REDUPLICATED CONSTRUCTIONS IN CHINESE AND QUESTIONS OF GENERATIVE POWER*

Jonathan Calder

(School of Epistemics,
University of Edinburgh)

1. Introduction

In the following pages, it is my intention to demonstrate the problems that some data from Mandarin Chinese pose for modern linguistic theories. The study focuses on the reduplication involved in a construction known as the 'A-not-A' construction and I shall not try to provide a grammar for a large fragment of Chinese. Instead, this paper will use the data and possible descriptions thereof to investigate the question of whether context-free languages, hereinafter CFL's, are sufficiently adequate for the task of describing certain linguistic phenomena. Taking Pullum (1984) as a starting point, I shall consider whether the Chinese data in question constitute a case against the claims of Generalized Phrase Structure Grammar (hereinafter GPSG).

In the rest of this introduction, I offer some methodological justification for dealing with this problem. I then proceed to an overview of the A-not-A construction. Section 3 examines methods that may be used to disprove claims of non-context-freeness, and their applicability to the Chinese data.

1.1 Context-freeness and the Methodology of GPSG

The fall and rise of context-free phrase structure grammars within linguistics has been well chronicled, (Gazdar (1982), Pullum and Gazdar (1982)), and I do not intend to repeat history here. Rather, I should like to comment briefly on the status of the question 'Is English a CFL?'

It has been assumed by many linguists in opposing camps that the framework of GPSG presupposes a positive answer to this question. Their view is that a demonstration of the non-context-freeness of natural languages constitutes a knock-down argument to the central claims and methodological tenets of GPSG. This is obviously the thrust of Higginbotham (1984), for example.

York Papers in Linguistics 12 (1986) 43-66

In fact, the paradigm of GPSG has always recognised the possibility of a demonstration that English, or some other natural language, could not be described by a context-free phrase structure grammar (hereinafter CFPSG), (cf Gazdar (1982:177) and more recently Pullum (1984)).

The current answer that GPSG offers to the question of whether English is a CFL is a provisional affirmative; none of the arguments claiming to prove the non-context-freeness of English have any substance (Pullum and Gazdar (1982)), although there may be other natural languages whose description does require supra-context-free power, (Pullum (1984), and below). Given this possibility, an important methodological advantage of GPSG is that future lines of research are, in a sense, predetermined; the constrainedness of the theory allows a very clear statement of the nature and status of counter-examples and the generative capacity of the theoretical apparatus may always be revised upwards, provided a proof is given of the necessity of such a change, along with good motivation for the proposed extension. In contrast with the various models of transformational grammar available, we may know that a particular grammatical description is the least extravagant possible. Viewed in the above terms, this study is, along with Pullum (1984), an inquiry into the limits of context-free description.

It is worth emphasising at this point that, during the course of this paper, we shall be concerned with weak, rather than strong, generative capacity. That's to say that we shall be interested in showing that it is technically feasible to give a context-free grammar for the constructions in question, not that the resultant grammars are the most linguistically satisfying. As Pullum and Gazdar note (1982:498), linguistic theory demands that we pay attention to the question of strong generative capacity, but in this paper I shall merely point to ways in which the less attractive consequences of opting for context-free description may be avoided.

2. Data on the A-not-A question

2.1 Preliminaries

Before launching into a description of the data, the following notes are in order.

First of all, the collection of consistent data for any aspect of Chinese linguistics is very difficult. In this essay, I have tried to arrive at grammaticality judgements relative to standard Mainland Chinese, known in the People's Republic of China as putonghua, and in the West as Mandarin. The situation is however complicated by several factors. Available modern work in linguistics typically takes as standard Mandarin as spoken in Taiwan

(kuoyu). Even within this literature, there are many contradictory judgements. Putonghua, in spite of considerable efforts at language reform, also contains considerable variation; standardization usually focusses on phonological, lexical and major syntactic features, rather than on the finer points of syntax. Variation among the judgements of my informants was in some cases so great as to leave no discernible pattern.

Secondly, the usage of the word 'Chinese' later in this essay, might be deemed by some to be a little loose, as all my discussion, unless otherwise stated, refers to Mandarin Chinese alone. This looseness may be justified by the fact that A-not-A constructions exist in Cantonese (cf Chao (1976:198)), and forms of sentential reduplication may be found in other dialects. For example, Zhang (1979) claims that the Chaoyang dialect of the Southern Min dialect group has reduplication of whole sentences for hyperbolic effect.

Thirdly, some general typological notes. Mandarin Chinese is canonically SVO (Li and Thompson (1981:23ff)). There are many constructions which do not adhere to this order, but they will only be of marginal concern to us here. VO is typically assumed to form a verb phrase (Li and Thompson (1981:139ff)). There is no inflectional morphology.

