

PHRASE STRUCTURE, LEXICAL INTEGRITY, AND CHINESE COMPOUNDS*

C.T. James Huang
National Tsing Hua University

1. Introduction

The purpose of this paper is to discuss two independent formal principles that figure prominently in Chinese grammar, and consider their implications for a proper analysis of certain types of verbal compounds in the language. The first principle requires that the internal structure of a sentence or verb phrase in Chinese be 'largely head-final but only trivially head-initial' in the sense to be made precise below. The second is a universal principle underlying much work in linguistic theory, which requires that the internal structure of a word or a lexical category be inaccessible to rules of syntax. In sections 2 and 3 below, these two principles, hereinafter to be referred to as the 'Phrase Structure Condition' and the 'Lexical Integrity Hypothesis' (the PSC and the LIH) respectively, will be shown to be relevant in the language and to be of important explanatory value for an adequate description of its grammar. In section 4, we will show that the two principles appear to run into certain difficulties when one considers sequences that have traditionally been identified as compounds, in particular 'resultative compounds' and 'verb object compounds,' in that they appear to be systematically contradicted by these sequences in certain environments. Section 5 discusses possible solutions to the paradox, and indicates their theoretical implications for a proper analysis of such sequences.

In the course of the discussion that follows, it will be seen that many of the facts are fairly well known in the literature. Our interest in this paper is mainly a theoretical one, however, and we will be concerned with trying to identify general principles that express significant generalizations underlying various observed facts, and with

deriving theoretical conclusions indicated at the end of the paper.

2. The Phrase Structure Condition

For the purposes of this paper, the PSC will be stated in the following form:

1) The PSC:

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

Put in a somewhat oversimplified way, this condition may be understood as saying that a verb in Chinese may be followed by at most one constituent, though it may be preceded by an indefinite number of constituents (including subject and adverbial modifiers). The verb of a sentence must occur near the end of the sentence, and only followed by at most one constituent, in the simplest cases. Therefore, the internal structure of a sentence or predicate phrase is largely head-final, and only trivially head-initial. Now I will try to motivate the PSC by showing that there is a clustering of (fairly well-known) facts whose joint existence may be most profitably made to follow from the postulation of some such condition as the PSC.

First of all, we know that ba constructions and bei constructions normally exist as optional variants of non-ba and non-bei sentences. This is shown by the fact that all of the sentences in (2) are well-formed.

- 2a) Tā piàn-le wǒ le.
he cheat-LE I LE
'He cheated me.'
- b) Tā bá wǒ piàn-le.
he BA I cheat-LE
'He cheated me.'

- c) Wǒ bèi tā piàn-le.
I by he cheat-LE
'I was cheated by him.'

However, in one type of construction (specifically what Lü Shuxiang called the 'retained object construction') the ba and bei constructions are obligatory: there is no well-formed non-ba or non-bei counterpart, as is shown in (3).

- 3a) *Wǒ [_v [_v buō-le pí] júzi].
I peel-LE skin orange
'I peeled the orange.'
- b) [Wǒ [bá júzi [buō-le pí]]].
I BA orange peel-LE skin
'I peeled the orange.'
- c) [Júzi [bèi wǒ [buō-le pí]]].
orange by I peel-LE skin
'The orange was peeled by me.'

If we consider the semantics of the ba sentence (3b) and the bei sentence (3c), it is reasonable to set up the form (3a) as their base (in traditional transformational terms--cf. Thompson 1973a). The NP júzi 'orange' would be the 'outer object' of the small verb phrase buō-le pí 'remove the skin (of),' which consists of the verb and the object 'skin,' the 'inner object.' This is on a par with the English example indicated in (4) below, where 'Bill' is the outer object of the little VP 'take advantage' meaning 'exploit,' and 'advantage' is the inner object of the verb 'take.'

- 4) John [_v [_v took advantage] Bill].³

The question relevant to our discussion is why (3a) in Chinese is not well-formed. Note that (3a) has exactly the structure that violates our condition PSC. As far as the linear structure is concerned, the verb buō is followed by two constituents, the inner object and the outer object. For a more precise picture about this sentence in relation to the PSC, consider its hierarchical structure. The small

VP buō-le pí is the head of the larger VP containing the outer object, and it branches to the left as the left daughter of the latter. Within the small VP the head verb further branches to the left, as the left sister of the inner object. This configuration is in direct violation of the PSC. The situation with the ba sentence (3b) and the bei sentence (3c) is different, however, since the verb is followed by only one constituent. The small VP occurs on the right branch of its parent node. The structures of (3b) and (3c) are in accordance with the PSC. The obligatoriness of the ba and bei construction in 'retained-object' sentences can thus be seen as a consequence of this condition.

Secondly, note that what we may call 'subject-verb inversion' may occur only with intransitive verbs and only when the verbs are not followed by any adverbial material. An example of legitimate inversion is given in (5).

