**Word order variation and *ba* sentences in Chinese***

Feng-hsi Liu
University of Arizona

**Abstract**

Despite extensive research on *ba* sentences in Chinese, the issue of when *ba* sentences are used in discourse has received little attention. This study examines the word order variation involving *ba* sentences by comparing three word orders: the canonical postverbal form, the *ba* form, and the topicalized preposed form. I show that the choice of the *ba* form depends on multiple factors, including information status, weight and topicality. The *ba* form is more likely to be used under two conditions: (a) when the *ba* NP carries old information but is not highly topical, (b) when the *ba* NP carries new information and is heavy. Further, my findings raise doubts on the *ba* NP's role as a topic in discourse.

## 1. Introduction

This paper examines *ba* sentences in Chinese with respect to the following question: When are *ba* sentences used in discourse? As one of the better-studied constructions in Chinese, the *ba* construction has been examined in just about every aspect: structural properties, semantic constraints, historical development, acquisition, the category status of *ba* itself, and grammaticalization of *ba*. Further, its properties have been shown to be significant for the study of typology and word order change (Li and Thompson 1974, 1975, Travis 1984), the Principle and Parameter Approach (Koopman 1983, Li 1985, Sybesma 1992), transitivity (Sun 1995, Thompson 1973), topic-comment (Hsueh 1987, Tsao 1987), and aspectual and event structure (Liu 1997,Yong 1993). However, curiously, to my knowledge the question of when *ba* sentences are actually used has received little attention. A number of studies (Li and Thompson 1975, Sun and Givón 1985) have looked at the relationship between word order and definite vs. indefinite NPs, including the *ba* NP; however, the specific question of under which conditions the *ba* form, rather than the canonical postverbal non-*ba* form, is used in discourse has not been examined. This paper is a first step toward understanding this issue.

In a *ba* sentence, the object which normally follows the verb occurs preverbally, marked by *ba*, as illustrated in (1).

(1) Yushui ba yifu nongshi -le
rain BA clothes make-wet-PERF
'The rain got the clothes wet.'

The preverbal order with *ba* and the canonical order, where the object occurs postverbally, sometimes give the same meaning, as (1) and (2), and sometimes not, as (3a) and (3b)[1]:

* I would like to thank George Bedell and Peter Shapley for helpful discussion. I also thank the anonymous reviewers for valuable comments. Thanks also to Mark Borgstrom at the Center for Computing and Information Technology, University of Arizona, for his help with statistics, and to Haihua Pan for providing me with part of the corpus data used in this project.

(2) Yushui nongshi -le yifu
rain make-wet-PERF clothes
'The rain got the clothes wet.'

(3) a. Wo xiang ba sange xuesheng songzou
I want BA three students send-away
'I want to send away three (particular) students.'

b. Wo xiang songzou sange xuesheng
I want send-away three students
'I want to send away (any) three students.'

In (3a) the object refers to three particular students, while in (3b) the object refers to any three students. Thus the two sentences have different meanings. When the same meaning is involved, as in (1) and (2), we have a case of word order variation. In addition, when an object occurs preverbally, it may or may not occur with *ba*; in the latter case, the object may occur before or after the subject or without the subject, as in (4a), (4b), and (4c) below:

(4) a. Nei feng xin wo du -le haoji bian
that CL letter I read -PERF several times
'That letter I read several times.'

b. Wo nei feng xin du -le haoji bian
I that CL letter read -PERF several times
'I that letter read several times.'

c. Nei feng xin du -le haoji bian
that CL letter read -PERF several times
'That letter (I) read several times.'

Thus variation also exists among preverbal forms, both in terms of word order and the marking of *ba*. This study will not consider the form represented by (4b), where the preposed object NP occurs after the subject. There are only a few instances of this form in my data. The preverbal form we will consider, besides the *ba* form, is one where the object NP occurs at the beginning of a sentence, followed by the subject, as in (4a), or without the subject, as in (4c). There are therefore three forms to examine in this study, represented by (1), (2) and (4a, c) respectively. We will call the three forms the *ba* form, the postverbal form, and the preposed form; the relevant object NPs will be called the *ba* NP, the postverbal NP, and the preposed NP. To find out when *ba* is used, we need to consider two issues: (a) What makes an object occur preverbally rather than postverbally? (b) For preverbal objects, what makes them occur with *ba* rather than at the beginning of a clause, as a topic? The first issue concerns word order variation, while the second concerns variation of both word order and *ba* marking. We will examine both issues in this study.

---

[1] Although most of the examples cited are from natural spoken or written discourse, for expository purposes I also use constructed examples. However, the constructed sentences are not included in the database, and play no part in the analysis.

In the literature word order variation is often discussed in the context of information status. Constituents carrying old information are placed earlier in a sentence, while constituents carrying new information are placed later in a sentence. The correlation between information status and the position of a constituent has been found in many languages (Birner and Ward 1998, Chafe 1976, Gundel 1988, among others). Our first question on the variation of *ba*, therefore, is whether information status affects word order in Chinese. Does a preverbal object, such as the *ba* NP, express old information?

Besides information status, it has also been shown that weight plays a role in word order variation (Hawkins 1994, Quirk et al. 1972). Constituents that are heavy tend to occur later in a sentence, whereas constituents that are light tend to occur earlier in a sentence. Our second question, then, is whether the preverbal vs. postverbal variation is affected by weight.

Arnold et al. (2000) show that both information status and weight influence word ordering in English heavy NP shift and dative alternation. We will therefore also compare information status with weight to see if both factors are needed to account for the word order variation of *ba* sentences.

We will then turn to the variation between the *ba* form and the preposed form. I will explore and propose factors that could affect the choice between the two forms. I will also briefly review the proposal that the *ba* NP is a topic (Tsao 1987), and consider whether the notion of topic is essential in characterizing the *ba* NP.

## 2. Preliminaries

### 2.1 Basic properties of *ba* sentences

First, a few words about *ba* sentences are in order. As is well known, *ba* sentences are subject to a number of syntactic and semantic constraints. Not all transitive sentences can be expressed in the *ba* form. Both the object and the predicate have to satisfy certain conditions. In general, it is assumed that the *ba* NP is definite or specific, and the predicate is complex, containing more than a bare verb. I will follow my earlier analysis (Liu 1997) that the occurrence of *ba* is subject to aspectual constraints. In particular, the *ba* predicate expresses a bounded event. In the above, (1) and (3a) both express a bounded event, and therefore can occur with *ba*; by contrast, (5a) expresses an unbounded event, and cannot occur with *ba*, as (5b) shows:

(5) a. Wo zai xie wode lunwen
I PROG write my dissertation
'I'm writing my dissertation.'

b. *Wo zai ba wode lunwen xie
I PROG BA my dissertation write
'I'm writing my dissertation.'

On the other hand, when a *ba* sentence is used, it does not mean that its postverbal counterpart is always possible. As Lü (1984) notes, in certain structural environments the object must occur preverbally. Two examples are given below:

(6) Ta shuo youde ren ba anhui de bengbu niancheng bangbu
he said some people BA Anhui DE Bengbu read-as Bangbu
'He said some people read "Bengbu" in Anhui (province) as "Bangbu".'

(7) Wode pengyou ba wo he yuanzhumin zuojia zuo bijiao
my friend BA me and native writer make comparison
'My friends compared me with the native writers.'

In (6), the two components of the verb compound *niancheng* 'read as' cannot be separated and there is another object following the compound. Therefore, the object *anhui* 'Anhui (province)' cannot occur postverbally. (7) is another case where there is no place for the *ba* NP postverbally. The object position is already occupied by another NP *bijiao* 'comparison'. Because a major concern of this study is to examine word order variation, sentences where the postverbal form is unavailable, such as (6-7), are excluded from my data.

**2.2 Previous studies**

As mentioned earlier, despite the large amount of literature on the *ba* construction, there have been very few studies that consider the issue of when *ba* is used in discourse. Li and Thompson (1981: 482-490) offer two conditions where it would be appropriate to use *ba*, given in (8):

(8) (a) When the *ba* NP is prominent— definite, specific or generic
(b) When the sentence expresses disposal— something happening to the entity referred to by the *ba* NP.

Whether the *ba* sentence will be used depends on the degree to which these two conditions are satisfied. The more strongly the two conditions are met, the more likely it is for *ba* to be used. Ho (1993:109-114) suggests that the *ba*-construction is used as a device of focalization and thematization. The object is moved preverbally so that whatever element is left at the end of a sentence can receive focus; meanwhile, once preposed via the *ba*-construction, the object can receive thematic status if it is further preposed to the sentence-initial position.