2.2 Why might the A-not-A construction be a problem?

One option for the speaker of Mandarin, if he wishes to ask a yes/no question, is to use the A-not-A construction (following the terminology of Rand (1969), Fenn and Tewksbury (1967) and most other sources). Li and Thompson's (1981) discussion of the construction occupies about a quarter of all the space they devote to questions; that is, it is not a marginal phenomenon.[1]

In the following section, I give a characterisation, in terms as neutral as possible, of the A-not-A construction, and potentially related constructions.

2.2.1 The simplest cases

In informal terms the construction involves a verb phrase and the negated repetition of that verb phrase, as the following illustrates:

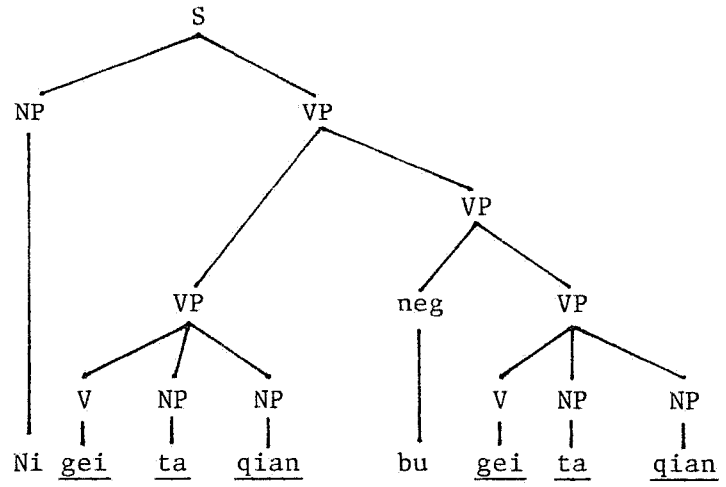
(1) Ta qu bu qu?
he go not go
'Is he going?'

(In example sentences, reduplicated sections are underlined.) The structure appears simple enough, when limited to sentences like that above, but there are numerous

complications in the behaviour of sentences involving aspectual marking or reduction in either verb phrase. A more complicated example is (2), the tree structure for which is given in (3).

(2) Ni gei ta qian bu gei ta qian?
you give him money not give him money?
'Do you give him money?'

(3)



Obviously, there are various details being glossed over here, for example the precise syntactic representation of the negation.

It should by now be clear why the A-not-A construction is so called; questions are formed by negating the repetition of some string A. We may also see why it is that such a construction might be problematical for certain linguistic theories; such constructions look very much like the **WW** languages, where **W** is a variable over strings of symbols, as described by, for instance, Aho and Ullman (1972:198). This source contains a proof that such languages may not be generated by a grammar using only context-free resources. Huang (1982b:281) has explicitly claimed that the same holds true for the Chinese construction under discussion. Let us consider the reasons for making such a statement, for which purpose we need to recall the definition of a context-free grammar. The following is taken from Hopcroft and Ullman (1979:79ff).

(4) A context-free grammar is a quadruple, V, T, P, S, where

(i) V is the finite set of variables (or syntactic categories),

(ii) T is the finite set of terminal symbols (or lexical items),

(iii) P is a set of rules which rewrite single elements of V as strings of symbols drawn from V and T, and

(iv) S is the 'start symbol', ie the syntactic category that appears at the highest node, or root, of a tree.

It is easy to demonstrate that grammars designed along the above lines may not generate a set of trees with reduplicated structure, as illustrated by the Chinese example above. An informal exposition may be based upon the following set of rules.

(5)
(i)	S	-->	NP VP	
(ii)	VP	-->	V NP NP	
(iii)	VP	-->	VP bu VP	
(iv)	NP	-->	ni	'you'
(v)	NP	-->	ta	'he' or 'him'
(vi)	NP	-->	qian	'money'
(vii)	V	-->	gei	'give'

While it is obvious that the sentence in (3) may be generated by the above grammar, it should be equally obvious that this grammar will generate other sentences without reduplication, such as (6).

(6) Ni gei qian ta bu gei ta ni
you give money him not give him you
'You give him to money (and) don't give yourself to him'

We may also note that the only reason why we succeed in reduplicating the verb gei in these cases is that there is no other choice. If we were to introduce another verb, then even this consistency would be lost. The above example allows an intuitive grasp of the problem that reduplication presents for context-free grammars. Before examining ways of circumventing this problem, let us take a further look at some relevant data.