- 5a) Yǔ xià-le.
rain fall-LE
'It has rained. Rain has fallen.'
- b) Xià-le yǔ le.
fall-LE rain LE
'It has rained.'

The verb xià 'fall' in (5a) is intransitive and is followed by nothing (ignoring le) prior to inversion. The inverted form (5b) is thus well formed. Consider now the sentences in (6).

- 6a) [Yǔ [xià de hen dà]].
rain fall COMP very big
'It rained very heavily.'
- b) *[[Xià yǔ] de hen dà].
fall rain COMP very big
- c) *[[Xià de hen dà] yǔ].
fall COMP very big rain

In (6a), the same verb 'fall' is followed by an adverbial modifier 'very heavily.' The result of inversion, either in the form of (6b) or in the form of (6c), is in violation of the PSC, because the verb is followed by an NP and an adverbial modifier. The fact that inversion is disallowed in this case can be seen as another consequence of the PSC.

Thirdly, consider the following sentences, in which a verb is followed by a resultative adverbial [as in (7a)], a descriptive adverbial [as in (7b)], a duration phrase [as in (7c)], or a frequency expression [as in (7d)].

- 7a) Tā pǎo de hen lèi.
he run COMP very tired
'He ran until he got very tired.'
- b) Tā pǎo de hen kuài.
he run COMP very fast
'He ran very fast.'
- c) Tā pǎo-le liǎngge zhōngtōu.
he run-LE two hour
'He ran for two hours.'
- d) Tā pǎo-le liǎng cǐ.
he run-LE two time
'He ran twice.'

Note that in each of these sentences, the verb is intransitive, not followed by an object in addition to the adverbial. If the verb is transitive and is followed by its own object and such an adverbial, the sentence is ill-formed, as shown in (8) in clear contrast to (7).

- 8a) *Tā qí mǎ de hen lèi.
he ride horse COMP very tired
'He rode a horse until he got very tired.'
- b) *Tā xiě xìn de hen kuài.
he write letter COMP very fast
'He wrote letters very fast.'

- c) *Tā kàn shū le liángge zhōngtòu.
he read book LE two hour
'He read books for two hours.'
- d) *Tā xué Yīngwén le liǎng cì.
he learn English LE two time
'He learned English twice.'

Note again that the ill-formedness of each of (8) can be attributed to the PSC, since the verb is in each case followed by more than one constituent.

Fourthly, a closely related fact is that there exists a process of verb copying whose existence may be plausibly taken as a function to save the ungrammatical strings in (8) from the PSC. Compare (8) with the sentences in (9), in which the verb has been reduplicated between the object NP and the adverbial material.

- 9a) Tā qí mǎ qí de hen lèi.
he ride horse ride COMP very tired
'He rode a horse until he got tired.'
- b) Tā xiě xìn xiě de hen kuài.
he write letter write COMP very fast
'He wrote letters very fast.'
- c) Tā kàn shū kàn le liángge zhōngtòu.
he read book read LE two hour
'He read books for two hours.'
- d) Tā xué Yīngwén xué le liǎng cì.
he learn English learn LE two times
'He learned English twice.'

Notice that the result of this reduplication in each case is a structure in which a verb is followed by exactly one constituent at a time: the original verb is followed by its object, and the reduplicated verb is followed by the adverbial material. This is exactly in accordance with the PSC, and the sentences are all well-formed.

Fifthly, the sentences in (8) may also be saved by preposing one of the two constituents following the verb in each case. For example, in (10), the object is topicalized, so that the verb is followed only by the adverbial material. Such sentences also satisfy the PSC, and no reduplication is necessary.

- 10a) Nèipi mǎ, tā qí de hen lèi.
that horse he ride COMP very tired
'That horse, he rode until he got very tired.'
- b) Nēifeng xìn, tā xiě de hen kuài.
that letter he write COMP very fast
'That letter, he wrote very fast.'
- c) Nèiben shū, tā kàn le liángge zhōngtòu.
that book he read LE two hour
'That book, he read for two hours.'
- d) Yīngwén, tā xué le liǎng cì.
English he learn LE two time
'English, he learned twice.'

Other examples may be constructed to show the same point. The reader may see for himself that the sentences in (8) may be saved by passivizing the object in each case, by turning the object into a preverbal *ba*-object, or otherwise removing the object from its postverbal position.

Finally, still another way to save the ill-formed sentences in (8) is to somehow turn the two constituents following the verb into one single constituent. For example, consider (8c) and (8d), each with the verb followed by two constituents. One can prepose the adverbial expression 'two hours' or 'two times' and make it a part of the preceding object NP, forming the sentences (11a) and (11b), respectively.

- 11a) Tā kàn le liángge zhōngtòu de shū.
he read LE two hour DE book
'He read books for two hours.'

- b) Tā xué le liǎng cì (de) Yīngwén.
 he learn LE two time DE English
 'He learned English twice.'