Neither proposal is based on a wide range of data. Li and Thompson's proposal essentially links *ba* to transitivity, since prominent NPs and strong sense of disposal are correlates of high transitivity (Hopper and Thompson 1980). High transitivity is indeed characteristic of *ba* (Liu 1999, Sun 1995, Thompson 1973); however, I believe transitivity has more to do with when *ba* **can** be used, rather than when it **is** used. In fact, virtually all of the tokens in this study are highly transitive, in both the *ba* form and the non-*ba* forms. In this study I will approach the issue of what influences the choice of the *ba* form as a case of word order variation. I will examine two factors that have been shown to be relevant for word order variation: information status and weight.

## 3. Data

### 3.1 Data source

The data used in this study consists of spoken and written language, both covering a variety of styles. The spoken data include conversations at TA meetings, speeches, TV interviews and commentaries, as well as conversations among friends (from Pan 1996). The written data consists of portions of a contemporary novel, on-line articles from China, and on-line articles published by a Taiwanese newspaper. These articles cover a wide range of genres, including commentaries, narratives, memoirs, letters, and news reports. Altogether this constitutes about 400,000 characters in transcripts, of which 150,000 characters are spoken data and 250,000 characters are written data.

### 3.2 Data selection

For inclusion as data, I collected sentences that have the following property: sentences that allow variation between the canonical postverbal form and the *ba* form without changing the meaning. Sentences with this property belong to one of the following three categories in (9):

(9) (a) *Ba* sentences that could be expressed in the canonical postverbal form.
(b) Sentences in the canonical postverbal form that could be expressed in the *ba* form.
(c) Sentences in the preposed form that could be expressed in both the *ba* form and the canonical postverbal form.

An instance of the first category can be seen in (1-2). (10a) is an example of a sentence in the postverbal form that could be expressed by the *ba* form, as in (10b), and (11a) is an example of a sentence in the preposed form, which could be expressed in the *ba* form, as (11b), or in the postverbal form, as in (11c):

(10) a. Ta fenxi -le yixia xianzai de qingquang
he analyze -PERF a-little now DE situation
'He analyzed the current situation a little.'

b. Ta ba xianzai de qingquang fenxi -le yixia
he BA now DE situation analyze -PERF a-little
'He analyzed the current situation a little.'

(11) a. Zeren nong qingchu
responsibility make clear
'Clarify the responsibilities.'

b. Ba zeren nong qingchu
BA responsibility make clear
'Clarify the responsibilities.'

c. Nong qingchu zeren
make clear responsibility
'Clarify the responsibilities.'

These three categories cover all possibilities where a choice of the canonical form or the *ba* form is possible. The selection results in 250 tokens in the *ba* form, 159 tokens in the postverbal form and 47 tokens in the preposed form. Altogether there are 456 tokens, which form the database for this study.

## 4. Preverbal vs. postverbal

I will take a two-layered approach in my analysis, examining first the postverbal vs. preverbal variation (this section), and then the variation between the *ba* form and the preposed form (section 5). In the analysis of the postverbal vs. preverbal variation, the *ba* form and the preposed form are combined as one category— preverbal. There are two reasons for doing so. First, in studies of Chinese word order (e.g. Chao 1968, Givón and Sun 1985, Li and Thompson 1974, 1975) the distinction between preverbal and postverbal has been considered a deciding factor for how an NP is distributed or interpreted. Li and Thompson, for example, claim that

nouns preceding the verb tend to be definite, while nouns following the verb tend to be indefinite. Therefore, a natural question to ask is whether the preverbal vs. postverbal distinction also plays a role in word order variation with respect to *ba* sentences. The second reason comes from statistical results. A loglinear analysis was performed, using SPSS version 12, to determine if the *ba* form and the preposed form can be combined without significantly changing the relationship between the three forms with regard to the factors of information status and weight. The likelihood ratio (LR) chi-square value for the three-category analysis is 254.19, with 12 degrees of freedom (df), and that for the two-category analysis is 244.82, with 7 df. The difference is 9.37, with 5 df. The criterion chi-square value for p of 0.05 and 5 df is 11.07. Since 9.37 is smaller than 11.07, we conclude that it is appropriate to collapse the *ba* form and the preposed form.

Given the justification from two perspectives, I will proceed to combine the *ba* form and the preposed form in the analysis of the preverbal vs. postverbal variation. The two factors that we will consider are information status and weight. Before we consider the variation, however, we will discuss how old vs. new is distinguished and how weight is measured.

### 4.1 Old vs. new

There have been many approaches on what is considered 'old' and what is considered 'new' in the literature. (Chafe 1976, Gundel 1988, Prince 1981, 1992, among others). I will adopt the criteria offered by Prince (1992), who makes a distinction between hearer-old/new and discourse-old/new. Hearer-old information refers to information that the speaker believes is known to the hearer, while hearer-new refers to information that the speaker believes is not known to the hearer. Discourse-old information is information that has already been evoked in the discourse; discourse-new refers to information that has not been evoked in the discourse. In addition, there is also information that has not been evoked but can be inferred. I will treat inferable information as old information. The main motivation for treating such inferable information as old comes from the fact that in my data the entities carrying inferable information pattern like entities carrying discourse-old or hearer-old information. This will be further discussed in 4.5.2.

There are thus four logical possibilities in terms of hearer vs. discourse and old vs. new: (a) hearer-old and discourse-old, (b) hearer-old and discourse-new, (c) hearer-new and discourse-new, (d) hearer-new and discourse-old. According to Prince, (d) is not found in natural data. For our purposes, if information is hearer-old or discourse-old, it is considered old. For convenience in what follows besides referring to information as old or new, I will also refer to the NPs that carry the information as old NPs or new NPs.

### 4.2 Weight

Next, we look at the second variable of word order variation— weight. In the literature different methods of measuring weight have been proposed, e.g. in terms of length (number of words) or syntactic complexity (number of nodes). (See Wasow 2002 for an evaluation of different measures.) Wasow (1997) finds that different measures actually don't give conflicting results; length and complexity are both strong predictors of variation involving heavy NP shift and dative alternation in English. In addition, Wasow (2002) demonstrates that complexity is a factor of weight independent of length. He also shows that both absolute weight and relative weight are relevant. In this study we will consider absolute weight only, and we will use length as the measure for weight, leaving aside other measures (relative weight and complexity) for

future study. A constituent is considered heavy if it is long. However, length in Chinese will be measured not by counting the number of words, but by counting the number of syllables. There are two reasons why we don't use word as a measuring unit. First, word is notoriously difficult to define; secondly, the orthography does not give a visual clue, as written words in Chinese are not separated from each other the way words in English are. Since in Chinese a syllable corresponds to a morpheme, and at the same time a syllable corresponds to a character, this measure amounts to counting the number of characters on transcripts, which is the standard way of measuring length in Chinese.

### 4.3 Coding

Each *ba* NP, postverbal NP, and preposed NP was coded for two properties: information status and weight. NPs were coded as old in the following situations:

a) NPs evoking entities already known to the hearer/reader.

b) NPs evoking entities that are discourse-old or inferable.

Other NPs were coded as new.

Weight was coded as 'heavy', 'medium,' and 'light'. The assignment was arbitrarily given as follows:

a) light: 1-5 characters

b) medium: 6-10 characters

c) heavy: 11 characters and above

### 4.4 Results

For the preverbal vs. postverbal distinction, I postulated information status and weight as two possible factors. In this section we will see if the two factors are significant. First, however, we observe some general patterns. Old NPs and new NPs have rather different distributions. As shown in Table 1, old NPs cluster around light NPs. This is the case for all three forms.

| | *ba* | % | preposed | % | postverbal | % |
|---|---|---|---|---|---|---|
| light NPs | 205 | 90.7% | 44 | 93.6% | 75 | 90.4% |
| medium NPs | 20 | 8.8% | 3 | 6.4% | 5 | 6.0% |
| heavy NPs | 1 | 0.4% | 0 | 0.0% | 3 | 3.6% |
| total | 226 | 100.0% | 47 | 100.0% | 83 | 100.0% |

Table 1: Distribution of old NPs

This fact is in accordance with the well-known correlation between information status and weight. That is, old NPs tend to be light. By contrast, new NPs exhibit two patterns between the *ba* form and the postverbal form, there being no new NPs in the preposed form. In the *ba* form, the new NPs are scattered through the medium and heavy NPs; in the postverbal form, the new NPs are mostly in the light and medium NPs. This is seen in Table 2:

| | *ba* | % | preposed | % | postverbal | % |
|---|---|---|---|---|---|---|
| light NPs | 0 | 0.0% | 0 | --- | 43 | 56.6% |
| medium NPs | 10 | 41.7% | 0 | --- | 27 | 35.5% |
| heavy NPs | 14 | 58.3% | 0 | --- | 6 | 7.9% |
| total | 24 | 100.0% | 0 | --- | 76 | 100.0% |

