17) is taken from his argument (p. 42).

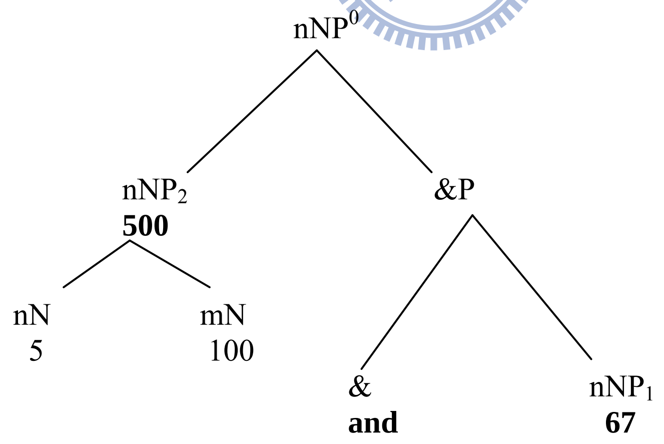
(17)



It is clear from this structure that the number 567 represents a sum of its two principal elements, 500 and 57. This diagram also shows the multiplicative process at work in attaining 500, what Hurford calls a Phrase, and the additive process at work in attaining 67, what Hurford calls a Number. The Multiplier 100 is also included in the structure, though the Multiplier 10 involved in calculating 67 is not included. Besides the fact that this structure represents the sum of 500 and 67 it also represents the fact that these two numbers are added. This is done by the inclusion of the conjunction "and." Here is where the problem occurs for proponents of X-bar theory. In Hurford's structure, the branch leading to "and" makes the entire structure ternary branching, not binary branching, as X-bar theory requires (Pollard 1984; Kayne, 1984; Kornai and Pullum, 1990). Duarte (1991, p. 33) states this explicitly: "A further requirement on syntactic configurations assumed in this framework [of X-bar theory] is binary branching: a mother node cannot have more than two daughters."

The solution I propose is based on Munn's (1993) argument for a phrasal projection of a conjunction such as "and" through right node adjunction. According to this approach, Hurford's number example 567 would be configured as in (18).

(18)



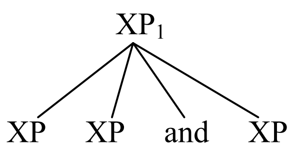
The first thing to note is that this structure incorporates a revision of the labeling of constituents from that presented in my thesis. In this new notation "nNP" signifies "numeric Noun Phrase," "mN" signifies "multiplier Noun," and "nN" signifies "numeric Noun." Finally, "&" signifies "and," and, following Hartmann (2000), "&P" signifies "and Phrase." From the example above, based on the generalization given in (5), it can be seen that the category "nN" includes numbers from 0-99. It also should be noted that this system of labeling makes no distinction between simple numeric Nouns (0-9) and complex numeric Nouns (those higher than 10). Nor does this system mark numerals as products or sums, since it is superficially evident which is which. This system does, however, retain my original assumption that all numbers are nouns.

Another important feature of my proposed adjunction analysis is that it combines the three major characteristics of English numerical expressions: the additive function, the multiplicative function, and the co-ordinate function of "and" – all within the framework of X-bar theory. Thus this phrasal structure not only solves the problem of trinite branching apparent in Hurford, but it also simplifies – and therefore improves – the system of labeling I used in my original revision of Hurford, while at the same time it retains my original improvement on Hurford's labeling by indicating syntactical categories where he had not done so. Before proceeding to apply this new analysis to numerical expressions in both English and Chinese it is necessary to outline and discuss Munn's (1993) treatment of co-ordination as adjunction.

4.3 Munn's (1993) Boolean Phrase (BP) Adjunction Analysis

Munn's (1993) analysis is based on his belief that co-ordination should be incorporated into X-bar theory. Jackendorf f (1977) presents co-ordination as a flat structure with either multiple heads or no heads, as in (19).

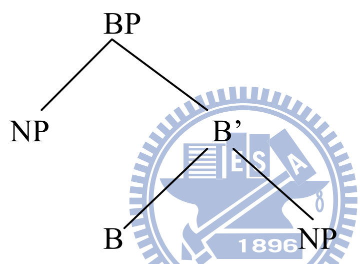
(19)



According this analysis "and" is syncategorematically linked to a series of XPs so that all the elements are equal. Munn observes that such a flat structure violates both binary branching (as we have already seen with Hurford) and endocentricity, two of the principal features of X-bar theory. Nevertheless, the flat structure analysis has had a long history and is still being advocated at the present time. Its proponents include Chomsky (1965), Dik (1968), Dougherty (1969), Gazdar *et al* (1985), Goodall (1987), Johnson (2002) and Phillips (2003). In the meantime, Munn's analysis has independently duplicated by Collins (1988) and subsequently supported by Bošković and Franks (2000) and Alharbi (2002).

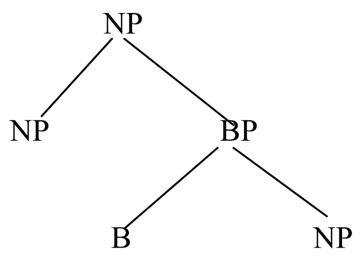
At the core of Munn's analysis is the conviction that the two conjuncts of a co-ordinate phrase are not equal semantically, nor is the conjunction empty of meaning. Following Ross (1967), Munn argues that the conjunction and the second conjunct form a phrasal constituent. Given this interpretation, Munn observes that there are two possibilities for configuration of what he calls the Boolean Phrase (BP). These are illustrated in (20).

(20) a.



Spec/Head BP

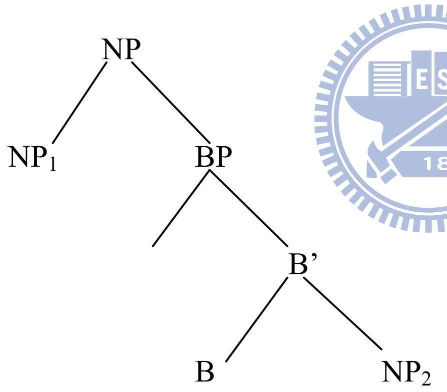
b.



Adjoined BP

The question is, where should the first conjunct be placed? Munn at first decided to place it in the Specifier position, while placing the second constituent in the complement position. This is illustrated in (20a). Munn's later choice was to adjoin the first conjunct to the constituent formed by the conjunction and the second conjunct. This is illustrated in (20b). Munn believes that adjunction supplies the most accurate interpretation of co-ordination, the principal reason being the asymmetry that exists between the two conjuncts. In Munn's adjunction analysis the head B and its complement, the second conjunct, form the maximal projection of the BP. This means that B, or "and," is the head of its own phrase. Moreover, the first conjunct NP_1 and the second conjunct NP_2 are of the same category and at the same bar level. The B and the second conjunct project to an X'' level, and the Specifier place is left empty as a landing site for the null operator. This is illustrated in (21)

(21)



Munn's argument for favouring the right node adjunction analysis of co-ordinate structures over the Spec/Head analysis is focused mainly on binding criteria. But first he points out that in co-ordinate structures the second conjunct – that is, the internal conjunct consisting of the co-ordinator and the second conjunct – can be extraposed, while the first, or external conjunct, may not be extraposed. This is illustrated in (22).

(22) a. John bought a book and a newspaper.

b. John bought a book yesterday and a newspaper.

c. *John bought a newspaper yesterday a book and.

d. *John bought a book and yesterday, a newspaper.

These examples show that the internal conjunct must be a maximal projection, since movement, such as exposition, applies only to maximal projections. This in turn means that the adjoined BP structure is preferable to the Spec/Head BP structure, since the internal conjunct of the adjoined BP is a maximal projection, while the internal conjunct of the Spec/Head BP is not a maximal projection.