2.2.2 Reduced forms

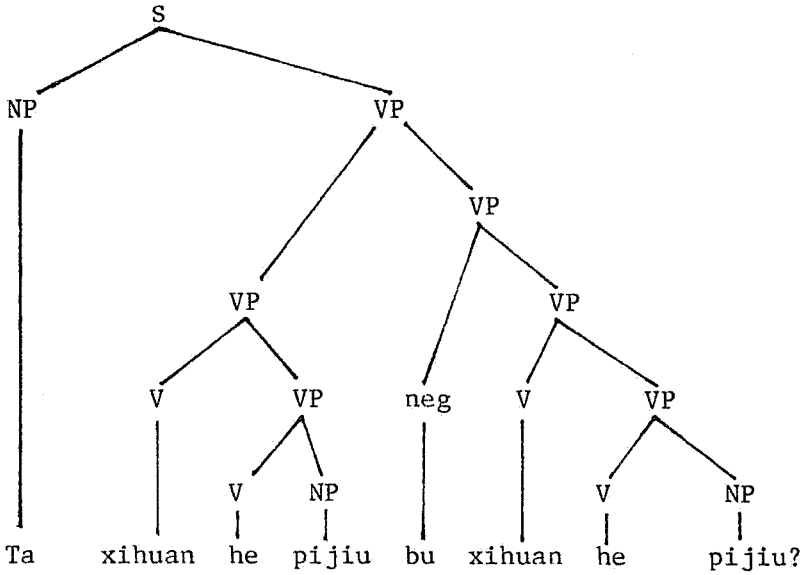
The sentences in (7) below represent a brief illustration of the possibilities of reduction. The unreduced form of the sentence in question appears on the line marked (iii) below, with the words in the reduced versions of the sentence aligned above their counterparts in the unreduced sentence. (i) and (ii) contain examples of possible reductions on the left-hand side and right-hand side (hereinafter LHS and RHS respectively). (8) is a possible tree structure for the unreduced sentence, (assuming that modals are dominated by a projection of V, cf Gazdar, Pullum and Sag (1982)).

(7) (i) Ta xi bu xihuan he pijiu?
Ta xihuan bu xihuan he pijiu?

(ii) Ta xihuan he pijiu bu?
Ta xihuan he pijiu bu xihuan?
Ta xihuan he pijiu bu xihuan he?

(iii) Ta xihuan he pijiu bu xihuan he pijiu?
NP modal V NP neg modal V NP
he like drink beer not like drink beer
'Does he like to drink beer?'

(8)



Note that all of the sentences given in (7) are grammatical, the first example in (7i) containing the usually bound morpheme xi.[2] Chao claims that this reduced form of the A-not-A question, involving what he terms 'ionised' morphemes, is a recent borrowing from Cantonese, (Chao (1976:198)). Such forms are common within this type of A-not-A questions; one informant deemed grammatical examples with 10 randomly chosen bisyllabic verbs. Such sentences are, however, completely ungrammatical when the reduction occurs on the opposite side, viz:

(9) *Ta xihuan he pijiu bu xi

The first sentence in (7ii) above illustrates the case where reduction of the RHS proceeds so far as to leave only the negation marker, a construction typical of the northern dialects, especially of colloquial Pekinese. The third sentence in (7ii) is one of the cases in which a reduction leaves more than the left-most constituent (or first syllable thereof). The inconsistencies in the data, alluded to in section 2.1, were found to be greatest in respect of sentences of this type. While some generalisations can be made over the data obtained (Calder (1984)), it is impossible to be sure that these are not the results of some

random factor. Given that the consensus within the literature and among my informants agrees on the grammaticality of unreduced and, what I shall call, fully reduced forms, I shall concentrate on these sentence types and not discuss, in any detail, the intermediate, semi-reduced forms. Neither shall I investigate the complex paradigms created by the interaction of the A-not-A phenomenon with the behaviour of aspect markers under negation.

2.2.3 What can substitute for A?

A more important concern for us than the description of the aspectually marked paradigms is the need to find some principled specification of the class of categories that may enter into A-not-A constructions. It is obvious, from (5) above, that the construction may consist of full VP's, but we need to know the categorial identity of the elements that appear in the reduced forms. In cases where a single word alone is repeated then the elements involved fall into the (English) classes of preposition (9)[3], adverb (11) and degree modifier (12), as well as into the classes of verb and modal verb (13) (there is a lot of lexical idiosyncrasy in the first two cases).

(10) PREPOSITION
Ni gei bu gei ta mai shu
NP P neg P NP V NP
you give not give him buy books
'do you buy books for him?'

(11) ADVERB
Ta zai bu zai ti zuqiu
NP ADV neg ADV V NP
he again not again play football
'Is he going to play football again?'

(12) DEGREE MODIFIER
Zhei ben shu tai bu tai zhong?
DET MW N DM neg DM ADJ
this MW book too not too heavy
'Is this book too heavy?'

(13) MODAL
Wang Xiansheng neng bu neng gei wo jie qian
PN Title modal neg modal P NP V NP
Wang Mr. able not able give me lend money
'Is Mr. Wang able to lend me money?'

There appears to be an absolute proscription on reduction in sentences containing reduplicated adverbs. Thus, although the sentence

(14) Ta manmarde pao bu manmarde pao?
he slowly eat not slowly eat
'Does he eat slowly?'

is considered grammatical by Li and Thompson (1981:538, example 83), any reduction is ungrammatical.

2.3 Other constructions involving reduplication

We may note other constructions that seem to require repetition of lexical material.[4] Obviously, if we decide to invoke any syntactic mechanism to ensure that the lexical material in A-not-A questions is faithfully reduplicated, this may also be of use in accounting for these constructions. Likewise we will want any analysis of the A-not-A construction to be readily extendable to these constructions.