The result of this maneuver is a surface construction like 'two hours books,' 'two times English,' with the duration and frequency expressions used analogously as the numeral-classifier phrases of the object NPs. The semantics of such sequences as NPs does not make much sense, but the existence of them each as a syntactic constituent may be seen as a consequence of the purely structural condition, the PSC.

In short, the six facts just mentioned seem to provide strong evidence for a principle along the lines of the PSC.⁵ We turn now to the second formal principle mentioned above.

3. The Lexical Integrity Hypothesis

It is an often observed fact that natural languages typically exhibit several 'autonomous' levels of structure, autonomous in the sense that the internal structure of one level is often inaccessible to the rules that apply at another level, or that rules applying at one level are often 'blind' to the internal structure of another level. The locution 'lexical integrity' refers to the hypothesis that information regarding the internal structure of words (or lexical categories) is often inaccessible to rules that apply in syntax to phrases (or phrasal categories). This hypothesis, which is a fundamental assumption underlying much work in current linguistic theory, is given in (12), following Jackendoff (1972), Selkirk (forthcoming).

- 12) The LIH
 No phrase-level rule may affect a proper subpart of a word.

The LIH may be motivated in a fairly easy way. For example, conjunction reduction may not apply to delete a proper subpart of a word. This is illustrated by the Chinese and English examples below.⁶

- 13a) [huǒ-chē] gen [qì-chē]
 fire-car and gas-car
 'train and automobile'

- b) *[huǒ gen qì] che
 fire and gas car

- 14a) New York and New Orleans

- b) *New [York and Orleans]

Secondly, semantic interpretation rules may not see the proper subpart of a word. This is why the phrase (15) is not taken to be anomalous.

- 15) yīkuài lǜsè de hēi-bǎn
 one green DE black-board
 'a green blackboard'

Thirdly, anaphoric rules may not apply to refer to a subpart of a word. Therefore, we have the examples given in (16) from Postal (1969):

- 16a) Chomsky likes Chomsky-ites/*him-ites.

- b) All Bloomfieldians like Bloomfield/*him.

These facts are, I think, sufficient motivation for the postulation of the LIH as a statement of a significant generalization about natural language. In the section that follows, we will look at some facts concerning Chinese compounds, especially as they pertain to the two principles just discussed.

4. Chinese Compounds

According to traditional treatments (as represented by Chao 1968, Lu 1964 and many other works), there are five types of so-called compounds in Chinese. These are classified based upon the 'internal syntax' of the

compounds as a) coordinative, b) subject-predicate, c) subordinative, d) resultative, and e) verb-object compounds. Each of these five types is briefly illustrated below for convenience.

17) Coordinative Compounds

- | | | | |
|----|------------|------------|----------|
| a) | chǐ-cùn | foot-inch | 'size' |
| b) | kāi-guān | open-close | 'switch' |
| c) | cháng-duǎn | long-short | 'length' |

18) Subject-Predicate Compounds

- | | | | |
|----|-----------|----------------|---------------|
| a) | tóu-téng | head-ache | 'headache' |
| b) | xīn-huāng | heart-worried | 'worried' |
| c) | ròu-má | flesh-creeping | 'sentimental' |

19) Subordinative Compounds

- | | | | |
|----|------------|--------------|--------------|
| a) | xiān-shēng | first-born | 'mister' |
| b) | tóng-qíng | same-feeling | 'sympathize' |
| c) | hēi-bǎn | black-board | 'blackboard' |

20) Resultative Compounds

- | | | | |
|----|---------|-------------|-------------------------|
| a) | dǎ-bài | hit-lose | 'defeat' |
| b) | tī-sǐ | kick-dead | 'kick to death' |
| c) | chī-bǎo | eat-full | 'get full by eating' |
| d) | hē-zuì | drink-drunk | 'get drunk by drinking' |

21) Verb-object Compounds

- | | | | |
|----|----------|-------------|----------------------|
| a) | bǎng-tuǐ | tie-leg | 'legging' |
| b) | dān-xīn | carry-heart | 'worry at' |
| c) | fù-zé | carry-duty | 'be responsible for' |

What are the criteria that have enabled or guided Chinese grammarians in their identification of these examples as compounds or words, and not as phrases? According to Chao (1968), an item is identified as a compound if one or more of the criteria in (22) are met.

- 22) a. Part of the item is a bound form.
 b. Part of it is neutral-toned.
 c. The meaning of the whole is not compositional of its parts.
 d. The internal structure is exocentric.
 e. The parts are inseparable from each other.

Let us consider first (22c), the criterion of semantic noncompositionality, or idiomaticity. Although Chao has used this as a possible criterion, it should be noted that it cannot be considered a sufficient condition for compoundhood, and must be excluded as a non-genuine criterion. This is because it is generally accepted that there may be idiom phrases, namely that idioms need not always be words. For example, the expression 'kick the bucket' is generally regarded as a phrase. The figurative expression guà yáng-tóu, mài gǒu-ròu 'hang a goat's head, sell a dog's meat' is also idiomatic, but it is in no sense a word or compound.