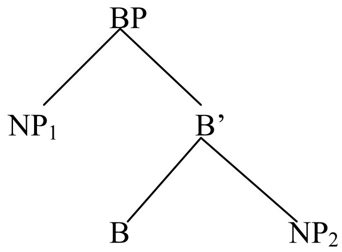
Munn goes on to argue that binding asymmetry is necessary for co-ordination. This is why flat structures are ruled out. The first conjunct must c-command the second conjunct, while the second conjunct must not c-command the first conjunct (Reinhart, 1976). This is illustrated with reference to numerical expressions in (23).

(23) a. seven hundred and twenty-three.

b. *and twenty-three seven hundred.

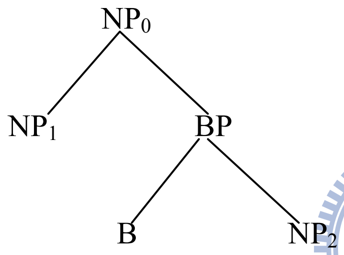
Munn points out that if NP_o in the adjoined structure, or BP in the Spec/Head structure, were simultaneously a projection of both NPs, $NP_{o,}$ or &P, would be the c-command domain for both conjuncts, but this would violate the principle that conjuncts may not be coreferent. Such a c-command argument is enough to dismiss the flat structure analysis of co-ordination, but to dismiss the Spec/Head analysis Munn turns to the concept of m-command (Chomsky, 1986) as an alternative to strict c-command. According to the expanded m-command definition, X m-commands Y if both are maximal projections. Consider the two structures again, given in (19) and repeated here as (24).

(24) a.



Spec/Head BP

b.



Adjoined BP

In the Spec/Head construction (24a) NP_1 does not c-command the BP (or &P, the internal conjunct) because it is dominated by BP. Also, NP_1 does not m-command B', because B' is not a maximal projection. Moreover, in the Spec/Head construction BP is the c-command domain for both NP_1 and NP_2, violating, as we have already seen, the rule against co-reference. But in the adjoined construction (24b) NP_1 does c-command the BP (or the &P, the internal conjunct) because they are sisters. Moreover, NP_1 m-commands the BP because they are both maximal projections. Furthermore, the stricture against co-reference is not violated because NP_0 is the c-command domain for NP_1, but BP is the c-command domain for NP_2.

For the purposes of applying X-bar theory to numerical expressions the adjunction analysis appears to be better than the Spec/Head analysis for two important reasons: the first conjunct both c-commands and m-commands the second conjunct, and the conjunction ultimately projects NP_0, the desired outcome, instead of BP (&P). These features are absent in the Spec/Head analysis. This means that the sample number 567, illustrated throughout this paper, would be an NP with the semantic additive sense of "five hundred *plus* sixty-seven." In other words, in keeping with the central proposal of this thesis, a numeral expression is both a sum and a nominal compound.

4.4 Alternatives to Munn

4.4.1 Zhang (2006)

Perhaps the most extensive defence of the Spec/Head construction of the &P appears in Zhang (2006). This author argues that co-ordinate structures, what she calls "coordinates complexes" (p. 176), are asymmetrical, binary branching phrases headed by coordinators, and she provides ample evidence that older flat structures are inadequate for analyzing co-ordination. In her version of &P, unlike Munn, she places the conjunction in the specifier position and argues at length that this is the only acceptable arrangement for &P.

There is, however, a problem with the logic of Zhang's argument. In the first place, although she discusses Munn's (1993) BP analysis, she does not address his arguments about binding, although these are the basis of his entire analysis. Moreover her own argument is based entirely on a dubious assumption. Throughout her paper Zhang insists that adjunction is only possible for subordination. Given this assumption, all she has to do is show that Munn's adjunction analysis applies only to co-ordinate structures, and then draw the conclusion that his argument must be wrong.

For example, Zhang uses the fact that external conjuncts cannot be stranded to disprove Munn's adjunction argument. In co-ordination internal conjuncts cannot move, but in subordination internal conjuncts can move. Munn's BP is an internal conjunct that cannot move; therefore, it cannot be part of a subordinate structure, and, because only subordinate structures can be adjoined, Munn's BP should not be adjoined. Zhang's other arguments against Munn are even more complicated. For example, she cites Ross's (1967) Co-ordinate Structure Constraint, along with Grosu's (1987) revision of this rule, whereby neither a conjunct nor an element of a conjunct may be moved. Zhang then argues that extraction may, in some instances, occur in co-ordinate structures. Extraction may not, however, occur in subordinate structures, mainly because they are islands. It must be remembered here that, according to Zhang's basic assumption, adjunction can only occur in subordinate structures. Therefore, since extraction can occur in co-ordinate structures (as exceptions to Ross's and Grosu's constraints), co-ordinate structures, such as Munn's BP, may not use adjunction.

Such tangled logic suggests that Zhang's argument is dubious. Not only does she fail to confront Munn on his own ground of binding theory, but she also depends too much on either/or thinking and probably creates a false dilemma when she insists that adjunction can only occur in subordinate structures, not in co-ordinate structures. In fact, Zhang appears to be begging the question, or assuming the validity of what needs to be proved, when she assumes as a fact the notion that adjunction is limited to subordinate structures. It is interesting to note here that Hurford (1975, pp. 174-175) does not recognize such a limitation in his own discussion of Chomsky-adjunction in relation to numerals. All he states is that the difference between Chomsky-adjunction and ordinary adjunction is that Chomsky-adjunction creates extra structure by copying downward of the node to which something is being adjoined. Hurford uses left-node adjunction in his example, but there is no reason to believe that he excludes right node adjunction as used by Munn for his BP structure.

4.4.2 Kayne (1994) and Zoerner (1995)

Both Kayne (1994) and Zoerner (1995) base their analysis of co-ordination on left node adjunction for the simple reason that Kayne's Linear Correspondence Axiom (LCA) bars right node adjunction. The LCA states that asymmetric c-command places liner order on all terminal elements. This means that all phrasal ordering is mapped into precedence relations. This, in turn, means that linear order is not parameterized Like Munn, Kayne combines the conjunction and the second conjunct to form a constituent, but, like Zhang, he places the first conjunct in the specifier position and makes the internal conjunct a simple bar level constituent. Zoerner, like Munn, regards the conjunction as the head of a phrase, which he calls &P.

Thus we see that both Kayne and Zoerner argue a Spec/Head or complementation structure for co-ordination that is similar to Zhang's interpretation, except that Kayne and Zoerner construct left-branching structures, whereas Zhang constructs (like Munn) a right-branching structure. (See 4.1 for trees) Interestingly, Zhang provides evidence to validate right-branching structures like Munn's and her own, while at the same time dismissing left-branching structures like Kayne's and Zoerner's. Zhang (p. 186) identifies English and Chinese co-ordinate structures on the one hand and Japanese co-ordinate structures on the other hand in (25).

(25) a. [[X] [&Y]] English and Chinese co-ordination

b. [[X&] [Y]] Japanese co-ordination

Because both English and Chinese co-ordinate structures are left branching, both Kayne's and Zoerner's right branching analyses can be dismissed for the purposes of my argument favouring Munn's right node adjunction analysis of co-ordination.