For instance, durative constructions with objects are formed by a verb phrase consisting of verb plus object, then the repeated verb followed by a noun phrase describing the length of time over which the action occurred. Thus (15), for which a possible structure is (16).

(15) Lao Deng xue Yingwen le, xue san nian le
Old Deng study English ASP, study three year ASP
'Deng has been studying English for three years'

(16) $[_S[_{NP}$ Lao Deng$]$

$[_{VP}[_{VP}$xue $[_{NP}$Yingwen$]$ le$][_{VP}$xue $[_{NP}$san nian$]$ le$]]]$

Very similar to such durative forms are the 'resultative complement' forms (cf Huang (1983), Tewksbury and Fenn (1967:130)). Again, if the verb takes an object and a resultative complement, then the verb must be repeated, first preceding the object, then preceding the resultative complement, as in:

(17) Ta zou lu zou de hen lei
He walk street walk till very tired
'He tired himself out walking'

Such sentences are discussed in greater detail in section 3.2.1. below.

2.4 Comparison with VP coordination

One possible analysis of the A-not-A construction is to view it as a special case of VP coordination; Wang (1964, 1965) offers precisely such a treatment. The following is a brief sketch of potentially relevant facts.

VP coordination is very common in Chinese, both as a result of syntactic processes, and as a discourse phenomenon when subject pronouns are dropped. Typical examples of the former are the durative and resultative complement constructions, described above. The latter do not concern us particularly here. An example of ordinary verb phrase coordination is (18):

(18) Lao Wang dao menkour, bu neng jin qu
Old Wang arrive door, not able enter go
'Wang got to the door and/but couldn't go in'

There is a very wide range of coordinating and subordinating conjunctions in Chinese. The most common of the former are ye, with the conjunctive meaning 'also', which can appear before both conjuncts and haishi, with the disjunctive meaning 'or', which typically appears only before the second conjunct. These are illustrated in (19) and (20) respectively.

(19) Zai Beijing de shihou, wo ye zuo shi ye
At Peking RM time, I also do business also
waryiwar.
play
'When I'm in Peking, I do business and have some fun'

(20) You kongr de shihou, wo kan shu haishi ting
Have free-time RM time I read book or listen
yinyue.
music
'When I have free time, I read or listen to music'

As illustrated by example (18) above, there is no requirement that an overt conjunction appear. If there is no overt conjunction, then the interpretation of the coordinated verb phrase is highly context-dependent. Thus the sentence:

(21) Ta chi fan chi mian
he eat rice eat noodles

may have either a conjunctive reading, 'he eats rice and noodles', or a disjunctive, interrogative meaning, 'does he eat rice or noodles?'

The most important difference between A-not-A questions and VP conjunction is that the reduction phenomena associated with the former do not occur in the latter. Informally, in the cases of left-hand side reduction in A-not-A questions, eg (22):

(22) Ni he bu he pijiu
you drink not drink beer
'Do you drink beer?'

we might wish to say that the noun pijiu is construed as the object of both occurrences of the verb he. A similar interpretation of a sentence with conjoined, non-identical Vs cannot use the same syntactic structure. Thus (23) is deemed ungrammatical with pijiu construed as the object of both niangzao, and he.

(23) *Ta niangzao bu he pijiu
he brew not drink beer
'He brews but doesn't drink beer'

(24) a) (Guanyu) pijiu, ta niangzao bu he
as for beer he brew not drink

b) Ta niangzao pijiu, bu he
he brew beer not drink

The reading discussed above is available if the noun in question is topicalised ((24a) above) or appears as the object of the first verb (24b), although the first rendering is preferred.

However, more than a little care must be taken here, as some verb coordination might seem to be possible, provided there is an overt conjunction: Li and Thompson (1981, 18.3.1 ex.57) offer the sentence:

(25) Ni chao haishi zheng zhei ge qingcai
you fry or steam this MW vegetable
'Do you fry or steam this vegetable?'

This example might be excluded from the data on the grounds of dialect difference, as Li is a native of Taiwan. However the situation is more complex than this; Huang (1982b:278-279), also from Taiwan, claims that 'coordinate V's' are 'quite unnatural', and stars the following sentences:

(26) *Zhangsan zhong (ye) mai gua
Zhangsan grow (too) sell melons[5]
'Does Zhangsan grow or sell melons?'

(27) *Zhangsan xie (haishi) mai shu
Zhangsan write (or) sell books
'Does Zhangsan write or sell books?'

My informants agreed with Huang's grammaticality judgements in the cases where there is no overt conjunction; one of them agreed with Li and Thompson in respect of sentences with overt conjunction. In what follows, I shall follow Huang in the case of the former, and Li and Thompson in the case of the latter.

Another respect in which the A-not-A construction is claimed to differ from ordinary verb phrase coordination is in the non-reversibility of constituents. Chao (1968:267-

270) notes that 'the order of items in coordination is grammatically reversible [...] but in the particular case of V-not-V [= A-not-A JC] questions the order is fixed ...'.