Table 2: Distribution of new NPs

Tables 3, 4, and 5 give the distribution in terms of weight. Table 3 shows that in the *ba* form and the preposed form, all of the light NPs are old; in the postverbal form, however, 41.4% of the light NPs are new.

| | *ba* | % | preposed | % | postverbal | % |
|---|---|---|---|---|---|---|
| old NPs | 205 | 100.0% | 44 | 100.0% | 75 | 58.6% |
| new NPs | 0 | --- | 0 | --- | 43 | 41.4% |
| total | 205 | 100.0% | 44 | 100.0% | 128 | 100.0% |

Table 3: Distribution of light NPs

Table 4 shows that medium NPs are mostly old in the *ba* form, but mostly new in the postverbal form.

| | *ba* | % | preposed | % | postverbal | % |
|---|---|---|---|---|---|---|
| old NPs | 20 | 66.7% | 3 | 100.0% | 5 | 15.6% |
| new NPs | 10 | 33.3% | 0 | --- | 27 | 84.4% |
| total | 30 | 100.0% | 3 | 100.0% | 32 | 100.0% |

Table 4: Distribution of medium NPs

Finally, Table 5 shows that heavy NPs do not occur in the preposed form and they are mostly new in both the *ba* form and the postverbal form.

| | *ba* | % | preposed | % | postverbal | % |
|---|---|---|---|---|---|---|
| old NPs | 1 | 6.7% | 0 | --- | 3 | 33.3% |
| new NPs | 14 | 93.3% | 0 | --- | 6 | 66.7% |
| total | 15 | 100.0% | 0 | --- | 9 | 100.0% |

Table 5: Distribution of heavy NPs

Together, Tables 3, 4 and 5 show the other side of the correlation between information status and weight: light NPs tend to be old, and heavy NPs tend to be new.

We now consider the first question: Is old vs. new a significant factor in the preverbal vs. postverbal variation? To answer this question, we check how old NPs are distributed and how new NPs are distributed separately. Table 6 gives the results:

| | preverbal | % | postverbal | % | total | % |
|---|---|---|---|---|---|---|
| old NPs | 273 | 76.7% | 83 | 23.3% | 356 | 100% |
| new NPs | 24 | 24.0% | 76 | 76.0% | 100 | 100% |

Table 6: Distribution of old and new NPs

76.7% of old NPs occur preverbally, while only 24.0% of new NPs occur preverbally. A chi-square value of 93.11 ($p < 0.001$) shows that this factor is highly significant. Therefore, we can conclude that information status is a factor in the preverbal vs. postverbal variation.

The second question concerns whether weight is also a factor. The findings are given in Table 7:

| | preverbal | % | postverbal | % | total | % |
|---|---|---|---|---|---|---|
| light NPs | 249 | 67.8% | 32.2% | | 367 | 100% |
| medium NPs | 33 | 51.6% | 48.4% | | 65 | 100% |
| heavy NPs | 15 | 62.5% | 37.5% | | 24 | 100% |

Table 7: Distribution of light, medium, and heavy NPs

Table 7 shows that more of the light NPs occur preverbally, and the same is true of medium and heavy NPs. This means regardless of the weight of the NP, it tends to occur preverbally. This tendency is stronger for light NPs; however, a chi-square value of 7.17 ($p < 0.028$) shows that the difference is only weakly significant.

Since the number of tokens in medium NPs and heavy NPs is small, a loglinear test was performed, using SPSS version 12, to see if the two can be combined. The results show that collapsing medium and heavy would significantly change the relationships among light, medium and heavy NPs. The chi-square value for the three-category distinction is 254.19, 12 df, while that for the two-category distinction is 232.95, 7df. The difference in value is 21.24, 5 df, and the criterion chi-square value for p of 0.05 and 5 df is 11.07. Since 21.24 is greater than 11.07, the difference between the three-category distinction and the two-category distinction is significant. Therefore, in Table 7 above, medium and heavy NPs are kept separate.

A further test was performed to see if any of the interactions between old/new, weight and word order (preverbal vs. postverbal) is significant. Hierarchical loglinear models (Fienberg 1978, p. 38) were fit to the data. This is given in Table 8. It includes eight models. A model that fits the data would have a LR chi-square value that is not significant. The first model is one where there is no interaction among the three variables. In the second model, preverbal/postverbal is independent, while old/new interacts with weight (indicated by 'X'). None of the eight models provide an adequate fit to the data, as indicated by the LR chi-square values. Therefore, the three-way interaction between preverbal/postverbal, old/new and weight is deemed the only appropriate model. That is, the three-way interaction of the variables is significant.

| Model | LR X2 | df |
|---|---|---|
| [pre/post][old/new][weight] | 244.82* | 7 |
| [pre/post][old/new X weight] | 140.73* | 5 |
| [pre/post X old/new][weight] | 151.96* | 6 |
| [pre/post X weight][old/new] | 237.91* | 5 |

| | | |
|---|---|---|
| [pre/post X weight][old/new X weight] | 133.81* | 3 |
| [pre/post X old/new][old/new X weight] | 47.87* | 4 |
| [pre/post X old/new][pre/post X weight] | 145.05* | 4 |
| [pre/post X old/new][pre/post X weight][old/new X weight] | 29.66* | 2 |

*Significant at the 0.05 level

Table 8: Loglinear models fit to the data

Figure 1 is a graphical depiction of the three-way interaction, and Table 9 provides the distribution of old/new and weight combination, as mapped out in Figure 1.

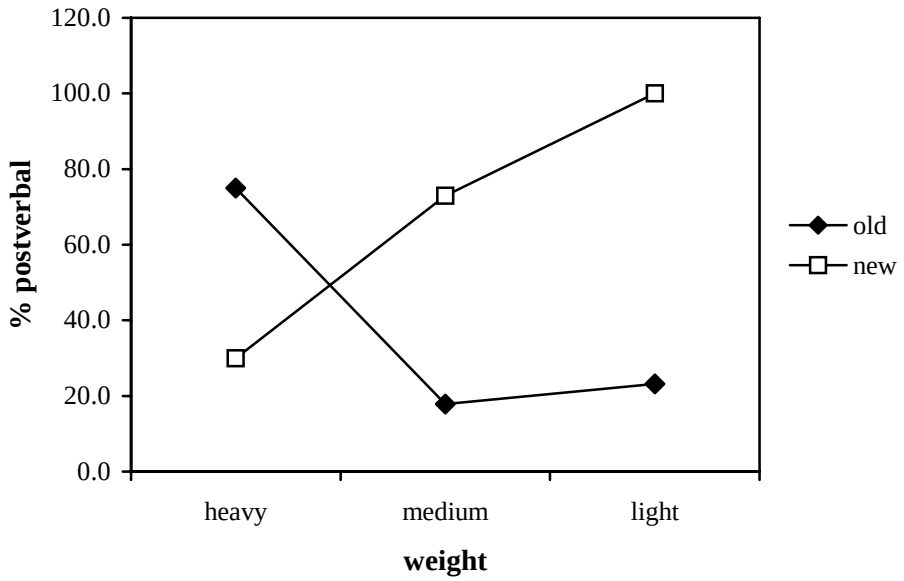


Figure 1: Three-way interaction of preverbal/postverbal, old/new and weight

| | preverbal | % | postverbal | % | total | % |
|---|---|---|---|---|---|---|
| old & heavy | 1 | 25.0% | 3 | 75.0% | 4 | 100% |
| new & heavy | 14 | 70.0% | 6 | 30.0% | 20 | 100% |
| old & medium | 23 | 82.1% | 5 | 17.9% | 28 | 100% |
| new & medium | 10 | 27.0% | 27 | 73.0% | 37 | 100% |
| old & light | 249 | 76.9% | 75 | 23.1% | 324 | 100% |
| new & light | 0 | 0.0% | 43 | 100.0% | 43 | 100% |

Table 9: Distribution of NPs with old/new and weight combined

## 4.5 Discussion

### 4.5.1 Analysis of results

The results indicate that information status (old/new) is a significant factor for the variation between the preverbal and postverbal word orders; however, weight by itself is only weakly significant. The results further show that the three-way interaction between word order,

information status and weight is significant. This means that although there is a relationship between information status and weight, neither factor can be ignored.

A number of observations can be made here. First, new, when interacting with light, shows a much stronger effect than new alone. Recall, from Table 6, that new NPs tend to be postverbal (76%), but Table 9 shows when new interacts with light, this tendency is strengthened to an absolute rule (100%). All 43 of the new & light occur postverbally. (12) is an example:

(12) Fanzheng manfenr shi wufenr, en, na chayidianr, ni sanfenr,
in-any-case full-score is 5-points hmm that less you three-points

erfenr dou xing. Yaoshi neng quick de, jiu jilu yixia tade
two-points all ok if can quick DE then record a-little his

wenti ye keyi.
problem also ok

'In any case the full score is 5 points. Hmm. If less than that, either you give 3, or 2 is ok. If you can be quick, writing down their problems would also be good.'