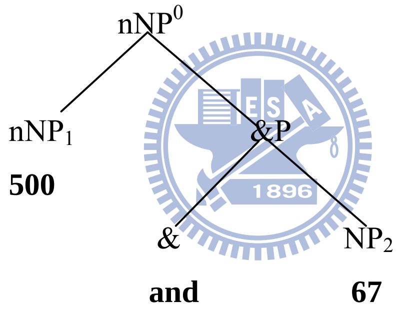
CHAPTER 5

MY PROPOSED ANALYSIS

5.1 Of English Numerical Expressions

My original example, given in (17) above, consists of a combination of two phrasal structures joined by Tree Adjunction Grammar Gazdar *et al*., 1985). The initial structure is illustrated again in (26).

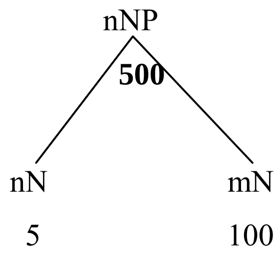
(26)



The phrasal structure shows the additive construction of the ultimate NP or nominal compound. NP_1 is pronounced "five hundred." This is immediately followed by &, pronounced "and," which is in turn followed by NP_2 , pronounced "sixty-seven." Put together, this numerical expression is pronounced "five hundred and sixty-seven." As it stands, this structure represents co-ordination as right adjunction, and it illustrates the fact that complex numerals are fundamentally sums or nominal compounds.

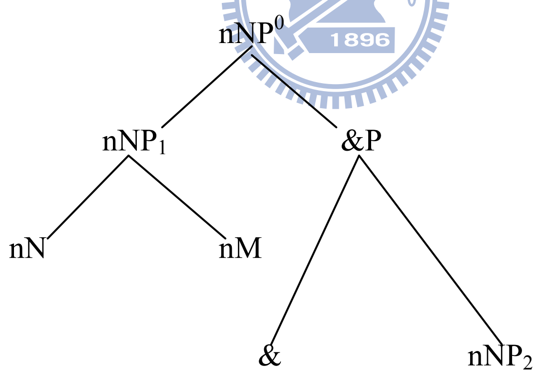
To account for the fact that most complex numbers are also composed in part by multiplication requires a second structure. Because of my proposal, stated in generalization (5) above, I consider the numbers 1-99 to behave in numeral expressions as if they were separate lexical entities. This is especially true of numbers following "and." Thus in (18) and (26) the number 67 is represented in the structure as a single entity, mainly because it is part of the additive process. The combination of the numbers 1-99 with decimal multipliers is a different matter. Because these numbers usually indicate the overall scope of the numeral, and because they represent the multiplicative process, I believe that they deserve their own structural representation, given in (27).

(27)



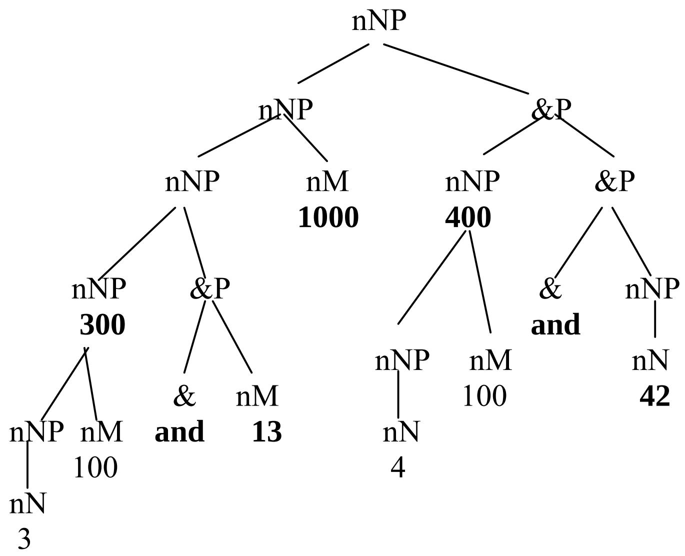
When this auxiliary structure is combined with the initial structure, the result is the complete phrasal structure for my analysis of numeral expressions, given in (28).

(28)



Moreover, as Munn observes, the adjunction structure for conjunction phrases can be iterated as much as desired. This, of course, is essential for large complex numbers. Each new conjunct must be preceded by an &-head that serves as a complement for it. An example of the phrasal structure for such a number is given in (29).

(29) English numeral 313,442

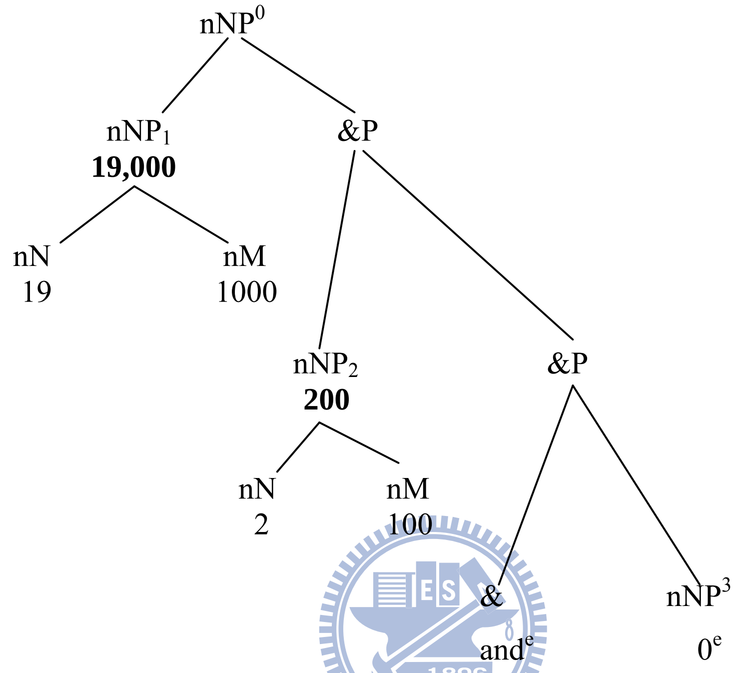


It should be noted that & ("and") appears twice in this structure, once in string of three digits at the thousands' level and once in the string of digits at the hundreds' level. It should also be noted that each occurrence of "and" follows the hundreds' place. There is a sense that "three hundred and thirteen thousand" means "three hundred thousand and thirteen thousand." That is, the numeral expression deletes the first "thousand" present at LF. The deleted number is, however, restored in the structure.

Since the tree diagram is drawn based on Munn's adjunction analysis, the numbers are projected to &P. In English numeral expression, it can be easily seen that there are three multipliers appearing in this number and those are 'digit', 'hundred' and 'thousand'. Thus, the number before 'thousand' is seen as a modifier for 'thousand' and the number before 'hundred' is seen as a modifier for 'hundred' and so on. Also, English numeral expressions contain both overt and covert conjunction "and." In the diagram there is an overt "and' appearing between '300' and '13' in '313000', and there is another overt "and" appearing between 400 and 42, but either or both of these conjunctions could also be covert. The pronunciation of the conjunction is optional in English numerical expressions, but it is usually overt.

Since the &P analysis is headed by &, a difficulty arises if there is no “and” in the numeral expression. An example of such a number is 19, 200, configured in (30).

(30)



In this numeral expression both the place for & and its complement NP are phonetically null for the simple reason that they are not needed semantically, though they are certainly present at LF. Such a numeral expression represents a small exception to the general pattern of &P adjunction analysis, so I believe it does not invalidate the use of this phrasal structure for English numerical expressions.