Thus Chao would rule out sentences such as (28):

(28) *Ni bu lai lai
you not come come
'Are you coming?'

Chao notes that this constraint does not apply in cases where there is an overt conjunction.

From the above discussion, it is fair to conclude that, in some dialects at least, certain verbal constructions are available only in A-not-A questions, and therefore when there is identity of lexical material. I shall return to this point in section 3.2.2, dealing with the adequacy of phrase structure approaches, after a formal statement of the descriptive problems posed by A-not-A questions and some discussion of ways to obviate these problems.

3. How should we deal with reduplication?

Pullum (1984) points out the major problems that the phenomenon of reduplication causes to grammatical formalisms which use only context-free resources. The essential nature of the difficulty is that a construction of the form

(29) $a_1\ a_2\ a_3\ \ldots\ a_n\ b_2\ b_3\ \ldots\ b_n$

requires supra-context-free devices for its generation, under the following conditions:

i) a_i and b_i are words or morphemes of a particular language

ii) there is some sort of dependency between a_i and b_i, and

iii) there is no upper bound on n.

In the following sections I shall first of all review what arguments may be used to demonstrate that the existence of such constructions does not necessarily compromise the thesis of the context-freeness of natural language, and then consider how the A-not-A construction may be handled within the syntactic framework of GPSG.

3.1 Dealing with apparent non-context-freeness

In this section I shall look at two of the ways in which GPSG may avoid invoking non-context-free syntactic

apparatus, while describing putatively non-context-free constructions. I shall avoid discussion of the more well-known aspects of GPSG, such as slash categories or the extensive use of features, assuming familiarity with the systems as introduced by Gazdar and others (e g Gazdar (1982), Gazdar and Pullum (1982), Gazdar, Klein, Pullum and Sag (1985) etc). Before examining the techniques that will help us in this, we should first of all consider the importance of the term 'stringset' in this discussion.

The notion of the stringset is used heavily by Pullum and Gazdar (1982) in their debunking of the published arguments claiming to demonstrate that descriptions of natural languages must invoke greater than context-free power. It involves treating natural languages 'purely as sets of strings of words' (op. cit. p471); in other words, the object of interest is the weak generative capacity of the apparatus required. The notion is crucially implicated, as the authors spell out, in their treatment of Dutch cross-serial dependencies (op. cit. p489), a construction which is putatively non-context-free (cf. Huybregts (1976), Bresnan, Kaplan, Peters and Zaenen (1982)). As they say, their context-free approach may not generate a description that will feed a compositional semantics. Nevertheless, the grammar they provide does succeed in generating the required sentences and does not overgenerate.

3.1.1 Expanding the data

One type of argument that may be invoked to avoid increasing the power of a grammar is that of increasing the data set that the grammar accounts for. Pullum and Gazdar (1982:490ff) use this approach to deal with what had been claimed by Postal to be a construction intractable for context-free methods. Postal's claim was that the process of 'noun incorporation' in Mohawk required a string-copying operation. Langendoen, in his reconstructed version of the argument (ibid. for references), intersects a finite state language with some example Mohawk sentences, and shows that the resulting language is of the **WW** type, as described in section 2.2.1 above, and therefore not a context-free language. This implies that the language with which the finite state language was intersected cannot be context-free.

The argument is sound, but Pullum and Gazdar note that the same may perhaps not be claimed of the initial premises. In particular, they contest the claim that the nouns involved in incorporation must be identical. They give data showing that in fact there are similar sentences on which the requirement for identity does not hold. Therefore, we may allow our grammar to generate sentences with any noun substituting for the category symbols dominating the positions in question. In certain cases the same noun may be substituted for both positions, giving rise to those

constructions that are thought to be context-sensitive. However we do not need to use context-sensitive apparatus to check that identity of insertion has occurred, as any non-identical insertions will be grammatical sentences of the other type of construction. In other words, we deny the first condition in section 3 for non-context-freeness, by claiming that there is no dependency between the morphemes in question. We shall revisit this particular strategy in looking at A-not-A questions in the GPSG framework.

3.1.2 Semantic filtering

Pullum (1984) gives examples showing that, if the semantic component of a grammar is allowed to rule out certain of the strings produced by a CFPSG, then the set of strings sanctioned by the grammar as a whole may be undescribable using simple context-free rules. His particular example uses a context-free syntax and filtering to turn a context-free stringset into a **WW** language. As with the above ploy, it will be of use in our descriptions of the A-not-A construction, and we shall return to it in section 3.2.3. It is obvious, but worth emphasising, that any argument invoking semantic filtering is worthless unless it makes explicit the means by which such filtering takes place.

The method used by Pullum is to associate a particular semantic type with each terminal symbol. Once a sentence has been generated, its semantic structure is built up, in accordance with the semantic specification for each rule: this much is standard GPSG. It may happen, however, that at some point the semantic rule calls for semantical objects of inconsistent types to be combined. In this case, no semantic interpretation of the sentence may be obtained, and the sentence is therefore not accepted by the grammar.