The object *tade wenti* 'his/their problems' is light and new.

Second, new & heavy, by contrast, imposes a drastically different change on new alone. The postverbal tendency exhibited by new alone is reversed when new interacts with heavy. It is much more likely to be preverbal (70%) than postverbal (30%). (13) is an example of new & heavy occurring preverbally, in the *ba* form:

(13) Zai yuequ tingzhi de shunjian, keyi tingjian cong beifang lai de
at music stop DE short-time can near from north come DE

feng wu wu de zai kongzhong fachu menxiang, bushide guozhe
wind wu wu de at air-in emit unhappy-noise occasionally wrap-DUR

yun ba chabuduo yibai ceng gao de huangjua yinhang dingduan
clouds BA about 100 stories high DE royal bank top

na juda de guaishou xing yinhang huiji tunmo.
that huge DE strange-animal shape bank logo swallow

'During the short time when the music stopped, you can hear the wind coming from the north, making the "wu wu" sound in the air. It occasionally wraps the clouds and swallows the huge bank logo in the shape of a strange animal, above the Royal Bank, about 100 stories high.'

The object in (13) is 25 characters long. This is an extremely heavy *ba* NP, partly because the example comes from written data.

Reversal of tendency can also be seen in old & heavy. Table 6 shows that old NPs tend to be preverbal (76.7%), but when old NPs happen to be heavy, the tendency is shifted to postverbal, as Table 9 shows. However, the number of tokens for old & heavy is too small (4) for this observation to be reliable. As for old & medium and new & medium, the trend reflects the same trend for the single factor old vs. new. This shows that medium weight does not have much impact on how NPs distribute, reflecting the same results shown in Table 7.

Finally, Figure 1 shows why medium and heavy should not be collapsed into one category. The two categories exhibit opposite patterns in the three-way interaction. While old & medium tends to be preverbal, old & heavy tends to be postverbal; similarly, while new & medium tends to be postverbal, new & heavy tends to be preverbal. This is also supported by the LR chi-square test, discussed in 4.4.

### 4.5.2 Inferable NPs

As mentioned earlier, inferable NPs, i.e. NPs that have not been evoked previously but can be inferred from the evoked NPs, are treated as old NPs. They constitute 15.8% (72/456) of all of the NPs included in the database, or 20.2% (72/356) of the old NPs. Their distribution across the three forms, in comparison to other old NPs, is given in Table 10:

| | *ba* | % | preposed | % | postverbal | % | total | % |
|---|---|---|---|---|---|---|---|---|
| Inferable | 47 | 65.3% | 10 | 13.9% | 15 | 20.8% | 72 | 100% |
| Other old | 179 | 63.0% | 37 | 13.0% | 68 | 23.9% | 284 | 100% |

Table 10: Distribution of inferable NPs and other old NPs

Table 10 shows that inferable NPs pattern like old NPs; their distribution across the three forms is comparable to the distribution of other old NPs.

In contrast, their distribution is drastically different from how new NPs are distributed, as given in Table 11:

| | *ba* | % | preposed | % | postverbal | % | total | % |
|---|---|---|---|---|---|---|---|---|
| Inferable | 47 | 65.3% | 10 | 13.9% | 15 | 20.8% | 72 | 100% |
| New | 24 | 24.0% | 0 | 0.0% | 76 | 76.0% | 100 | 100% |

Table 11: Distribution of inferable NPs and new NPs

In the preposed form, while there are 10 inferable NPs, there are no new NPs at all; further, the relative distribution over the *ba* form and the postverbal form is reversed between inferable NPs and new NPs.

These comparisons support grouping inferable with old.

## 5. The *ba* form vs. the preposed form

We now turn to the second layer of the analysis and examine the two preverbal word orders. The issue is the following: if an object occurs preverbally, when is it more likely to occur with *ba*, and when is it more likely to be preposed? This question concerns variation between the two preverbal forms— the *ba* form and the preposed form, both exhibiting a non-canonical word order. In the preposed form, the object functions as a topic, at the initial position of a sentence, followed by the subject, which may or may not be expressed. In the *ba* form, the object is marked by *ba* and follows the subject, which also may or may not be expressed. Although the choice between the two forms also involves choice of word order, it is not the only thing that matters. The issue, it seems, has more to do with when an NP is marked by *ba*, and conversely, when an NP can be a topic. A related issue is whether the *ba* NP is itself a topic, as Tsao (1987) suggests. We will return to this issue in section 6.3.

First, however, it should be noted that not all *ba* sentences can be expressed in the preposed form. When the object is animate, sometimes the preposed form is unavailable. (14) is an example:

(14) a. Zhejuhua ba Lanxi renao
this-sentence BA Lanxi annoy
'These words annoyed Lanxi.'

b. *Lanxi zhejuhua renao
Lanxi this-sentence annoy
'These words annoyed Lanxi.'

The preposed form of (14a) is (14b), but it is not acceptable. If the subject is non-overt, then the preposed form may also be unavailable, as illustrated in (15):

(15) a. Ba wo ji -guoqu
BA me squeeze-over
'(He) squeezed me over.'

b. Wo ji -guoqu
I squeeze -over
'I squeezed over.'

In (15), (b) is not the preposed form of (a), since (b) does not have the same meaning as (a). In (15a), *wo* 'I' is the object, but in (15b), it can only be interpreted as the subject, not the object. Therefore, the preposed form is not available to (15a). Similarly, in (16), (a) and (b) have opposite meanings.

(16)a. Zuotian san bi yi ba laohuidui sheng le
yesterday three compare one BA tiger-team win PERF
'Yesterday three to one (we) beat the Tigers.'

b. Zuotian san bi yi laohuiduii sheng le
yesterday three compare one tiger-team win PERF
'Yesterday three to one the Tigers won.'

In (16a) the Tigers lost, but in (16b) they won. Thus the preposed form is also not available to (16a). On the other hand, (17) illustrates a *ba* sentence with an animate object that could be expressed in the preposed form:

(17)a. Bie ba haizi danwu le
don't BA child delay CRS
'Don't delay your child.'

b. Haizi bie danwu le
child don't delay CRS
'Don't delay your child.'

As for sentences in the preposed form, they could all be expressed in the *ba* form. Recall in 3.2 a criterion of data selection is that a sentence has the potential of being expressed in both the *ba* form and the postverbal form.

There are 250 *ba* sentences in the data, of which 57 contain animate objects, and 27 of them could not be expressed in the preposed form. The latter are therefore excluded from the second layer analysis. The data for this part of the analysis, then, includes 223 tokens in the *ba* form as well as 47 tokens in the preposed form.

As we will see below, a number of factors are involved in the variation between the *ba* form and the preposed form, each of which accounts for a portion of the data. In 5.1 I look at how new NPs distinguish between the two forms. In 5.2 I review three notions of topic: discourse topic, topic continuity (topicality) and topic-comment. In 5.3 and 5.4 I propose when the *ba* form and the preposed form are each likely to be used.

### 5.1 New and non-light

As we compare the *ba* NPs with the preposed NPs, we immediately observe one difference. While the *ba* NPs are mostly old, with a small percentage of new (10.8%), all of the preposed NPs are old. This suggests that between the two forms, if an object is new, the *ba* form will always be used. In fact, when this happens, the *ba* NP is invariably non-light, as Table 2 shows. (13) above is such an example.

As for the preposed NPs, the fact that they never carry new information suggests that preposed, topic NPs are subject to stricter constraints with respect to information status. Only NPs carrying old information can be topics. This finding is also consistent with Ward and Prince's (1991) analysis of topicalization in English. Ward and Prince propose that only constituents expressing discourse-old information can be topicalized in English. The constraint seems to hold in Chinese as well. The preposed NPs in my data include definite, specific indefinite and bare NPs with definite or generic interpretation, but they all carry old information. This difference between preposed NPs and *ba* NPs, however, only accounts for a small portion of the data: 24 of the 223 *ba* sentences are new. The majority of the data, 199 of the *ba* form and 47 of the preposed form, must be explained some other way.

### 5.2 Three notions of topic

I would like to suggest that the main factor for the variation between the *ba* form and the preposed form is topicality. In general, if the object is less topical than the subject, then the *ba* form is more likely to be used; elsewhere the preposed form is more likely to be used.