5.2 Of Chinese Numerical Expressions

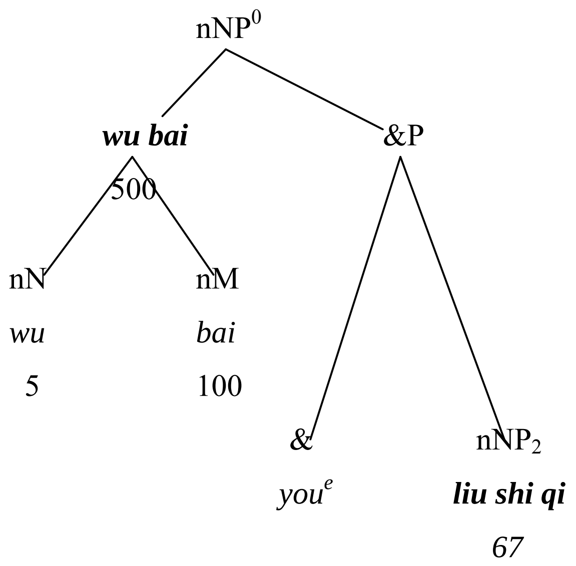
What about Chinese numerical expressions?   Since there is no equivalent of English “and” in modern Chinese numerical expressions, it might seem paradoxical to propose that Chinese numerals can be analyzed as & Phrases. Nevertheless, I think this approach can be justified.

The &P analysis appears to work well for English numerical expressions, and Chinese numerical expressions are very similar to their English counterparts. We first need to concentrate on the absence of an equivalent of "and" in Chinese numerical expressions.

How can a numerical expression be represented as an &P if there is no conjunctive morpheme present in that expression? To answer this we must recall what Hurford has taught us: Numbers are sums. In other words, numbers consist of at least two smaller numbers being combined through addition. Logically, in terms of syntax, this suggests co-ordination. There is no doubt, then, that numerals are what Zhang (2006) calls "co-ordinate complexes." I call them nominal compounds. What all this means is that the logic of the syntactical composition of numerical expressions is the logic of co-ordination. This in turn means that we might predict that a conjunction similar to English "and" *should* appear in Chinese numerical expressions.

Brainerd and Peng (1968) remind us that in Archaic Chinese the morpheme *you* or *yu* was used as a conjunction in numerical expressions. Liu and Peyraube (1994) argue that *you* was originally a verb that grammaticalized, first, to a preposition, then later, to a conjunction. It is possible, therefore to conclude that *you* is still present at LF in Chinese numerical expressions. Yang (2005) makes this assumption, and I agree with her. Winter (1995, p. 7) identifies the condition of *you* in regard to numerals perfectly. "There exist languages with zero conjunction because morphemes like *and* are not necessary for conveying logical connection." The fact that *you* is not present at PF in Chinese numerals does not mean that it was not once present at PF, and that it is not still present at LF. In short, the logic of the composition of numerical expressions in Chinese demands co-ordination, and co-ordination is normally expressed with conjunctions such as "and" or *you*. Therefore, I propose that, since phrasal structures are definitively rooted in Deep Structure – that is, language before transformations occur – *you,* though it is phonetically null and marked *you*^{e}, may be considered as a legitimate holder of the place & in an &P. This is illustrated in (31).

(31) Chinese 567

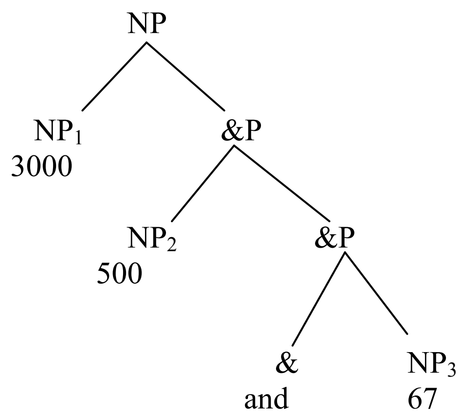


It has been argued throughout this thesis that numbers are primarily notations for a system of arithmetical operations. Moreover, the central operation of this system is addition. Numbers are sums. We have also seen in Section 2.4 that the division of each string of three digits in English into two distinct parts, a hundreds' place followed by a tens' and units' place, mediated by the conjunction "and," suggests the syntactical binary-branching phrasal structure of X-bar theory. Indeed, my main argument has been that numerical expressions, at least in English, can be adequately accounted for by the requirements and conventions of X-bar theory. Moreover, because the arithmetical structure of numerical operations is obviously addition, it follows that the syntactical structure of numerical expressions should be co-ordination. As Cormack and Smith (2005) have observed, co-ordination might well be the most primitive of all syntactical operations. If this is true, co-ordination seems to be well suited to accounting for the similarly primitive system of notation to be found in numeral expressions. At the core of this matter is the indisputable fact that both the arithmetical operation of addition and the syntactical operation of co-ordination encode the relation of parts to wholes. Because of the centrality of this relationship to both operations, cumulative conjunctions seem to be indispensible for both addition and co-ordination. Indeed, Hudson (2003) argues that in English "and" does not express dependency but the relationship of parts to wholes in what he calls word strings. All of these facts suggest that if any X-bar phrasal structure can account for numerical expressions, it must be co-ordination. Moreover, co-ordinate phrasal structures require conjunctions. Therefore, in an English cumulative co-ordinate phrasal structure "and" must appear, either overtly or covertly. Similarly, in a Chinese cumulative co-ordinate phrasal structure, *you*, or at least some conjunction, *should* appear, either overtly or covertly. Why then is *you* phonologically null in the adjunctive co-ordinate structure I have proposed for Chinese numerical expressions?

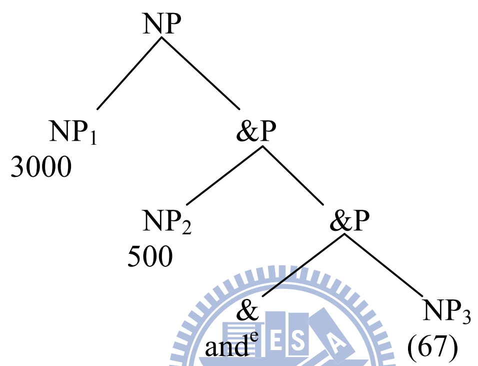
The answer is not likely to be found in standard government and binding rules. The empty category principle, for example, does not apply to e-& in numerical expressions for the simple reason that there is no movement involved in this instance, so there is no trace to be identified. Furthermore, & does not assign case in numerical expressions. In fact, case is not marked on the elements of either English or Chinese numerals. Hurford (2003, p. 65) observes that in some natural languages, such as Finnish, numerals are marked for case agreement, but even in those instances it is not a conjunction that assigns case but the context of the sentence in which the number appears. Kayne (1994) claims that abstract X^0 – in effect, e-& – licenses the first conjunct in a co-ordinated phrase, but he makes very little of this in relation to the second conjunct. Alharbi (2002, p. 76) points out that in the phrasal structure of recursive co-ordination e-& occurs as deletion or ellipsis at LF, and it is there, in recursive e-&, that a clue might be found to explain why *you* is phonologically null in Chinese numerical expressions.

A close examination of the empirical data reveals that & is frequently deleted in recursive co-ordinate structures such as complex numerical expressions. The right node adjunction analysis I have been using throughout my argument easily accommodates recursivity in various arrangements, as in (32).

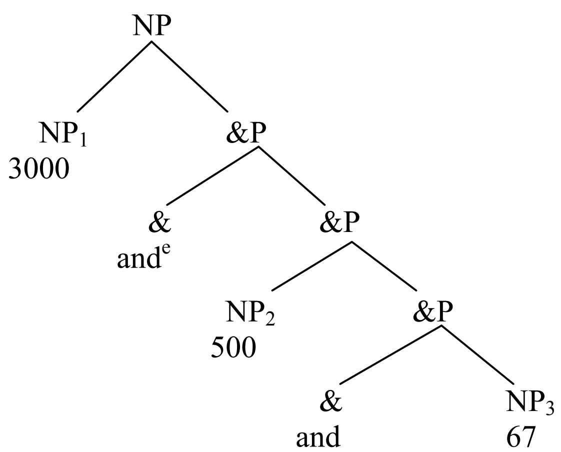
(32) a. 3, 567



b. 3, 567



c. 3, 567



In (32a) we see the standard structure for the English numeral 3, 567. In this version "and" occurs in the middle of the final string of three digits, immediately after the hundreds' place, as is usual in such expressions. In (32b), however, "and" is deleted in this place. Leaving e-& as in Chinese numerical expressions. Dik (1968, p. 272) admits that such an expression as "three thousand, five hundred sixty-seven" is acceptable in English, especially in American English. In fact, this phrase is the English equivalent of the Chinese numerical expression "*san qian wu bai liu shi qi*."