3.2 Chinese within the GPSG framework

I shall now turn to possible treatments of the Chinese data within the framework of GPSG. Armed with the techniques described above, we shall look first at some suggestions made by Huang (1983) in respect of the resultative complement construction, and then consider how we might approach other potentially troublesome constructions involving reduplication.

3.2.1 Resultative complements

Chu-Ren Huang (Huang 1983) offers phrase structure rules in the immediate dominance and linear precedence (hereinafter ID/LP) format of Gazdar and Pullum (1982:18f) to deal with the resultative form of verb reduplication described in section 2.4. above. He contests the assertion made by C-T James Huang, during a lecture to the Linguistic

Institute, 1983, that such sentences require more than context-free power for their generation. Thus, sentences of the form

(30) Ta [$_{VP}$ qi ma] [$_{VP}$ qi de hen lei]
he ride horse ride till very tired
'He rode the horse until he was very tired'

where there is repetition of lexical material, may be assimilated to those of the form

(31) Wu da niang [$_{VP}$ shang jie] [$_{VP}$ zou de hen lei]
Wu big wife up street walk till very tired
'Eldest Wu's wife got very tired by walking downtown'

(Glosses and translations are from Huang (1983:11)). Huang claims that these are instances of a general schema of verb phrase conjunction, of which the immediate dominance (ID) statement is:

(32) VP --> VP, VP
[+VN] [+COMPL]

where the different features are required to allow for the occurrence of de in the second VP. The feature [VN] forces the expansion of the VP to contain an object. The accompanying linear precedence (LP) rule is:

(33) VP < VP
[+VN] [+COMPL]

As Huang mentions, this is a very specialised LP rule and one which it is hard to motivate with respect of the rest of the grammar. (We may note however that it will be of use in dealing with the durative constructions of section 2.3.) Thus the formal argument is identical to that alluded to in section 3.1.1 above, namely we deny the thesis that there is a dependency between the constituents in question.

In order to facilitate later discussion, I repeat in (34) to (39) below some of Huang's rules. As mentioned above, these are given in the ID/LP format, due to Gazdar and Pullum (1982). For our purposes, the most important of these are the rules for the expansion of S, and VP, which are given below.

(34) a. S --> XP[PRED]

b. [$_{VP}$ xia yu le]
fall rain ASP
'It is raining'

(35) a. S --> NP, XP[PRED]

b. [$_{NP}$ hua] [$_{AP}$ hen mei]
flower very beautiful
'Flowers are very beautiful'

(36) a. VP --> V, NP

b. [$_{V}$ da] [$_{NP}$ ta]
hit him
'hit him'

(37) a. VP --> V

b. [$_{V}$ fei]
fly
'fly'

We also need the following linear precedence statements:

(38) a. V < NP < AP < N
b. NP < VP

and the following Feature Co-occurrence Restriction:

(39) [PRED] ⊃ [+V]

which guarantees that expansions of S have as their predicate either verb or adjective phrases.

With Huang's rules as our basis, let us look at the potential for context-free descriptions of other examples of reduplication.

3.2.2 Context-free treatments of reduced A-not-A questions

Given the data in section 2, we may now see how various attempts to show that Chinese is not a context-free language may be defused. I shall first of all consider the fully-reduced forms of A-not-A questions, i e those of which one side has been reduced to a single lexical or sub-lexical element. I shall then investigate the implications of the full forms of A-not-A questions for the power of our theoretical devices.

Consider first of all the fully reduced forms of A-not-A questions. In these cases one might suppose that a proof of the following form demonstrates the non-context-free nature of this construction. (James Huang (1982b:281) suggests this possibility, but does not provide any formal proof. The following is one possible elaboration of his comments.)

> Verb conjunction in Chinese is allowed exceptionally in the case of A-not-A questions, hence the grammaticality judgements for examples (23) section 2.4, repeated below:
>
> (40) a) Ta niangzao bu niangzao pijiu?
> he brew not brew beer
> 'Does he brew beer?'
>
> b) *Ta niangzao bu he pijiu.
> he brew not drink beer
> 'He brews but doesn't drink beer'
>
> c) NP V_i neg V_j NP
>
> Assuming a representation like c), we must invoke a rule introducing V_i and V_j which imposes an identity condition on the rewriting of these elements; such a rule is obviously context-sensitive and so this construction may not be generated using only context-free resources.