Excluding idiomaticity from further consideration, note next that the other four criteria listed in (22), in so far as they may be used as valid criteria, actually may be reduced to one single criterion, i.e., seen as special cases of the LIH just discussed. First, consider (22a), (22b), and (22e). Neutral-toned elements are usually bound forms, so (22b) may be reduced to (22e), the criterion of inseparability. Finally, note that this criterion is just a special case of the criterion of lexical integrity. Now, let us consider the criterion of exocentricity, indicated in (22d). An example that meets this criterion is the verb-object compound indicated in (21a), bǎng-tuǐ, used as a noun meaning 'legging.' This is exocentric in that a verb-object combination makes up a noun, in violation of the general principles of phrase structure, which require that well-formed phrase structures must be endocentric. Now, by calling the string bǎng-tuǐ a compound, we make it a word, and make its exocentric internal structure inaccessible to the endocentricity principle, which is a principle of phrasal, not lexical, structure. This is making crucial use of the notion of lexical integrity. In

short, the LIH is the single most important hypothesis underlying much work on Chinese compounds.

Although the LIH has proved to be successful in identifying a large body of Chinese compounds, there are two kinds of systematic exceptions to this criterion. The first is represented by the fact that most so-called verb-object compounds are separable in certain positions, a fact fairly well known since Chao (1968). For example, consider the sentences in (23):

- 23a) Tā hen dān-xīn zhèjian shì.
he very worry this matter
'He is very worried about this matter.'
- b) Tā dān shénme xīn?
he carry what heart
'He was worried about what?'
(Lit. 'He carried which heart?' or
'He wor- about which -ry?')
- c) Tā dān-le sān nián de xīn.
he carry-LE three year DE heart
'He worried for three years.'
(Lit. 'He wor- ed three years -ry.'
- d) Xīn, wǒ yīdiǎn dou bù dān.
heart I a-bit all not carry
'Worried, I am not at all.'
(Lit. '-Ry, I do not wor- at all.')
- e) Xīn, wǒ xiǎng tā shì huì dān de.
heart I think he be will carry DE
'Worry, I think he will.'
(Lit. '-Ry, I think he will wor-.'

In (23a) the two syllables in dān-xīn 'worry' are not separated, which is in accordance with the LIH. Furthermore, since dān-xīn is followed by an object 'this matter,' it must be regarded as a compound, not a phrase. If dān-xīn were a phrase, then the verb would be followed by an inner object xīn 'heart' and an outer object 'this matter.'

This would violate the PSC which we motivated earlier, and would be excluded as ill-formed on a par with strings like (3a). By calling it a compound, we render the internal structure of dān-xīn inaccessible at the syntactic level, and the sentence is well-formed. Therefore, by both the LIH and the PSC, dān-xīn is a word in (23a).

Consider now (23b, c, d, e). In each of these examples, xīn has been separated from dān. In (23e), in fact, it has been separated in an unbounded fashion by long-distance topicalization. By the LIH, dān-xīn would then be a phrase in each of these latter examples. Furthermore, since in each of these sentences dān is not followed by an outer object like 'this matter' of (23a), the PSC does not force one to call it a compound. Therefore, by both the LIH and the PSC, dān-xīn is more appropriately regarded as a phrase, in particular an idiom phrase, given its semantic non-compositionality. We have a paradox here, with dān-xīn being a word and a phrase.

A similar example is given in (24):

- 24a) Wǒ chángchang yōumò tā.
I often tease he
'I often teased him.'
- b) Wǒ yōu-le tā yi mò.
I hu-LE he one -mour
'I teased him once.'
(Lit. 'I hu-ed him one -mour.')
- c) Zhèzhong mò, wǒ xiǎng nǐ zuìhǎo haishi búyào yōu.
this-kind -mour I think you better still don't hu
'This way of teasing, I think you'd better not do.'
(Lit. 'This kind of -mour, I think you'd better not hu-.'

In (24a) yōu-mò is followed by an object 'him,' and the two syllables are not separated. By both the PSC and the LIH yōu-mò is a word in (24a). However, in (24b) and (24c), yōu and mò are separated and may be appropriately regarded

as elements of a phrase.⁷ Again we have a paradox.

Still another example is given in (25). Xiǎo-biàn might be rightly taken as a compound in (25a) consistent with both the PSC and the LIH, but must be regarded as an idiom phrase by both counts in (25b) and (25c).

- 25a) Tā qù xiǎo-biàn le.
he go urinate LE
'He went to urinate (have a minor convenience).'
- b) Tā xiǎo-le èrshí fēnzhōng de biàn.
he minor-LE 20 minutes DE convenience
'He had a twenty minute "minor convenience."
(Lit. 'He minored a twenty minute convenience.')
- c) Zhège biàn, tā yòu xiǎo bu chéng le.
this convenience he again minor not successful LE
'This minor convenience, he again can't have.'
(Lit. 'This convenience, he again can't minor.')