We should note here what seems to be a trivial empirical fact, although it is, I think, important to this discussion. In the orthographic or written form of the English numeral in (32b) a comma is used to mark the intonational pause introducing the final string of the digits. In the written form of the English translation of this numeral no comma is used, respecting the fact that Chinese numerals are arranged in strings of four digits – not three, as in English. This suggests that there is no need to mark this Chinese numerical expressions with punctuation because there is no string break in it. I realize, of course, that in Chinese orthography or written form commas are not ever used to mark intonational pauses, but, I submit, such pauses are still present in Chinese at PF. Moreover, if the numeral in (32) is expanded to five digits, then it would be appropriate to write it in an English translation of the Chinese numeral with a comma included, for example: *ba wan, san qian wu bai liu shi qi* (8, 3567). English punctuation simply supplies a means of identifying such phonological facts in the translation of Chinese numeral expressions.

In (32c) "and" is present as e-& after NP_1 (3000). This coincides with the break before the introduction of the final string of three digits. This break is indicated by a comma in both the arithmetical notation and the written form of the numerical expression. The comma also indicates an intonational pause at this point. All of this suggest that & appears, either overtly or covertly, in co-ordination wherever string intonational pauses are required. In English numerical expressions & appears overtly after the hundreds' place in each string of three digits, and it also appears covertly between the strings of three digits. In Chinese numerical expressions & only appears covertly – possibly between the strings of four digits, but mainly, I propose, between the hundreds' place and the tens' and units' places in the final string of four digits. If this is true, then Chinese numerical expressions resemble English numerical expressions in that they can both be accounted for by analyzing them as adjoined conjunction phrases (&P).

But why should & appear covertly at this place in the final string of four digits in Chinese numerical expressions? This claim is not merely an assumption of convenience for my argument. I believe there is a principled reason for proposing this arrangement. Dik (1968, pp. 41, 58) states that if there is to be one conjunction in a recursive co-ordinate structure, it must come before the last conjunct. As we have already seen, & is deleted between the strings of three digits in English numerical expressions, and & is deleted not only between the strings of four digits in Chinese numerical expressions but also at the end of the numeral before the last conjunct. Dik's placement rule is more obviously true in non-numerical co-ordinate constructions. Consider the common English phrase "Tom. Dick, and Harry." Our knowledge of co-ordination allows us to predict that the meaning of this phrase is "Tom *and* Dick *and* Harry." The important thing to note here is that the commas in the written form of the phrase indicate two things. First, the commas mark intonational pauses at PF, and secondly, the commas alerts us to the fact that there is an e-& between the first two NPs. Taken together these two facts suggest that in recursive co-ordinate structures – especially in numerical expressions – either commas or intonational pauses, or both, signify e-&.

Are there intonational pauses in Chinese numerical expressions? As we have already seen, such pauses are clearly apparent in English numerical expressions. These pauses are marked by the commas between strings of three digits and by the appearance of "and" after the hundreds' place in each string. Commas are not used in transcriptions of Chinese numerals. Nevertheless, although it is not noticeable in everyday speech, in *formal* spoken Chinese there *can be* intonational pauses after each decimal multiplier. These pauses might be marked with commas in transcription, and the commas, in turn, might indicate intonational pauses, as in (33).

(33) a. *ba wan, san qian, wu bai, liu shi qi*

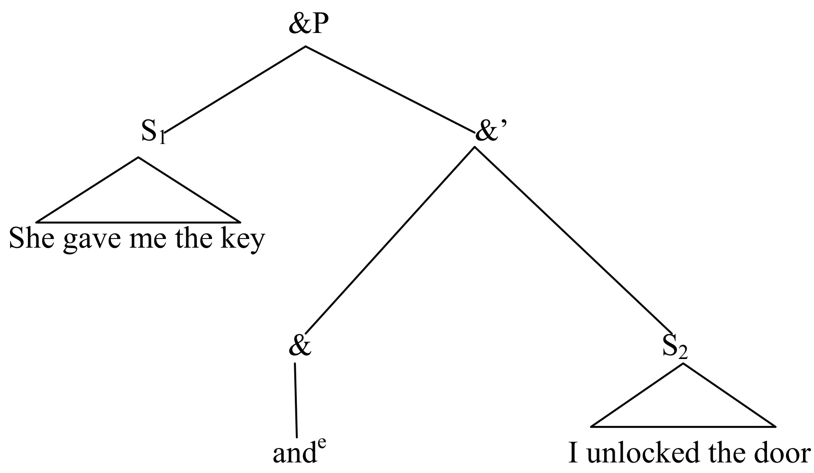
b. *ba wan* $_{\text{[pause]}}$ *san qian* $_{\text{[pause]}}$ *wu bai* $_{\text{[pause]}}$ *li shi qi*

The pause is longer between strings of four digits, but pauses also *can* occur, in formal speech, after each multiplier throughout the number, and – I believe – the pause before the tens' and units' place in the last string of four digits is usually a little longer than the other pauses in the expression – except, of course, the pauses between strings of four digits. It is important to note here that there is no pause between the tens' place and the units' place in the final part of the string. If my assumption is correct here, the intonational pause at PF in formally spoken Chinese before the final conjunct both mirrors English usage and validates the &P analysis of Chinese numerical expressions.

Hurford (1975, 2003) points out that numerical expressions consist of very rigid word orders. As we have seen throughout this thesis, numerical expressions are also paratactical constructions based on the mutually compatible logical assumptions of addition and co-ordination. For these reasons it is possible to regard a numerical expression as an independent form of discourse, with each successive part of the overall number being an element of the discourse. Specifically, numerical expressions are concatenations of numbers. As Hurford (1975, p. 30) observes, "The operations of addition, multiplication and exponentiation are all defined in terms of simpler operations, and ultimately all in terms of the basic arithmetical operation, counting or incrementing iteratively by 1." Moreover, numerical forms of discourse are unique in that they consist totally of nouns and conjunctions, which may be either overt or covert

Tokizaki (2005), a proponent of an hierarchical, binary-branching e-& analysis of paratactic discourses, argues for the existence of a strong isomorphism between phonology and syntax. Earlier studies had already predicted this. In an empirical study Grossjean, Grossjean, & Lane (1979) discovered that the pauses used by speakers performing sentences were not task specific, but were, in fact, related to the grammar of the sentences. But first Tokizaki analyzes co-ordination between sentences as an asymmetrical binary phrase. Tokozaki's tree structure is the same as the &P we have seen throughout this thesis, except that it uses the Spec/Head complementation arrangement instead of adjunction. This is illustrated in (34).

(34) Sentences 1 and 2: She gave me the key. I unlocked the door.