Before we continue, it will be useful to clarify the notion of topic. Vallduví (1992: 30-35) provides a concise review of three notions related to topic: discourse topic, topicality, and topic-comment. Discourse topic applies to a stretch of text or conversation; it refers to an entity or an event that is the topic of discussion. Structurally, a discourse topic need not be expressed, and if it is, there is no constraint on where it occurs in a sentence. Topicality, or topic continuity, on the other hand, is a property that all participants in a clause have. This notion was developed by Givón (1983). Participants exhibit high or low degrees of topicality, depending on recency of mention and persistence. Topicality has more to do with referent encoding and tracking, and not with information packaging. Finally, topic-comment is sentence-based, and has received the most attention in the literature. It is a relational notion: topic is what a sentence is about and comment is what is said about the topic. This is the sense in which I have been using the term 'topic' so far.

For the variation between the *ba* form and the preposed form, it is the first two notions, discourse topic and especially topicality, that will be more relevant. Givón (1983) offers three measurements of topicality (topic continuity): (a) referential distance, which measures the distance between the previous mention of an entity and its current mention, (b) potential interference, which measures the number of other elements in the directly preceding discourse that are compatible with the predicate, and (c) persistence, which measures how far an entity continues to be mentioned after its current mention. For our purposes, I will make use of (a) and (c) to measure relative topicality.

### 5.3 The *ba* form

In general, the *ba* form is used when the subject has higher topicality than the object, and this holds regardless of whether the object is a discourse topic or not. However, there are a few complicating factors that override this general tendency. I will illustrate the general trend first, and then consider one of the complicating factors— where *ba* is used even when the condition is not met. The other factors have to do with the reverse situation— where *ba* is not used even when the condition is met. They will be discussed in 5.4.

(18) is an example of the subject having higher topicality while the *ba* NP is a discourse topic:

(18) Zai xuangou dipi shi, ta kandao yige nongchang you yiqian
at choose-buy land time he saw one-CL farm has a-thousand

duozhu shishu. yinwei ta xiai na dapian
more-CL persimmon-tree because he loves that large-area

biejufengge de shishu, bian ba zhe shiyuan
with-special-character DE persimmon-tree then BA that persimmon-farm

mai-le xialai Ta jingxin sheji-le gaijian fangan, jinliang bu
buy-PERF DIR he do design-PERF remodel plan try-best not

fadiao yizhu shishu. Yiqie bantuo zhihou, ta jiang ziji de
cut one-CL persimmon-tree all done after he OBJ self DE

xin yusuo quming wei ‘bade yuan’
new residence name as ‘eight-virtue farm’

‘When he was looking for a lot, he saw a farm with more than a thousand persimmon trees. Because he loved the persimmon trees, which covered a wide area and were rather special, he purchased the persimmon farm. In his remodel design, he did his best to avoid cutting any trees. When all was done, he gave his new home the name “Eight Virtue Farm.”’

This passage is about a well-known Chinese artist buying a persimmon farm in South America. The *ba* NP is *zhe shiyuan* ‘the persimmon farm’, which is also the discourse topic. The subject of the sentence (beginning with *yinwei* ‘because’) is *ta* ‘he’, which is unrealized in the clause where *ba* occurs. The unrealized subject has high topicality, in terms of both referential distance and persistence. The referent is mentioned in the immediately preceding clause, and continues to be mentioned for five more clauses. In contrast, the object has low topicality— there is no persistence and the previous mention is three clauses away.

(19), from Pan (1996: 453-454), is an example of the subject having higher topicality while the *ba* NP is a non-discourse topic:

(19) Pingshi zuoye zhiyou liushi fenr, ta gei-le wo qishi fenr. Ta
regular homework only 60 points he gave me 70 points he

shangci ketangshang jiang le. Ta shuo: "Dajia xianzai dou dui,
last-time in-class say PERF he said everyone now all check

dui yixia ou. Ruguo ni liuci zuoye dou jiao le, ni
check a-little PRT If you six-times homework all turn-in PERF you

xianzai yinggai shi liushi fenr" Zuihou wo yikan wo zhe shangmian
now should be 60 points finally I looked I the top-side

zenme yijing qishi le. Wo jiu, wo jiu xialai wen ta.
how-come already 70 CRS I then I then go-down ask him

Ta shuo budui, yinggai shi liushi ya. Wo jiu ba <u>wo nage</u> gei
he said not-right should be 60 PRT I then BA my that-CL gave

ta. Ta qu cha-le bantian. "Oh", ta shuo, "wo duo gei-le
him he go check-PERF long oh he said I extra give-PERF

ni shi fen, keep it".
you 10 points keep it.

'There are 60 points for the assignments. He gave me 70 points. He said it in class last time. He said "Now everyone check, check a little. If you have handed in all six assignments, you should have 60 points." I took a look. How come (I) already have 70 points here? I then, I then went down to ask him. He said that's not right. It should be 60 points. I then gave <u>that thing of mine</u> to him. He went and checked for a long time. He said: "Oh, I gave you 10 points extra. Keep it."'

The passage is about the speaker receiving extra points in the assignments. It is not clear exactly what the *ba* NP *wo nage* 'that (thing) of mine' refers to. It could refer to the record kept by the speaker himself, or it could be the progress report that the instructor passed out to each student. Nonetheless, it is clear that wo *nage* 'that (thing) of mine' is what the speaker looked at, alluded to a few clauses back but not mentioned. The subject has higher topicality than the *ba* NP, because the subject *wo* 'I' is recently mentioned, while the object *wo nage* 'that (thing) of mine', although understood from context, is the only mention of the record/report. Such 'once-only' *ba* NPs are not uncommon. 29.3% (73/250) of all *ba* NPs are in this category.

We now consider a situation that overrides the tendency. It concerns cases where the condition of the subject having higher topicality is not met and yet the *ba* form is used. Seven tokens are in this category. (20) is an example:

(20) Qutang Xia menkou nakuai shitou wo kan cong yanse xiang
Qutang Gorge gate that-CL boulder I see from color like

tie yiyang, buran ni xiangxiang, ta bu keneng zai nar name
iron same otherwise you think-think it not possible at there so

duonian, yili zai nar. Wushi niandai shi ba <u>ta</u> zhadiao de.

many-years standing at there 50's period is BA it bomb DE

'The boulder at the gate of Qutang Gorge, it seems to me, is iron, based on its color. You think about it; otherwise, it couldn't have been there for so long, standing there. In the 50's, it was bombed.'

The *ba* NP *ta* 'it' refers to the boulder; it has higher topicality than the unexpressed subject, whose referent is mentioned only once in the passage. Thus (20) does not follow the general tendency. This may have to do with the fact that the *ba* NP is a pronoun. Six of the seven tokens include a pronominal *ba* NP. In contrast, none of the preposed NPs is a pronoun. This indicates that pronominal objects are unlikely to be topics; when the object is a pronoun, the *ba* form is used regardless of topicality.

Overall, the condition of the subject having higher topicality accounts for 192 out of 199 tokens in the *ba* form. Actually, subjects generally tend to have higher topicality than objects, as reported in the findings for various languages in Givón (1983). Therefore, the condition for the *ba* form simply follows this general tendency. In 5.4, however, we will see that this tendency is not followed in the preposed form. Another observation we can make concerns discourse topic status. Only 23.1% (46/199) of the *ba* NPs are discourse topics. Therefore, the *ba* NP as a non-discourse topic is the most common use of the *ba* form, accounting for 153 of the 199 tokens (76.9%) in the *ba* form.

### 5.4 The preposed form

As mentioned earlier, in the preposed form the object NP is at the initial position of the clause and acts as topic. Given what we said about the environment for *ba*, we would expect the preposed form to be used where the *ba* form is not, that is, when the object has a higher topicality than the subject. This is indeed one of the environments for the preposed form. But the preposed form is also used in three other environments, two of which are overriding factors.

First, (21) is an example where the object has higher topicality than the subject:

(21) Zhunbei kaoshi de feiyong qishi shi keyi jiesheng xialai de,
prepare test DE fees actually be can save DIR DE

fangfa jiushi—bu canjia hong taiyang yingyu peixun ban.
method is not attend Red Sun English training class

Dangran, zhebiqian wo shi laolaoshishide fu -le de.
Of course this-fee I be honestly pay-PERF DE

Dan wo shenbian ye you bushao ren shengxia-le yibufen,
but I body-side also there-are not-few people save -PERF some

biru qishier xiaoshi, jiu meiyou shang TOEFL peixun
for-example seventy-two hours then not attend TOEFL training

ban, dan kaode feichang bucuo
class but test-RES very not-bad

'The fees for preparing the test actually can be saved. The way to do it is not to sign up for the English classes offered by the Red Sun. Of course, this fee, I honestly

paid all of it. But there are also people who saved some of it. For example, "72hours" did not take the TOEFL class, but he did very well in the test.'