Note that it is entirely reasonable to regard dān-xīn in (23a) and yōu-mò in (24a) each as a word, and this is further supported by the fact that they are not separable precisely in these sentences, where each of them is followed by an outer object, as one can see below:

- 26a) *Tā dān-le sān nián de xīn zhèjian shì.
he carry-LE three year DE heart this matter
'He worried about this matter for three years.'
- b) *Xīn, wō yìdiǎn dou bù dān zhèjian shì.
heart I a-bit all not carry this matter
(Lit. '-Ry, I don't wor- about this matter.')
- 27) *Wō yōu- le yi -mò tā.
I hu- LE one -mour he
'I teased him once.'

We have illustrated one type of problem that arises when certain so-called compounds are so identified on the basis on the LIH and the PSC. The second type of problem

occurs with the resultative compounds. Consider first the examples in (28).

- 28a) Tā qí mǎ qí de [Ø hen lèi].
he ride horse ride COMP very tired
'He rode a horse until he got very tired.'
- b) Mǎ bèi tā qí de [Ø hen lèi].
horse by he ride COMP very tired
'The horse was ridden by him until it got tired.'
- c) Tā bá mǎ qí de [Ø hěn lèi].
he BA horse ride COMP very tired
'He rode the horse until it got tired.'

The three sentences vary with respect to the reference of the missing subject of the resultative clause. In (28a), the rider got tired, and in (28b) and (28c), the horse got tired. It appears that this may be accounted for by the following principle, stated in (29):

- 29) The missing subject of a resultative complement must refer to a ba-NP in the main clause if there is a ba-NP. Otherwise, it must refer to the main clause subject.

Now let us consider sentence (30):

- 30) Tā dǎ-sǐ le Zhāngsān.
he hit-dead LE
'He hit Zhāngsān dead.'

Note that the element si does not refer to the subject ta. This means that the interpretation rule (20) must be prevented from applying here to give rise to the unavailable reading. But this is OK, because in this sentence dǎ-sǐ is followed by the object Zhāngsān, and must be regarded as a word by virtue of the PSC. And, by virtue of the LIH, the interpretation rule (29) does not apply to a proper subpart of this word. Therefore, we correctly allow (30) as grammatical, even though si does not refer to the subject.

A problem arises when we consider the sentences in (31):

- 31a) Tā chī-bǎo le fàn le.
 he eat-full LE rice LE
 'He ate rice (and got) full.'
- b) *Fàn bèi tā chī-bǎo le.
 rice by he eat-full LE
 *'The rice was eaten full by him.'
- c) *Tā bǎ fàn chī-bǎo le.
 he BA rice eat-full LE
 *'He ate the rice and made the rice full.'

As first observed by Hashimoto (1971), the contrast in grammaticality between (31a) and (31b-c) is clearly related to the contrast in interpretation between (28a) and (28b-c). That is, (31b-c) are ill formed because the rule (29) forces the absurd interpretation that, as a result of his eating the rice, the rice becomes full. Now, note that chī-bǎo must be regarded as a word in (31a), by virtue of the PSC. But in (31b-c) the same sequence must not be analyzed as a word, because in that case the LIH would prevent the rule (29) from having any effect. In order to allow (29) to rule out (31b-c), then, we must regard chī-bǎo as a phrase in these sentences. (The view about chī-bǎo in (31b-c) is also consistent with the PSC, because chī-bǎo is not followed by an object.) We thus have a paradox again, where some verb-result strings must be regarded as words under some circumstances, but as phrases under others.

5. Further Discussion

The problem we are facing is how to resolve the paradox. The most plausible idea, apparently, is to say that the kinds of verb-object and verb-result sequences that we have seen have a dual status, either as words or as phrases. Whether it is actually a word or a phrase is

determined by independent principles of the grammar and by the context of its occurrence. In particular, when such a sequence is followed by an object, the independent principle PSC requires it to be a word and to exhibit properties consistent with the LIH. When it is not followed by an object, the PSC does not require it to be a word, and it may exhibit phrasal properties inasmuch as the LIH is irrelevant. This idea appears to be right, leaving us with the question of how to execute it.

There are three possible ways to instantiate this idea. First of all, we may assume that all so-called verb-object compounds and resultative compounds are listed in the lexicon as both words and phrases. In other words, they are listed twice, or entered in the lexicon with parallel structures. Independent principles will guarantee that they are inserted into the right position with the right structure. In particular, such sequences must be inserted each as a compound when there is a following object, as in (23a), (24a), (30) and (31a). However, these elements must be inserted as phrases when they occur in final position, so that they may be separated by syntactic rules as in (23b-e) and (24b-c), and so that the interpretation rule (29) may take effect as in (31b-c).