The flaw in this argument is the assumption that the identity condition can not be expressed in the rule itself, which would only be the case if the class of items introduceable by V_i and V_j was infinite. (This would contravene the definition of context-free grammar, cf (4) i) above.) However, calculations suggest that there is a finite number of phonologically and syntactically distinct verbs in Chinese,[6] and we may use this information to demonstrate that certain forms of the A-not-A question are tractable within a context-free framework. Thus if we assume a rule of the form

(41) $V_i \longrightarrow V_i$ bu V_i

where V_i stands for that preterminal category that may be rewritten only as niangzao, then we may use rules of this form to generate all and only the legal conjunctions of Chinese.[7] In the cases where the verb is repeated to the right of the full VP, we may borrow Chu-Ren Huang's tactic of assimilating such forms to ordinary verb phrase coordination, in which case there is no identity requirement, as example (24b) shows, repeated here:

(42) Ta niangzao pijiu, bu he.
he brew beer not drink
'He brews, but doesn't drink beer'

Thus, James Huang's assertion that '[constructions having lexical or smaller-than-lexical categories as conjuncts] must be generated by context-sensitive rules' (1982b:281) is completely groundless.[8]

3.2.3 Unreduced forms

Having seen that reduced forms of the construction may be successfully handled by context-free rules, let us turn to consideration of the unreduced forms, i e those consisting of whole VP's. The obvious course of action is to assimilate A-not-A questions to ordinary verb coordination, again adopting Chu-Ren Huang's arguments to demonstrate that no identity condition need be stated. This would be fine, were it not for Chao's observation (1968:269) on the non-reversibility of A-not-A question forms (see section 2.4 above). The import of this is that, while sentences of the form X-not-X, X-not-Y, and Y-not-X (where $X \neq Y$) are all permissible, sentences of the form not-X-X are not. This opens up the possibility of a demonstration that Chinese is a non-context-free language, the intuitive basis for which is as follows. If we have a rule for VP coordination that does not specify that the right-hand conjunct is to be negated, then the grammar overgenerates producing sentences of the form not-X-X; on the other hand, if we do impose such a condition, then we may not generate the grammatical sentences of the form not-X-Y. In formal terms we may describe the situation thus. Let L be the set of grammatical Chinese sentences, and L_{cf} be that set of sentences generated by free interaction of VP conjunction and negation. L differs from L_{cf} in containing sentences of the type not-X-X. Let L_{cs} be the set of these sentences defining a language which is provably context-sensitive. We now need to know the formal properties of the relative complement of L_{cf} and L_{cs}. At present the result of this operation is unknown (Hopcroft and Ullman, (1969:132)), and I shall therefore assume that the result in question is not provably context-free.

The above demonstration requires a lemma to the effect that the sentences involved may be infinitely long, otherwise we could invoke similar argument to those used in the VP conjunction case above. In other words, we should need to know that A-not-A questions may consist of reduplicated VPs of unbounded length.

However, demonstrating the required premise is not so easy to do. The key to the techniques that might be of help is the concept of recursion. If we can show that, for example, it is possible to have a relative clause in the object NP of an A-not-A question, then we may use the argument, due to Miller and Chomsky (1963:470ff), that any bound on the depth of recursion is arbitrary, and a matter of psychological, rather than linguistic, theory. However, sentences involving recursion were not well received by my informants. Let us look first at the structure of Chinese relative clauses.

Chinese relative clauses are pre-head, consisting of an embedded sentence, followed by the relative marker de.

Thus, the 'John loves Mary' of Chinese relative clauses:

(43) Ni renshi nei ge [$_S$[$_{VP}$ dai yanjing] de] xuesheng
You know that MW wear glasses DE student
'You know the student wearing glasses'

Noun phrases may be modified by several relative clauses, as in (44):

(44) Ni renshi nei ge [$_S$ dai yanjing de],
You know that MW wear glasses DE

[$_S$ bijiao gao de] xuesheng
rather tall DE student

'You know the rather tall student wearing glasses'

However, A-not-A questions involving this sort of recursion were very badly received, and even one level of embedding led to judgements of ungrammaticality, as in (45):

(45) *Ni renshi neige dai yanjing de xuesheng
You know that wear glasses DE student

bu renshi neige dai yanjing de xuesheng?

'Do you know the student wearing glasses?'

Similar reactions were found to A-not-A sentences involving adjectival modification; usually adjectival modification takes the same syntactic form as the relative clauses described above. Thus (46)i) is unacceptable, unlike ii):

(46) i) *Biaozhun fayin hen kequ

ii) Biaozhun de fayin hen kequ
standard DE pronunciation very desirable

'Standard pronunciation is very desirable'

There is a small class of adjectives that may modify nouns directly, but even these were found to be ungrammatical in A-not-A question forms. Hence:

(47) *Ni you jiu shu meiyou jiu shu
you have old book not-have old book
'Do you have old books?'

The above sentences obviously do not provide the basis for a proof of the unbounded length of the VP's involved in A-not-A questions. One further possible source of recursion is the embedding of sentential complements. The sentences in (48) might be claimed to be such examples:

(48) a) Ni renwei wo ben bu renwei wo ben?
you think I/me stupid not think I/me stupid
'Do you think I'm stupid?'

b) Lao Li rang ta qu Shanghai bu
Old Li let he/him go Shanghai not

rang ta qu Shanghai?
let he/him go Shanghai

'Does Li allow him to go to Shanghai?'