It is important to note here that Tokizaki has no difficulty positing an e-& in this binary branching phrasal analysis of sentences or discourses – precisely what I have been arguing throughout this thesis. Tokizaki has this to say: "A covert conjunction combines a sentence with a paragraph, which in turn is combined with another sentence or discourse constituent to make a larger discourse constituent. This merging process continues to apply until all the sentences in the whole discourse are combined to make a tree." It seems to me that this description of the importance of covert conjunctions to the construction of discourses in general applies very well to the analysis of numerical expressions. Because numerical expressions have such a rigid word order – and perhaps because Chinese depends upon word order even more than it does upon transformations – we can easily process the semantic information encoded in numerals by paying attention solely to the word order and the intonational pauses. We simply do not need conjunctions to understand numerical expressions. Nevertheless, as Tokizaki points out, the word order is held together – the parts are made whole – throughout the structure by covert cumulative conjunctions. This is reason enough for me to accept the existence of e-& as a key feature of an &P analysis of Chinese numerical expressions.

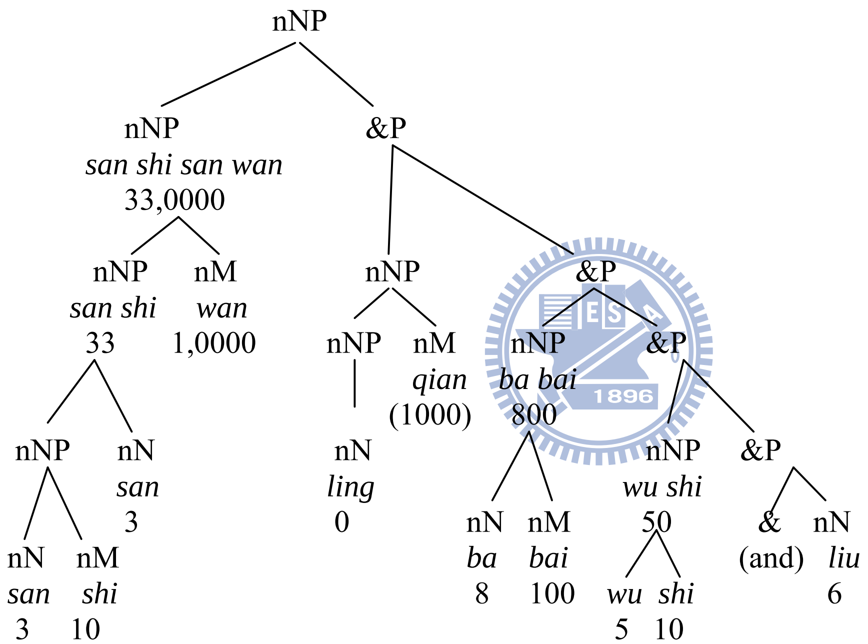
5.2.1 The Matter of *ling*

How well fitted adjunction phrasal structure is to Chinese numeral expressions is even more obvious when the number is very large, as in (35).

(35) Chinese numeral 33, 0856

san shi san wan ling ba bai wu shi liu

(三 十三萬 零八百五十六)



This Chinese numeral expression is also drawn based on Munn's adjunction analysis, so all the numbers are adjoined from the right and start with &P. However, here, it can be easily seen that there are five multipliers and they are 'digit', *shi*, *bai*, *qian,* and *wan*. Thus, the tree is drawn according to the multipliers. The number appearing before *wan* is seen as modifier for *wan* and the number appearing *qian* is seen as a modifier for *qian,* and so on. In this case, since there is no number existing before *qian*, *ling* appears as a place holder.

This numeral also offers a fine example of the usage of *ling*. Since there is a 0 in the thousands place at the beginning of the terminal string of four digits, this fact is indicated by *ling*.

If a 0 occurs in a medial position of an English numeral the morpheme "zero" is not used. For example, 5002 would be pronounced "five thousand and two." In this case the value of the hundreds' place is zero, but this fact is not indicated in the expression of the number. It is clear, then, that although *ling* and "zero" both mean 0, they are not used in the same way in numerical expressions. Interestingly, the number 5002 would be pronounced *wu qian ling er* in Chinese. Here *ling* appears to be parallel to "and" in English, but this is not really so. Although Radzinski (1991, p. 281) observes that *ling* sometimes appears to function as "a type of conjunction," a close examination of the facts does not support this claim, for it is only ever used in numeral expressions to indicate that medial multipliers have the value of 0. In other words, *ling* is a place holder and nothing else. What is especially noteworthy in (35) is how easily *ling* is incorporated into the adjunctive phrasal structure I have proposed for numerical expressions in both English and Chinese.

Ling appears where it does in the structure because it is functioning as a number that is adjoined to the recursive &P. Just as 33,0000 and 200 appear in similar positions before they are adjoined to the first &P containing the conjunct 56. It should be remembered that *ling* is a number, always. *Ling* is not a conjunction similar to English *and*, even though it does appear in a similar position in the structure and it is not an adjective. Neither *ling* nor 33,0000 *modifies* 256; instead, 33,0000 is *added* to 256, and *ling* indicates that the *qian* position is empty. Moreover, in this structure *ling* is adjoined to the number following it (256), not to the number preceding it (33,0000). This is correct, because *ling* indicates the zero that introduces the second string of four numerical places that is conventional for Chinese numerical expressions.

We should not forget here that Hurford (2003, p. 53) states that the word order of numeral expressions is rigid in all natural languages, and that other lexemes almost never interrupts the flow of numerical lexemes. The only possible exception is additive co-ordinate conjunctions, such as *and* in English numerical expressions. Interestingly, Hurford (1975. p. 246) equates *and* with *ling* in Chinese numerical expressions, as in (36), but this is misleading:

(36) a. *iqian* ***ling*** *er shisi*
one-thousand zero two-ten-four
one thousand **and** twenty four

b. san bai **ling** er
three-hundred zero two
three hundred **and** two

In these examples *ling* holds the empty place of 100 in (36a) and 10 in (36b), but “and” does not hold these empty places in the English versions of these numerical expressions. As we have already seen in (4b), “and” always occurs in complex English numerical expressions after the hundreds’ place in each string of three digits, and it always has the semantic value of addition. Thus, “and” would still appear in comparable English numerical expressions, if the hundreds’ place and the tens’ place were not empty, as in (37):

(37) a. one thousand three hundred **and** twenty-four

b. two hundred **and** forty-two

The same is not true for the Chinese expressions of these numerals, where *ling* would not appear at all. This appears to be conclusive evidence that Chinese *ling* is not the equivalent of English “and”, either semantically or syntactically.

5.3 A Brief Summary of My Proposal for Co-ordinate Adjunction

According to Dik (1972: p. 272), “The completely unspecific combinatory value of *and* is . . . the basis of its use in number-names. Indeed, whether we say *one hundred twenty-five* (as in American English) or *one hundred and twenty-five* (as in British English), the *and* adds no more than its purely combinatory value.” Nevertheless, what Dik calls “combinatory value” might be taken to mean that “and” expresses the function of addition itself. Indeed, this is what I have argued throughout this thesis: Numbers are sums and their proper syntactical from is additive co-ordination, accounted for within the framework of X-bar theory as &P adjunctions. Interpreted thus, “and” also has semantic value, and it means something specific, namely that the number that follows “and” is added to the number that precedes it. In fact, there seems to be a special quality about numerical expressions that merges them with the intrinsic nature of numbers as a unique psychological system.

Rutkowski (2003, p. 19) argues that the internal word order of numerical expressions at LF follows neither the rules of syntax nor the rules of semantics. Instead, numerical expressions should be viewed, according to Rutkowski, as arithmetical imports originating outside linguistics proper. If this means that numerical expressions give semantic, syntactic, and phonological form to the arithmetical operation of addition, I wholly agree with him.