A second possible solution is to list all V-O and V-R combinations as words in the lexicon, and invoke a rule of reanalysis which relabels them as phrasal categories in appropriate environments, specifically, when they occur sentence-finally. This is the view implicit in Chao's (1968) treatment, where he calls the phenomenon 'ionization.' In terms of the X-bar theory of phrase structure, we may formalize ionization as a rule of reanalysis that turns an X-zero category into an X-one-bar category. Note that this rule of ionization must be obligatory when it can apply, when a given combination occurs sentence-finally. This is because the sentences in (31b-c) are ungrammatical, and an optional rule of ionization would not suffice to rule them out.

The third possible solution is the opposite of ionization. According to this hypothesis, we list all V-O and

V-R combinations only as phrases in the lexicon. The term 'compound' is thus inappropriate to them in the lexicon. Rather, dān-xīn, etc. would be listed as idiom phrases. Then, when these phrases are inserted into sentence-final position, nothing need take place. But if inserted into sentence-medial position, with an object following, they would undergo a process of lexicalization, by which a V-one-bar category is reanalyzed as a V-zero category, namely a phrase becomes a word. This then will prevent the PSC from taking effect, and save the sentence from the PSC. This rule of lexicalization can be seen as a synchronic reflex of the historical process by which many compounds have been derived. Indeed, this rule may be seen as another phenomenon which, like the other phenomena mentioned in section 2 (verb copying, etc.) functions to save an otherwise illegitimate phrase structure from the PSC. Note that the rule need not be made obligatory. We may simply assume that it simply has the option of applying. The existence of the independently motivated PSC will ensure that only when reanalysis takes place will the sentences be grammatical.

The three possible solutions do not differ as far as their empirical coverage is concerned. From a theoretical point of view, arguments that are available in favor of one or the other of these instantiations also appear to be quite subtle. Nevertheless, we would like to suggest that the third solution appears to be the most plausible. There are two reasons that lead to this belief. First, this solution does not require lexicalization to be an obligatory rule. Its application is required only by the independently motivated PSC. When no such motivation exists, it simply does not apply. The ionization rule, on the other hand, must be stated as obligatory when a given V-O or V-R combination occurs sentence-finally. But this stipulation must be made in an ad hoc manner. The parallel structure hypothesis is also defective in the same way, in that we do not have any independent principle which forces one to insert a phrase rather than a word in sentence-final position. (There is no reason why we cannot insert a V-O or V-R compound as an intransitive verb sentence-finally.) The second reason has to do with the fact that the

phenomenon of ionization as a rule of grammar appears to be extremely exotic. Ionization involves the increasing of the complexity of a structure, or increasing the depth of embedding in a tree. Lexicalization, on the other hand, has the effect of regularizing a more complex structure into a simpler one: making a simple word out of a phrase. The latter phenomenon has been reported in practically every language, but not the former. On grounds of plausibility, therefore, I would prefer the third solution. I would, in other words, claim that what is often called a V-O compound and a resultative compound is really basically a phrase in the lexicon, often an idiom phrase. It is only in certain limited environments that such phrases become compounds.

We will end this paper with a remark on the status of the vast number of verb-object and verb-result combinations in a proper theory of the synchronic grammar of Chinese. Historically, it is a well-known fact that V-O compounds and resultative compounds, like other compounds, are derived from phrases having the same internal structure.¹⁰ It is also well agreed that this process of word-formation has affected various items in various degrees. Thus, some items have completely been turned into words, as shown by their complete inseparability and their ability to take an object. Such items are relatively few, including for example, gūan-xīn 'be concerned about,' zhù-yì 'pay attention to,' chū-bān 'publish,' tí-yì 'propose,' dē-zuǐ 'offend,' etc. Some items have the ability of taking an outer object, but are separable when they do not take such an object. These include the examples like dān-xīn 'worry,' bāng-máng 'help,' fù-zé 'be responsible for.' So-called resultative compounds also belong to this group, since they may be followed by an object, but when they are not followed by an object they may be affected by the interpretation rule (29). Finally, a vast number of items do not have either property of the first groups of items. They may not take an outer object even though they may be semantically understood to be transitive, denoting an action that affects some object. And they may be separated in a number of ways. This class appears to include a majority of what are listed as V-O compounds by most writers (cf. Chao 1968, Lu 1964, and Li and Thompson 1981):

buō-pí 'peel,' dǎ-ěrguāng 'slap,' bǎng-piào 'kidnap,' shēng-qì 'be angry at,' sù-kǔ 'complain,' lì-fǎ 'give a haircut to,' kāi-wánxiào 'make fun of,' chī-dòufu 'tease,' kāi-dǎo 'operate on,' xiào-lǎo 'serve.' Many V-O combinations that are intransitive in meaning also belong to this group, because they may not be followed by an adverbial complement like 'two times,' or 'three days, and they cannot be directly followed by a resultative or descriptive complement without their first syllable being reduplicated immediately before the complement: shuì-jiào 'sleep,' wán-shuǎ 'play,' zuò-gōng 'work,' yǎn-xì 'act,' diǎn-tóu 'nod,' tiào-wǔ 'dance,' niǎn-shū 'read,' etc.