In this part, the writer offers ways of saving money when preparing for the TOEFL test. The object *zhebiqian* 'this fee' is the discourse topic. It has higher topicality than the subject *wo* 'I'; the latter exhibits low topicality here, as it is mentioned neither in the preceding context nor afterwards. As expected, the preposed form is used.

(22) is an example of the object as a non-discourse topic having higher topicality:

(22) W: ...Bazhi jiu zai zhekuai huagangyan shang. Wo jide
dam-location right at this-CL granite on I remember

shang- ci dao daba lai de shihou, tamen songgei wo
last- time to big-dam come DE when they give me

yikuai jinian de shitou jiushi yikuai bazhishang kaizao chulai
a-CL souvenir DE rock it-is one-CL dam-at dig out

de ba fenghua qudiao de, shijishang jiushi yikuai shizha.
DE BA erosion remove DE actually it-is one-CL rock-sediments

Danshi tamen baozhuang de henhao, meigeren nadao yihou, dou
but they package -DE well everyone receive after all

juede shi yige feichang hao de liwu.
fell is one-CL very good DE gift

G: Zhe yeshi yongyuan meiyou le, yihou <u>zha</u> qudiao
this also-is forever not-have CRS afterwards sediments remove

yihou, ba jiu zai dixia. zhe wannian daji
after dam right at bottom the ten-thousand-years big-foundation

jiu kao zhege.
then depend-on this-CL

'W: ...The location of the dam is right on this granite boulder. I remember last time when (I) came to the dam, they gave me a rock for souvenir, which was dug out at the dam, with the erosion removed. Actually (it) is a sedimentary rock. But they packaged it very well. Everyone who received it thought it was a very nice gift.

G: There won't be any more of that. In the future, after <u>the sediments</u> are removed, the dam will be at the bottom. It will be the foundation for thousands of years.'

The preposed object *zha* 'sediments' is not a discourse topic; rather, the discourse topic is the Three Gorges Dam. The word *shizha* 'sedimentary rock' serves as a previous mention of the preposed NP *zha* 'sediments'. There is no more mention of sediments in the rest of the conversation. The subject of the clause containing *zha* 'sediments' is unrealized and unimportant, and has no continuity at all. Therefore, (22) also follows the general tendency.

Higher topicality of the object accounts for 17 out of the 47 tokens of the preposed form (36.2%). The rest of cases fall into one of the three situations below.

### 5.4.1 Parallel structure

In the data, when two or more objects occur in clauses which are parallel in terms of structure and meaning, the preposed form is used regardless of relative topicality. (23) is an example of subject having higher topicality:

(23) Ailuete xie-guo yilianchuan Shaweng shidai juzuojia de pinglun,
Elliot write-EXP a-series Shakespeare period playwright DE criticism

wo ye genzhe du tamen. Pirushuo Ailuete xie -le pian
I also follow read them for-example Elliot write -PERF CL

Malu de duanlun, ershi fengzhong jike dubi, dan yao
Marlowe DE short-essay 20 minutes can read-finish but want

zhenzheng linglue qi jianjie zhi jingshen, ni ziji ye fei du
really appreciate its view DE depth you self also must read

Malu buke. Wo zai Shanghai qijian, Malu quanji zaoyi
Marlowe not-can I at Shanghai duration Marlowe collection already

du-le, Peng qiangsheng juben ye chabuduo du-le, Shaweng
read-PERF Ben Johnson plays also almost read-PERF Shakespeare

quanji ye du-le daban.
collection also read-PERF most

'Elliot wrote a series of essays on the playwrights during the Shakespeare period. I also read them. For example, Elliot wrote a short article on Marlowe. (I) finished it in twenty minutes, but to really appreciate the depth of his viewpoint, (you) have to read Marlowe yourself. When I was in Shanghai, the Marlowe collection (I) already read; Ben Johnson's plays (I) also basically read; the Shakespeare collection (I) also read more than half.'

The clauses that contain the preposed NPs *Malu quanji* 'the Marlowe collection' and *Peng Qiangshang juben* 'Ben Johnson's plays' and *Shashibiya quanji* 'the Shakespeare collection' are parallel in terms of structure, semantic content and discourse function. The subject *wo* 'I' has higher topicality than any of the three objects.

(24), from Pan (1996: 424), is another example of parallel structure, but in this case the subject has low topicality:

(24) A: Ta xianzai guanjian shi shenme ne, reading taiduo.
she now source is what PRT reading too-much

Zaiyige ta ting, ting ke ting bu dong.
another she listen listen class listen not understand

Ta shuo ta cha zidian shizai shi cha, cha bu guolai
she say she check dictionary really is check check not DIR

B: Mhm
mhm

C: Na shi kending de ma
that is certain DE PRT

A: Xiu bu xialai
take not down-DIR

C: You hao, youxie zi, you, youxie ci dou shi xie...
there-are many some words some some words all are some

A. Wo guji keneng jiushi youxie ci dagai gei ta shuo
I guess possible it-is some words generally to her explain

yixiar. Ta jiu buyong cha zidian Zaiyige ne,
a-little she then not-have-to check dictionary another-one PRT

meiyipianr dayi gei ta tong yixia
each summary to her straighten-out a-little

'A: Right now what is her main problem? There is too much reading. Another thing is that she cannot understand (the professor). She said she uses the dictionary (to the point) she cannot handle it.

B: Mhm.

C: That is for sure.

A: (She) cannot handle it.

C: There are many, some words, some, some words are all...

A: I guess possibly it is that for some of the words (you) explain to her a little. (This way) she won't have to use the dictionary. Another thing is for the summary of each article, (you) straighten her out a little.'

A graduate student was having trouble in an English literature class. In this part of the conversation, the three participants discuss the areas of difficulty faced by the student, which are also areas where she could use some help. This is represented by the two juxtaposed phrases: *youxie ci* 'some of the words', and *meiyipianr dayi* 'summary of each article'. It is clear that a parallelism exists between the two clauses, in terms of structure (both employing a PP and the quantified phrase *yixia* 'a bit'), semantic content (both about what the student needed), and discourse status (both are part of the discourse topic). Both phrases have limited topicality; while *youxie ci* 'some of the words' was mentioned in the immediately preceding sentence, *meiyipianr dayi* 'summary of each article' refers back to the reading, mentioned at the beginning. On the other hand, the unexpressed subject, referring to a potential tutor, is not mentioned either before or after the parallel structure, and therefore has no topicality.

Parallelism as a factor for preposing can also be seen in English. Birner and Ward (1998) show that topicalization of an adjective phrase requires a pair of clauses, not just a single clause. However, they also note (p. 47) that in their data, topicalization of other grammatical categories (including NP, VP and PP), mostly (79%) concerns single clauses only. In Chinese, parallelism seems to play a slightly more prominent role. In my data, it accounts for 14 of the 47 tokens (29.8%) of the preposed form. This includes both cases where the subject has higher topicality and cases where the object has higher topicality.

### 5.4.2 Unimportant participants

Another situation that results in the preposed form has to do with unimportant participants. If the participant expressed by the subject is unimportant, the preposed form is likely to be used even if the subject has higher topicality. (25) is an example:

(25) Dangshi yinwei zhua wo de shihou, wo yijing zhidao yao zhua
that-time because arrest me DE time I already know will arrest

wo le, Yinwei wo tongan xian zhua jinqu le, dangshi wo
me CRS because my partner first arrest in-DIR PERF that-time I

qi motuoche, jingche zai houmian genzhe, jingche
ride motorcycle police-car at behind follow-DUR police-car

yi tingxia, wo juede danao hong de yisheng.
as-soon-as stop I feel brain 'hong' DE one-sound

Wo zhidao wo zuoguo shenme shiqing.
I know I do-EXP what things

Zhe yao fu duoda de zeren.
this will pay how-big DE consequence

'At that time because when (they) arrested me, I already knew (they) were going to arrest me, because my partners were already arrested. At that time I was riding a motorcycle, the police car was following me. As soon as the police car stopped, I felt my head go "hong". I knew what I did. What consequences I would have to pay!'

In (30) the speaker describes how he was arrested for the crime he committed. The object *wo tongan* 'my partners' has lower topicality; its previous mention is many clauses before, but there is no mention of it afterwards. The unexpressed subject, referring to the police, has slightly higher topicality, as it was mentioned in the immediately preceding clause. Although it is clear that it is the police that arrested the speaker and his partners, the referent is not expressed and is not important in the discourse. The relative unimportance of the subject referent contributes to the use of the preposed form. Nine tokens (19.1%) are in this category.