CHAPTER 6

CONCLUDING REMARKS

6.1 Summing Up

This thesis is based on the assumption, derived from Hurford, (1975, 1987, 2003), that numbers are sums. This means that numbers are constructed by the operation of addition. Often it is said that numbers are constructed by a combination of addition and multiplication. In fact, Hurford establishes his entire structure for numerical expressions on distinguishing between three arithmetical categories: Numbers (simple digits from 0 to 9), base Multipliers (multiples of 10), and Phrases (sums or combinations of sums). On closer examination, though, it is revealed that all numbers larger than 10 are sums, because even multiplication is a kind of iterated addition. Moreover, it is immediately apparent that this system is mostly – if not entirely – concerned with arithmetical operations. In other words, numerals embody a system that exists beyond the system of language. Thus, the creation of numerical expressions is an attempt to combine arithmetical operations with syntactical, semantic, and phonological operations. This is precisely what Hurford, who is the acknowledged expert on the language of numbers, has done.

Nevertheless, I began this thesis with an ambition to improve on Hurford's analysis of numerical expressions. I intended to do this by identifying the syntactical categories of numbers as abstract nouns. Most authors, Hurford included, have assumed that small lexical numbers behave like adjectives, while decimal multipliers and larger complex numbers behave like nouns. I think, however, that all numbers – as numbers – are nouns. I have held this view from the beginning, and I still hold it. Secondly, in my attempt to integrate arithmetical operations with syntactical operations, I have argued, again from the beginning, that numerals are nominal compounds. Finally, in recognition of the arithmetical origin of numerical expressions, I have argued that numerals are a unique form of discourse: paratactic word strings that consist entirely of nouns, with the exception of the cumulative conjunction & ("and" in English, *you* in Chinese). Numerical word strings are concatenations that exhibit rigid word order.

In my original attempt to expand and revise Hurford's analysis of numerical expressions I relied too heavily on Hurford's structural analysis, although I did improve on the terminology that should be used in phrasal analysis. I did this by indicating that all numbers are nouns and then identifying numbers according to their arithmetical functions in the composition of numerals. But even after I had done this, two serious problems remained. The principal problem associated with Hurford's phrasal structure is that it has to become trenary to accommodate the "and" of English numerical expressions, thus violating the binary branching principle, one of the most important rules of X-bar theory. A second problem is that when Hurford's structure is applied to Chinese numeral expressions, the morpheme *ling*, or "zero," is forced to project as what Hurford calls a Phrase, when logically it should project as what he calls a Number. These issues were not resolved satisfactorily in the first draft of my thesis.

The current version of this thesis does, I believe, solve these problems more convincingly. The reason for the improvement is that, while maintaining my original refinements of Hurford's terminology for syntactical categories, I have adopted Munn's (1993) adjunction analysis of the Conjunction Phrase (&P), what he calls a Boolean Phrase. I was led to this analysis by discovering, through close examination, that the internal composition of English numerical expressions exhibits what might easily be regarded as binary structure. Within each string of three digits in English numerals there is a hundreds' place, followed by the conjunction "and," followed by a combination of the tens' and the units place. Such an arrangement led me to believe that the most appropriate syntactical categorization for numerical expressions is co-ordination.

In this new analysis, achieved within the framework of X-bar theory, and following Munn's argument for co-ordinate adjunction, the conjunction "and" conjoins with its complement, the second conjunct, to form the maximal projection of the &P at the X'' level. This phrase is then Chomsky-adjoined to the first conjunct, leaving the Specifier position empty. The ultimate result is NP^0 containing the entire structure. Applied to numerical expressions, this adjunctive &P analysis goes a long way to solving all the problems associated with Hurford's analysis. There is no more need for trenary branching for the conjunction, since the structure, according to X-bar theory requirements, is hierarchical, not flat. Similarly, *ling* is easily accommodated within this structure when it is applied to Chinese numerical expressions. In short, the &P analysis offers an especially convincing account of English numerical expressions.

The only remaining problem is the hard fact that there is no longer a conjunction equivalent to “and” at PF in Chinese numerical expressions. There is evidence that in the past the morpheme *you* was used as a conjunctions in Chinese numerals, but that is no longer the case. How can Chinese numeral expressions be analyzed as &P if they do not contain the category &? My answer is that the conjunction *you* is still present at LF. Moreover, the internal logic of the composition of numerals in all languages is essentially arithmetical addition, which in turn suggests syntactical co-ordination. If this is true, *you* can be used as a conjunction in Chinese numerical expressions as long as it is marked as phonetically null. If this provision can be accepted, the &P analysis as a whole, based on Munn’s application of right adjunction, provides a satisfactory account of both English and Chinese numerical expressions.

Positing a phonetically null conjunction to head a phrase is not entirely radical. Kayne (1994), for instance, does this to license the first conjunct in his co-ordinate structure, but I am trying to justify the merging of e-& with the second conjunct. Most authors would probably question this possibility. In order to prove that Chinese numerical expressions are headed by a covert conjunction, or e-&, I have relied on two related arguments. The first is related to the concept of isomorphism between syntactical and phonetic forms. There is, no doubt, a possible interface there, and I have argued that intonational pauses between segments of numerical expressions in Chinese allow us to identify or recover the deleted conjunction *you*. I have also argued, following Tokizaki (2005), that covert conjunction can be used to integrate the various parts of a discourse, even entire sentences, according to the asymmetrical properties of phrasal co-ordination under standard X-bar theory. Since I also argue that numerical expressions are a special kind of paratactic and concatenated form of discourse, I believe that I am justified in concluding that *you* is covertly present in the & slot of &P in Chinese numerical expressions. We probably should remember here that, as Haegman (1994, pp. 7-8) observes, what is acceptable in language is not necessarily what is grammatical – and *vice versa*. An overt & might not be acceptable in Chinese speech, but it still might be grammatical.

6.2 Postscript: Hurford Revisited

As I have just demonstrated, this thesis is an attempt to come to terms with Hurford's (1975, 1987, 2003) analysis of numerical expressions. It first seemed to me that his structural interpretation of numerals was syntactically vague and incomplete. And so it is. I have improved it by identifying all the elements as nouns and by combining arithmetical and syntactical operations in my analysis. I have also managed to present what I think is a solid – and at the same time original – argument for analyzing numerical expressions as asymmetrical co-ordinate phrasal structures. Moreover, I have demonstrated that this analysis works for both English and Chinese numerical expressions. Nevertheless, I am left with a profound suspicion that perhaps Hurford really has said all there is to say about numerical expressions, and that he was wise to keep the grammar of numerals separate from the general grammar. One thing is certain: numerals are special words.

REFERENCES

Alharbi, Abdullah H. (2002). The syntax of coordinate constructions. *Language and Translation*, 14, 67-87. Retrieved February 11, 2009. http://digital.library.ksu.edu.sa/V14M172R450.pdf

Bošković, Željko and Franks, Steven. (2000). Across-the-board movement and LF. *Syntax*, 3 (2), 107-128.

Brainerd, Barron and Peng, Fred C.C. (1968). A syntactic comparison of Chinese and Japanese numerical expressions. In Hugo Brandt Corstius, (Ed). *Grammars for Number Names.* Reidel: Cordrecht, 53-81.

Carston, Robyn and Blakemore, Diane. (2005). Introduction to coordination: Syntax, semantics and pragmatics. In Diane Blakemore and Robyn Carston (Eds.). *Coordination: Syntax, Semantics, and Pragmatics*. *Lingua,* 115 (4), 353-358.

Chomsky, Noam. (1965). *Aspects of the Theory of Syntax.* Cambridge: MIT Press.

Chomsky, Noam. (1986). *Barriers* (Linguistic Inquiry Monograph 13). Cambridge: MIT Press.

Collins, Christopher. (1988a). Part 1. Conjunctions adverbs. Manuscript. Cambridge: MIT.

Collins, Christopher. (1988b). Part 2. Alternative analysis of conjunction. Manuscript. Cambridge: MIT.