In view of this, we would like to point out a methodological error in traditional synchronic treatments of the majority of these items. According to Chao and others, all of these items are described as 'compounds,' and their differences in separability being described, in effect, by saying that some compounds are not separable while some others are separable in some limited ways. Recall that the single valid criterion for the identification of compounds is the LIH. By saying that some compounds need to obey the LIH while others do not, the theory destroys its only principle, and is at a loss as to what after all is a compound. Obviously, this is too great a loss, for the LIH has in other areas proved to be a valid principle for compound identification, e.g. in the identification of subject-predicate, coordinate, subordinate compounds, and of V-O compounds that are exocentric.

What we would like to suggest is that most of the so-called V-O compounds and resultative compounds are not compounds, but are rather phrases, many of them idiom phrases. First, consider the third types of items enumerated above. If these items are phrases, then it is a natural consequence of the PSC that they cannot be followed by an outer object, nor by a duration or frequency adverbial, nor by a resultative or descriptive complement without the verb undergoing reduplication. It is also entirely consistent with the LIH that they are separable. Consider now the second class of items. These items are separable in some limited ways. The most important

situation under which separation cannot take place is when they are followed by an outer object or an adverbial complement (see 26) and (27). But this is precisely when they must be recognized as words by virtue of the PSC. Where the PSC does not require us to identify them as compounds, they may be naturally expected to exhibit phrasal properties. In other words, the second class of items are best described as phrases inherently (again many as idiom phrases) which syntactically become compounds under certain syntactic environments. The difference between the second and the third types of items is that while an item of the second type is specified in the lexicon as an idiom phrase that can undergo the lexicalization rule referred to above, an item of the third type is not so specified, and by an appropriate marking convention, is incapable of undergoing the rule. Finally, as for the first type of truly inseparable compounds, we may, of course, list them as true compounds in the lexicon. Or alternatively, they may be listed as idiom phrases that obligatorily undergo the lexicalization rule. It is clear that within the view advocated here, most (if not all) so-called V-O and resultative compounds are phrases in the lexicon, some of which become compounds only when they enter into relations with other elements in a syntactic tree. Under such a view, linguistic theory need not tolerate a contradictory notion of compounds which defines them in terms of the LIH but speaks of the existence of some 'separable compounds.'

NOTES

*This is a slightly revised version of the paper read at the Chinese formal grammar panel of the 1983 AAS meeting. I am grateful to Professor Hsin-I Hsieh for presenting this paper on my behalf at the conference. The contents of this paper have also been orally presented at a linguistics conference on East Asian Languages held at USC in October 1983, and are scheduled to appear in its proceedings. I am grateful to Professors R. Cheng, Y.C. Li, and S.H. Teng for useful comments on earlier versions of the paper.

¹More precisely, this condition should be understood as restricting the hierarchical structure of a phrasal category, which marks left-branching structures of the following kinds as ill-formed: [[[...]]...] or [...[...]]..., where each pair of less inclusive brackets exhaustively contain the head of the phrase included in the next more inclusive pair of brackets. The first structure is ill-formed because the head branches to the left twice, and the second is ill-formed because the intermediate pair of brackets, which occurs on the left branch of its parent node, is not on the lowest level of phrasal expansion. This more precise formulation of the PSC is defended at length in C-T. J. Huang (1982).

²We have presented the PSC here only as a descriptive principle. In a more rigorous formal treatment, it is possible that it may be derived as a theorem of other independent principles of Universal Grammar and independently motivated parameters pertaining to Chinese. For a plausible attempt at deriving this condition as a theorem, see Li (1984).

³The English example is not well formed as it stands, unless the preposition of is inserted before the outer object Bill. It is a familiar assumption, however, that this requirement has little to do with the semantics of the example, and the preposition is inserted as a case-marker for purely grammatical reasons, and therefore may be 'invisible' at the semantic level. We believe that the preposition ba has a similar status in Chinese.

⁴Our view here regarding the motivation of verb-copying as a purely formal one may remind the reader of Chao's (1968) treatment. In Chao's explanation, a verb is reduplicated when followed by an object and an adverbial complement because each of these following elements has to be close enough to the verb. Chao's account does not, however, extend to some of the other phenomena that we have taken to follow from our account. For example, the motivation for reduplication is removed in case the object is topicalized, as we show below in the text, even though the topicalized object is now far from its verb, in

violation of his functional principle. Furthermore, Chao has stated his principle in purely linear terms, which is empirically inadequate in the face of 'double object constructions' and the like, cf. note 7. We follow Mei (1978) and assume that adverbial complements of the kind shown in (9) are 'verb phrase complements' (and not 'verb complement') in that they are sisters to the verb-object phrases preceding them.

⁵Some of the facts mentioned in this section have also been accounted for along similar lines in the literature, but as far as we know, ours is the only explicit account that attempts to give a unified explanation for all of them, with the exception of S.F. Huang (1983), who has independently proposed a linear version of ours (1). Before the existence of this latter account came to our knowledge, however, we have explicitly argued against it on grounds of its empirical inadequacy, in C-T. J. Huang (1982). An example that shows the superiority of our hierarchical account over S.F. Huang's linear account is the sentence (24b). Cf. note 7 below.