### 5.4.3 Other cases

Besides parallel structure and unimportant subject participants, there are seven tokens that occur in contexts where the preposed form is used when one would expect the *ba* form, where the subject has higher topicality. One of them is given in (26):

(26) Wo wen: "daodi you shenme shi?" Ta shuo: ni xiang zou-le
I ask finally there-is what matter she say you want leave-CRS

shi ba? zheli you gui yao chi-le ni!" Wo buhaoyisi, zuoxialai
be PRT here there-is ghost will eat-PERF you I embarrassed sit-down

shuo: "shao dian shui pao bei cha lai chi, kou kesi le"
say boil some water make cup tea dir eat mouth thirsty-dead CRS

Ta qu shao-le shui lai shuo: "qishi ni keyi zai deng

she go boil-PERF water come say actually you can more wait

liangnian na-le gongminquan zai zou, luka bei langfeidiao
two-years get-PERF citizenship then leave greencard don't waste

le. you-le huzhao, laiqu jiu ziyou le shenmeshihou xiang
CRS have-PERF passport come-go then free CRS when want

lai jiu lai." Wo shuo: "hai deng liang nian? liangge yue dui
come then come I say still wait two year two-CL month to

wode yizhi dou shi yige kaoyan"
my will all be one-CL challenge

'I asked: "What is the matter?" She said: You want to leave, don't you? There is a ghost who wants to eat you?" I feel embarrassed. Sitting down, I said: "Could you boil some water and make some tea? I'm thirsty." She went to make water, and as she returned, she said: "Actually you can wait for two more years and get your citizenship, then you can go. The green card, don't waste it. Once you have the passport, you can come and go freely. Anytime you want to come you can come." I said: "Wait for two more years? Two months is already a challenge to me."'

In this passage the speaker's friend tries to persuade him to stay and get Canadian citizenship before returning to China. The preposed NP *luka* 'green card' has limited topicality, being mentioned again 14 clauses later. The unexpressed subject, referring to the speaker, clearly has higher topicality than the object. According to my analysis, it should be compatible with the *ba* form, but the preposed form is used instead. This might have something to do with *luka* 'green card' being part of the discourse topic; in the conversation that follows, the friend continues to talk about citizenship, passport and green card. Five of the seven tokens that don't conform to my analysis contain discourse topics. In fact, most of the preposed NPs (68%, 32/47) are discourse topics. However, being a discourse topic does not necessarily mean the preposed form will be used. Recall earlier in 5.3 it was observed that in the *ba* form 46 of the 199 *ba* NPs are also discourse topics. So it is not the case that if the object is a discourse topic, the preposed form is likely to be used, although the reverse does hold. I will take examples like (26) as cases of variation that fall outside of any of the factors proposed.

In general, then, the preposed form is used under three environments: lower topicality of the subject, parallel structure, and unimportant subject participants. In cases where topicality matters, the tendency of subjects having higher topicality is not followed; rather, it is the object that has higher topicality.

## 5.5 Summary

In this section we have considered the factors that affect the choice between the two preverbal forms. First, between the two forms, if the object NP is new, then the *ba* form will be used. Elsewhere, topicality is a major, but not the only, factor. The *ba* form is more likely to be used when the subject has higher topicality than the object. As for the preposed form, it is mainly used when the object has higher topicality, in a parallel structure or when the subject is unimportant. Most of the *ba* NPs are non-discourse topics, while most of the preposed NPs are discourse topics.

## 6. Discussion

In this section we will consider three issues that arise from my findings. In 6.1 I take a closer look at the factors for the usage of *ba*; in 6.2 I compare my findings to principles of word order that have been proposed in the literature; and in 6.3, I consider whether the *ba* sentences are related to the notion of topic.

### 6.1 Factors for the usage of *ba*

My findings suggest that the use of the *ba* form depends on multiple factors. When all is considered, the *ba* form is more likely to be chosen under two situations: (a) when the object carries old information and is less topical than the subject, (b) when the object is new & not light. In the first situation, information status and topicality are relevant; in the second situation, what matters is information status and weight.

What do these findings tell us about *ba* sentences' functions in discourse? They suggest that *ba* sentences do not perform a unified function in discourse. On the one hand, *ba* sentences are used as a device to express information that is old and yet is not highly topical; on the other hand, *ba* sentences are also a device that allows speakers to place heavy material before the verb when it is new information. My findings also indicate what the *ba* NP is not intended to do— it is not intended to mark the discourse topic, as most of the *ba* NPs are not discourse topics.

Sun and Givón (1985) suggest that the OV construction, whether it is marked by *ba* or not (including the *ba* form and the preposed form) is a marked, contrastive and emphatic device. They arrived at this conclusion by comparing two measures between the OV categories and VO categories: referential distance and potential interference. In comparison with the VO order, the OV order has low referential distance value and high potential interference value. My study did not look at these measures for the VO and OV variation. Rather, we showed that the variation between VO and OV has to do with old vs. new, weight, and the interaction between old vs. new and weight. It should be noted, however, that the two studies are not really comparable. While Sun and Givón considered all OV sentences in their database, I have considered only sentences that can potentially vary between the *ba* form and the postverbal form. Most of the OV examples cited by Sun and Givón don't meet this criterion. Further, while my study is concerned with when the *ba* form (or the postverbal and preverbal forms) is more likely to occur, Sun and Givón are concerned with distribution of VO and OV sentences in terms of types of NPs.

### 6.2 Principles of word order

According to Gundel (1988), at least two different pragmatic principles have been proposed in the literature concerning word order: (a) Old before new, (b) Most important information first. Gundel points out that the two principles are sometimes in conflict. It is often the case new information is the most important information. How do my findings bear on these principles? My data clearly supports the first principle, as discussed in section 4. As for the second principle, since this study is not concerned with important information, the data neither supports nor challenges the principle.

As for weight, there is evidence that the placement of heavy material varies cross-linguistically. English is known to position heavy material toward the end of a sentence, e.g. Heavy NP Shift. On the other hand, Hawkins (1994) shows that in Japanese and Korean heavy material tends to be placed earlier in a sentence. The Chinese data paints a mixed picture. Table 7 shows that NPs tend to be preverbal, regardless of weight. The situation becomes much clearer, however, when weight is considered together with information status. As Figure 1 shows, while

light & old NPs tend to be preverbal, light & new NPs are always postverbal; similarly, heavy & old NPs are mostly postverbal, while heavy & new NPs are mostly preverbal. Thus although weight alone is not sufficient to predict where NPs will occur in *ba* contexts, when weight combines with old/new, it is a rather strong predictor— light & old and heavy & new tend to go before light & new and heavy & old.

### 6.3 The *ba* NP as a topic

In the literature the *ba* NP has been associated with the notion of topic one way or another. Givón (1978:313) considers the *ba* sentences as devices for marked topicalization, while Tsao (1987) proposes that it is a secondary topic. Is the *ba* NP indeed a topic? In what sense is it a topic? To answer these questions we will briefly review Tsao's proposal.

Tsao (1987) proposes that *ba* NP is a secondary topic, which together with the comment that follows, makes a comment about the primary topic at the initial position of a sentence. By applying the same criteria that are used to identify the primary topic to the *ba* NP, Tsao says that the *ba* NP has most of the properties of a topic, given below:

a) The *ba* NP invariably occupies the S-initial position of the first S in a *ba* topic chain.[2]

b) The *ba* NP is most often definite or generic but can be specific, especially when the regular topic is in the first person

c) The *ba* NP has some discourse properties as well. It can extend its semantic domain to more than one S.

d) The *ba* NP is in control of all the pronominalization and coreferential NP deletion processes in a *ba* topic chain.

Of the four criteria, two are sentence-based, and two discourse-related. If the criteria for the *ba* NP as topic are mainly sentence-based, then there isn't much need to look at its behavior in discourse— the *ba* NP already possesses the sentence properties regardless of where it occurs in discourse. However, it will be useful to see if the *ba* NP also has a role to play in discourse with respect to the two discourse-related criteria, that the *ba* NP's semantic domain often extends beyond its own clause and that it can head a topic chain. Both have to do with the topicality of the *ba* NP. The supporting evidence provided by Tsao, however, is based on constructed examples. It is therefore appropriate to reevaluate the proposal against natural data. Earlier in 5.3, we saw that the *ba* NP is generally less topical than the subject; here we would like to find out to what degree the *ba* NP's domain extends beyond its own clause, and how often it forms a topic chain.