Corbett, Greville G. (1978). *Number*. Cambridge: Cambridge University Press.

Cormack, Annabel and Smith, Neil. (2005). What is coordination? In Diane Blakemore and Robyn Carston (Eds.). *Coordination: Syntax, Semantics, and Pragmatics*. *Lingua,* 115 (4), 395-418.

Dik, Simon C. (1968). *Coordination: Its Implications for the Theory of General Linguistics.* Amsterdam: North-Holland.

Dougherty, Ray C. (1969). Review of *Coordination: Its Implications for the Theory of General Linguistics.* By Simon C. Dik. *Language,* 45 (3), 624-636.

Duarte, Ines. (1991). X-bar theory: Its role in GB theory. In J. Siekmann (Ed.), *Natural Language Processing.* Lecture Notes in Computer Science. Berlin: Springer-Verlag. 25-54. http://www.springerlink.com/content/p61w361384p8504r/fulltext.pdf

Gazdar, Gerald; Klein, Ewan; Pullum, Geoffrey; and Sag, Ivan. (1985). *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell

Goodall, Grant. (1987). *Parallel Structures in Syntax: Coordination, Causatives and Restructuring.* Cambridge: Cambridge University Press.

Greenberg, Joseph H. (1978). Generalizations about numeral systems. In Joseph H. Greenberg (Ed.). *Universals of Human Language. Vol. 3: Word Structure.* Stanford: Stanford University Press.

Grosjean, François; Grosjean, Lysiane; and Lane, Harlan. (1979). The patterns of silence: Performance structures in sentence production. *Cognitive Psychology*, 11 (1), 58-81.

Grosu, Alexander. (1973). On the nonunitary nature of the coordinate structure constraint. *Linguistic Inquiry*, 4 (1), 88-92.

Haegeman, Liliane. (1994). *Introduction to Government and Binding Theory*. Oxford: Blackwell.

Hartmann, Katharina. (2000). *Right Node Raising and Gapping.* Philadelphia: John Benjamins.

Haspelmath, Martin. (2005). Coordination. In Timothy Shopen (Ed.). *Language Typology and Linguistic Description.* Cambridge: Cambridge University Press. Accessed April 27, 2009. http://email.eva.mpg.de/~haspelmt/coord.pdf

Hudson, Richard A. (2003). Coordination. In *An Encyclopedia of English Grammar and Word Grammar*. http://www.phon.ucl.ac.uk/home/dick/enc-gen.htm

Hurford, James R. (1975). *The Linguistic Theory of Numerals.* Cambridge: Cambridge University Press.

Hurford, James R. (1980). A note on Corbett's numeral universal. *Lingua*, 50 (3), 247-248.

Hurford, James R. (1987). *Language and Number: The Emergence of a Cognitive System*. Oxford: Basil Blackwell.

Hurford, James R. (2001). Numeral systems. In *International Encyclopedia of the Social and Behavioral Sciences*, edited by Neil J. Smelser and Paul B. Baltes. Pergamon, Amsterdam. pp. 10756-10761. http://www.ling.ed.ac.uk/~jim/numsys.pdf

Hurford, James R. (2003). The interaction between numerals and nouns. In Frans Plank (Ed.) *Noun Phrase Structure in the Languages of Europe* (one of ten volumes to emerge from the European Science Foundation EUROTYP project) pp. 561-620. http://www.ling.ed.ac.uk/~jim/eurofinal.pdf

Jackendoff, Ray. (1977). *X' Syntax: A Study of Phrase Structure*. Cambridge: MIT Press.

Johannessen, Janne Bondi. (1998). *Coordination*. Oxford: Oxford University Press.

Johnson, Kyle. (2002). Restoring exotic coordinations to normalcy. *Linguistic Inquiry*, 33 (1), 97-156.

Kayne, Richard. (1984). *Connectedness and Binary Branching*. Dordrecht: Foris

Kayne, Richard. (1994). *The Antisymmetry of Syntax*. Cambridge: MIT Press.

Kornai, András and Pullum, Geoffrey K. (1990). The X-bar theory of phrase structure. *Language,* 66 (1), 24-50.

Kubo, Yoshihiro. (2007). A note on typology of coordination. Retrieved May 3, 2009. http://www.adm.fukuoka-u.ac.jp/fu844/home2/Ronso/Jinbun/L38-4/L3804_1229.pdf

Li, Charles N. and Thompson, Sandra A. (1981). *Mandarin Chinese: A Functional Reference Grammar*. Berkeley and Los Angeles: University of California Press.

Liu, Jian and Peyraube, Alain. (1994). A history of some coordinate conjunctions in Chinese. *Journal of Chinese Linguistics,* 22 (2), 179-201.

Liu, Yuehua et al. (1996). *Modern Chinese Grammar*. Shi Da Shu Yuan.

Menninger, Karl. (1969). *Number Words and Number Symbols: A Cultural History of Numbers.* Dover.

Munn, Alan. (1993). *Topics in the Syntax and Semantics of Coordinate Structures*. College Park: University of Maryland dissertation. Retrieved January 26, 2009. https://www.msu.edu/user/amunn/psfiles/thesis.pdf

Phillips, Colin. (2003). Linear order and constituency. *Linguistic Inquiry,* 34 (1), 37-90.

Pollard, Carl. (1984). *Generalized Phrase Structure Grammars, Head Grammars, and Natural Language*. Stanford: Stanford University dissertation.

Poncinie**,** Lawrence. (1993). Aristotle and Bressan on a number of things. *Erkenntnis,* 39 (2), 129-144. Retrieved January 17, 2009. http://www.springerlink.com/content/wn885007r416q102/fulltext.pdf

Radzinski, Daniel (1991). Chinese number names, tree adjoining languages, and mild context-sensitivity. *Computational Linguistics,* 17 (3), 277-299. Retrieved January 6, 2009. http://acl.ldc.upenn.edu/J/J91/J91-3002.pdf

Reinhart, Tanya. (1976). *The Syntactic Domain of Anaphora*. Cambridge: MIT dissertation. Retrieved April 14, 2009. http://dspace.mit.edu/handle/1721.1/16400

Ross, John Robert. (1967). *Constraints on Variables in Syntax.* Cambridge: MIT dissertation. http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/38/b3/9d.pdf

Rutkowski, Pawel. (2003). On the universal neuropsychological basis of the syntax of numerals. *Journal of Universal Language,* 4 (2), 147-182.

Sag, Ivan; Gazdar, Gerald; Wasow, Thomas; and Weisler, Steven. (1985). Coordination and how to distinguish categories. *Natural Language and Linguistic Theory*, 3, 117-171.

Tokizaki, Hisao. (2005). Pause and hierarchical structure in sentence and discourse. *Proceedings of IDP05*. Retrieved May 19, 2009. http://aune.lpl.univ-aix.fr/~prodige/idp05/actes/tokizaki.pdf

Winter, Yoad. (1995). Syncategorematic conjunction and structured meanings. *Proceedings of Semantics and Linguistic Theory*. Retrieved March 2, 2009. http://www.cs.technion.ac.il/~winter/papers/syncat.pdf

Yang, Henrietta Shu-Fen. (2005). *Plurality and Modification in Mandarin Nominal Phrases*. Austin: University of Texas dissertation. Retrieved April 13, 2009. http://repositories.lib.utexas.edu/bitstream/handle/2152/2198uamgd17828.pdf

Zhang, Niina. (2006). On the configuration issue of coordination. *Language and Linguistics,* 7 (1), 175-223.

Zoerner, Cyril Edward III. (1995). *Coordination: The Syntax of &P*. Irvine: University of California Dissertation.