⁶A-not-A questions like NÍ gāo-bu-gāoxìng? 'Are you happy?' appears to be a counterexample to this statement. But this is a counterexample only if one assumes that such questions are derived via a syntactic rule of conjunction reduction, along the lines of Wang (1967). There is good reason, however, to believe that what is involved in such constructions is a phonological process of reduplication. And we assume that the LIH restricts only phrasal syntactic and semantic rules, but not phonological or morphological rules. For some arguments for a reduplication analysis of A-not-A questions, see C-T. J. Huang (forthcoming).

⁷Note that in (24b) the element yōu- (which is now analogously used as a verb) is followed by two constituents, 'him' and 'one-mour.' This structure is allowed by our PSC understood in the hierarchical sense, because it may be analyzed as a sentence on a par with 'double-object constructions' like 'I give him a book.' Both the two constituents following the verb are sisters of the verb, which therefore occurs only once on the left branch of its

rent node and on the lowest level of expansion. Such examples as these argue against a linear interpretation of a PSC.

⁸The principle has been given here only as a descriptive principle. We believe that this again can be derived as a theorem from more general principles regarding the inference of missing elements, within an optimal universal theory of anaphora. In particular, if we assume that the preposition ba is not present at the semantic level (given its semantic emptiness), then what (29) says is that the antecedent of an empty argument is the 'closest' NP that c-commands it, where A c-commands B if the first branching node dominating A also dominates B (cf. Reinhart 1976). This more general principle has been shown by many to be motivated on other grounds.

⁹A possible objection to this third approach may arise on consideration of examples like yōu-mò 'tease' and liǎo-biàn 'urinate.' Since yōu-mò is coined from English 'humour' and xiǎo-biàn is exocentric, with an A-N combination being used as a verb, they must start out at some point in the lexicon as words. This argues for listing them as words in the lexicon and invoking an ionization rule that accounts for their separability in the syntax. Note, however, that such examples are sporadic, and cannot be used to argue for general hypotheses. It is plausible to assume that they get ionized as idiom phrases first in the lexicon, after which they then behave like other V-O combinations that are listed as phrases in the lexicon, and are subject to the same treatment as the latter.

¹⁰Considerations of this fact argue against a purely lexical approach to resultative compounds such as proposed by Thompson (1973b). If resultative compounds are formed entirely within the lexicon, without access to information about the syntax of the language, why is it that the order of elements within the resultative compound mirrors precisely the word order of verb-result phrases? Similarly, if V-O compounds are formed entirely in the lexicon without access to syntactic information, why do we not have O-V compounds? It appears that any adequate theory of compound

formation must rely on the syntactic component as a possible input to a compound rule.

REFERENCES

- Chao, Y.R. 1968. A Grammar of Spoken Chinese. University of California Press, Berkeley, California.
- Hashimoto, A.Y. 1971. Mandarin Syntactic Structures. Unicorn 8, Chinese Linguistics Project, Princeton University.
- Huang, C-T.J. 1982. Logical Relations in Chinese and the Theory of Grammar, doctoral dissertation, MIT.
- Huang, C-T.J. (forthcoming) 'A Modularized Approach to Annot-A Questions in Chinese.'
- Huang, S.F. 1983. 'Morphology as a Cause of Syntactic Change,' in S.F. Huang, Papers in Chinese Syntax, Literary Crane Press, Taipei, Taiwan.
- Jackendoff, R. 1972. Semantic Interpretation in Generative Grammar, MIT Press, Cambridge, Massachusetts.
- Li, A. 1984. 'Case Assignment and Word Order in Chinese,' mimeographed, USC.
- Li, C. and S. Thompson. 1981. Mandarin Chinese: A Functional Reference Grammar, University of California Press, Berkeley, California.
- Lu, Z.W. 1964. Hanyu de Gouci Fa. Peking.
- Mei, K. 1978. 'On Verb-Phrase Complements in Chinese,' in A Festschrift for Professor Chu Wan-Li.
- Postal, P. 1969. 'Anaphoric Islands,' Proceedings of the 5th Meeting of Chicago Linguistic Society.

- Reinhart, T. 1976. Syntactic Domains of Anaphora, doctoral dissertation, MIT.
- Selkirk, E. (forthcoming) Phonology and Syntax: The Relation between Sound and Structure, MIT Press, Cambridge, Massachusetts.
- Thompson, S. 1973a. 'Transitivity and the Ba-Construction in Mandarin Chinese,' Journal of Chinese Linguistics 1.2.
- Thompson, S. 1973b. 'Resultative Compounds in Mandarin Chinese: A Case for Lexical Rules,' Language 49.2.
- Wang, W. 1967. 'Conjoining and Deletion in Mandarin Syntax,' Monumenta Serica 26.