To find out the answers, I did two counts. First, I considered a *ba* NP's domain as more than one clause if the same referent occurs anywhere in the immediate contexts— in a stretch of ten clauses, five preceding and five following, regardless of the linguistic devices (including zero-anaphora) used to mark the referent. Second, I counted the occurrences of a *ba* NP occurring in a chain with three or more links, regardless of whether it heads the chain. It turns out just about half of the *ba* NPs (51.2%, 128/250) extend their domains to other clauses within the ten-clause range, and 11 of them (4.4%) form a topic chain of three clauses or more. This tells us that the *ba* NP does play a role in discourse in that its semantic domain extends beyond its own clause half of the time. However, it also suggests that having a domain wider than its clause or forming a topic chain is not a central characteristic of the *ba* NP, as the other half of the

[2] This statement is obviously inaccurate, since the *ba* NP does not normally occupy the S-initial position. A better statement would be: The *ba* NP occupies a preverbal position.

times the referent of the *ba* NP is mentioned only once in a ten-clause range. Another piece of evidence that also suggests that the *ba* NP does not have high topicality is that, as mentioned in 5.3, 29.2% of the *ba* NPs (73/250) are mentioned only once anywhere in the data. Such high number of once-only occurrences of the *ba* NPs casts doubt on the *ba* NP's clause-linking function.

Returning to the earlier question, is the *ba* NP a topic? The answer depends on how topic is defined— whether it is sentence-based and whether continuity in discourse is an essential characteristic of topic. Chu (1993, 1998), for example, takes clause linking as a primary attribute of topic. However, based on my data, the only sense in which the *ba* NP could be considered a topic is at the sentence-level, which is probably what Tsao intended in the first place. As for its behavior in discourse, all we can say is that the *ba* NP exhibits limited continuity in discourse.

## 7. Conclusion

In this study, I have examined how *ba* sentences are used in discourse. In particular, I considered when the *ba* form is likely to be used, as opposed to the postverbal form and the preposed form. My study suggests that the choice of the *ba* form depends on multiple factors, including information status, weight and topicality. The *ba* form is more likely to be chosen under two situations: (a) when the *ba* NP carries old information but is not highly topical, and (b) when the *ba* NP carries new information and is heavy.

A rather striking finding of this study is that while the *ba* NP mostly carries old information, it does not play a significant role in discourse, as it exhibits only limited topicality in discourse and most of the *ba* NPs are not discourse topics. Therefore, *ba* sentences are not devices of clause linkage, neither are they devices of expressing topics in discourse.

Two rules were discovered from the data. First, if the *ba* NP carries new information, it is not light; the *ba* NP is never new and light. Second, the preposed form is used only when the object carries old information; preposing the object to the sentence-initial position always concerns old NPs.

It is also interesting to note that the variation between preverbal and postverbal word order is affected by both information status and weight. In particular, the interaction between information status and weight is significant, suggesting that both are important factors of word order variation in Chinese.

## ABBREVIATIONS

| | |
|---|---|
| BA | *ba* |
| CL | classifier |
| CRS | currently relevant state marker |
| DE | *de*, possessive, relative, modifier marker |
| DIR | directional complement marker |
| DUR | durative marker |
| EXP | experiential aspect |
| OBJ | object marker |
| PERF | perfective aspect |
| PROG | progressive aspect |
| PRT | sentence final particle |
| RES | resultative marker |

REFERENCES


Arnold, Jennifer; Wasow, Thomas; Losongco, Anthony; and Ginstrom Ryan. 2000. Heaviness and newness: The effects of complexity and information structure on constituent ordering. *Language* 76: 28-55.

Birner, Betty J.; and Ward, Gregory. 1998. *Information status and noncanonical word order in English*. Amsterdam: John Benjamins.

Chafe, Wallace. 1976. Giveness, contrastiveness, definiteness, subject, topic and point of view. *Subject and topic*, Li, Charles N. (ed.), 25-55. New York: Academic Press.

Chao, Yuen-ren. 1968. *A grammar of spoken Chinese*. Berkeley: University of California Press.

Chu, Chauncey, C. 1993. The prototypicality of topic in Mandarin Chinese. *Journal of Chinese Language Teachers Association* 28: 25-48.

Chu, Chauncey, C. 1998. *A discourse grammar of Mandarin Chinese*. New York: Peter Lang.

Fienberg, Stephen, E. 1978. *The analysis of cross-classified categorical data*. Cambridge: MIT Press.

Givón, Talmy. 1978. Definiteness and referentiality. *Universals of human language*, Vol. 4: *Syntax*, Greenberg, Joseph. H.(ed.), 291-330. Stanford: Stanford University Press.

Givón, Talmy (ed.). 1983. *Topic continuity in discourse: A quantitative cross-language study*. Amsterdam: John Benjamins.

Gundel, Jeanette. 1988. Universals of topic-comment structure. *Studies in syntactic typology*, Hammond, Michael; Moravcsik, Edith; and Wirth, Jessica (eds), 209-239. Amsterdam: John Benjamins.

Hawkins, John A. 1994. *A performance theory of order and constituency*. Cambridge: Cambridge University Press.

Ho, Yong. 1993. *Aspects of discourse structure in Mandarin Chinese*. Lewiston: Mellen University Press.

Hopper, Paul J.; and Thompson, Sandra A. 1980. Transitivity in grammar and discourse. *Language* 56: 251-299.

Hsueh, Feng-sheng. 1987. The *ba* construction in Mandarin: An integrated interpretation." *Language Teaching and Linguistic Studies* 1: 4-22.

Koopman, Hilda. 1984. *The syntax of verbs*. Dordrecht: Foris.

Li, Audrey Y.-H. 1985. *Abstract case in Chinese*. Ph.D. Dissertation. University of Southern California.

Li, Charles, N.; and Thompson, Sandra A. 1974. Historical change of word order: A case study in Chinese and its implications. *Historical linguistics*, Anderson, John; and Jones, Charles (Eds.), 199-217. Amsterdam: North-Holland.

Li, Charles, N.; and Thompson, Sandra A. 1975. The semantics function of word order: A case study in Mandarin. *Word order and word order change*, Li, Charles (Ed.), 163-195. Austin: University of Texas Press.

Li, Charles, N.; and Thompson, Sandra A. 1981. *Mandarin Chinese: A functional reference grammar*. Berkeley: University of California Press.

Liu, Feng-hsi. 1997. An aspectual analysis of *ba*. *Journal of East Asian Linguistics* 6: 51-99.

Liu, Feng-hsi. 1999. Transitivity and Structure Preservation. *Functionalism and formalism in linguistics II, Case studies.* Darnell, Michael; Moravscik, Edith; Noonan, Michael; Newmeyer, Frederick and Wheatly, Kathleen (eds.), 175-202. John Benjamins Publishers.

Lü, Shu-xiang. 1984. Ba zi yong fa de yanjiu (Study of *ba* sentences). *Hanyu yufa lunwen ji* (Studies of Chinese syntax). 176-199. Beijing: Shangwu Yinshuguan

Pan, Junlin. 1996. *Occurrence/non-occurrence, distribution, and interpretation of zero-anaphora in Chinese conversational discourse*. Ph.D. Dissertation, University of Arizona.

Prince, Ellen. 1981. Toward a taxonomy of given/new information. *Radical pragmatics*, Cole, Peter (Ed.), 223-254. New York: Academic Press.

Prince, Ellen. 1992. The ZPG letter: Subjects, definiteness and information status. *Discourse description: Diverse analyses of a fundraising text*, Thompson, Sandra A.; and William Mann (Eds.), 295-325. Amsterdam: John Benjamins.

Quirk, Randolph; Greenbaum, Sidney; Leech, Geoffrey; and Svartvik; Jan. 1972. *A grammar of contemporary English*. London: Longman.

Sun, Chaofen. 1995. Transitivity, the *ba* construction and its history. *Journal of Chinese Linguistics* 23: 159-195.

Sun, Chaofen; and Givón, Talmy. 1985. On the so-called SOV word order in Mandarin Chinese: A quantified text study and its implication. *Language* 61: 329-351.

Sybesma, Rint. 1992. *Causatives and accomplishments: The case of Chinese ba*. Ph.D. Dissertation. Leiden University.

Thompson, Sandra A. 1973. Transitivity and the *ba* construction in Mandarin Chinese. *Journal of Chinese Linguistics* 1:208-221.

Travis, Lisa. 1984. *Parameters and effects of word order variation*. Ph.D. Dissertation. MIT.

Tsao, Feng-fu. 1987. A topic-comment approach to the *ba* construction. *Journal of Chinese Linguistics* 15: 1-54.

Vallduví, Enric. 1992. *The informational component*. New York: Garland Publishing.

Ward, Gregory L.; and Prince, Ellen F. 1991. On the topicalization of indefinite NPs. *Journal of Pragmatics* 16:167-177.

Wasow, Thomas. 1997. Remarks on grammatical weight. *Language Variation and Change* 9: 81-105.

Wasow, Thomas. 2002. *Postverbal behavior*. Stanford: Center for the Study of Language and Information.

Yong, Shin. 1993. *The aspectual phenomena of the ba construction*. Ph.D. Dissertation, University of Wisconsin, Madison.