Neither of these sentences was well liked by my informants. However, only sentence b) was reckoned ungrammatical, and only then by one informant. One might want to claim that this does not in fact vitiate the claim that, as a stringset, Chinese is context-free, as one may argue that we do not need to analyse such constructions as containing embedded sentences, but rather adopt an approach which assigns the following structure to the reduplicated portions of the above strings:

(49) $[_{VP}$ V NP VP]

This treats rang as the head of the construction, which might not be desirable, if we wish to assimilate it to prepositions. On the other hand, given that varying patterns of control phenomena are found with particular coverbs, we might well wish to have individual rules of the above kind for introducing them (in the manner of Gazdar (1982:148ff)).[9]

The grammaticality of other sentences which might be analysed as having embedded sentential complements has not been tested in this study. A relevant example is:

(50) Ta shuo $[_S$ ta shi Meiguo ren] bu
he say he COP American man not

shuo $[_S$ ta shi Meiguo ren]
say he COP American man

'Does he say he's American?'

My intuition is that this sentence is as problematic as examples (45) and (47) above, although this of course requires verification from a native speaker. If it were found to be the case that such examples are wholly ungrammatical, then we should have no justification for accepting as a premise the idealisation that reduplicated strings in Chinese may be infinitely long or for accepting the thesis that these cases of reduplication necessarily implicate grammars of greater than context-free power. It should be noted that the elimination of recursion from certain portions of the grammar is easy to achieve; the

addition of a feature may prevent a rule that introduces a recursive element from applying.

On the other hand, if a good case can be made for such constructions being potentially unbounded, then this is the point at which it makes sense to invoke Pullum's notion of semantic filtering (Pullum (1984) and section 3.1.2 above). In this case, however, the constraint that forbids a certain set of strings and perhaps renders the stringset non-context-free is presumably more pragmatic than semantic, as it will need to explain how X-not-X is readily interpretable as a question, while not-X-X is necessarily contradictory.

4. Conclusions

The preceding discussions have looked in depth at the problems that the phenomenon of reduplication can cause for linguistic formalisms with restricted expressive power. We have seen under what assumptions the A-not-A construction in Chinese may be dealt with by such formalisms. In particular, we have been able to prove that certain forms of the construction (and some similar phenomena) are provably tractable within the framework of CFPSG's despite previous assertions to the contrary.

The main drawback of the treatment offered above is the proliferation of features. It is not very appealing linguistically to have to add a new feature value each time we add a word to the grammar. One improvement that could be made would treat the reduced form of the construction as involving some sort of lexical reduplication (cf note 4). That's to say, there will be a rule in the lexicon stating that, for every verb X, there is a syntactically identical verb X-bu-X, which has a different semantics. This move will not take us out of the realms of context-freeness, as it can be viewed as a purely local constraint, and has the following advantages. Firstly, it copes more easily with the occurrence of 'ionized' morphemes, and is presumably more adaptable to statements of lexical idiosyncracies in non-verbal classes, than would be a structurally based approach. Secondly, it might form the basis of an explanation for the ungrammaticality of conjoined non-identical verbs (cf examples (22) and (23)).

FOOTNOTES

* I should like to express thanks to Ewan Klein and Mark Steedman for their comments and criticisms on this paper, and to Chen Xiaoying and Zhu Shensheng for providing the data. All errors and omissions are of course my own. The work reported herein was supported by a Science and Engineering Research Council Advanced Course Studentship. This paper is a revision of my MSc thesis (Calder (1984)).

1 Taxonomic and transformational analyses are given by Simon (1958) and Wang (1964, 1965) respectively. Wang (1964, 1965) opts unsurprisingly for a conjunction reduction analysis, as does Rand (1969). Li and Thompson (1981, 18.3.2) offer examples, with emphasis on the pragmatics of their use. Huang (1982a, 1982b) gives analyses of A-not-A and related constructions within the theory of GB.

2 In classical Chinese, the morphemes, xi and huan, were both free, and very close in meaning. As the northern dialects lost syllable-final stop consonants, and with the concomitant threat of a very large number of homophonic clashes, the fusing of near synonyms to form a bisyllabic word was a very common diachronic process. There are other cases in which one might wish to argue that the combination is based on a 'specific + general' verbal structure; thus niangzao, meaning 'brew', consists of the syllable niang which was a free morpheme in classical Chinese, with a meaning very similar to the modern compound, and the syllable zao, a verb with the very general meaning of 'make'. Niang, unlike either syllable of xihuan is still 'semi-free', occurring in the fused verb-object compound

i) niang jiu
brew alcoholic-drink
'brew (beer)' or 'make wine'

Most modern bisyllabic words have histories similar to those of xihuan and niangzao, Li and Thompson (1981:14).

3 Often termed 'coverbs' in Chinese linguistics. Prepositions or coverbs have their historical origin in verbs and the occurrences of gei in (2) above and (10) below represent the same etymological item. Li and Thompson (1973) claim that coverbs are 'syntactically and semantically [...] simply prepositions'. In this essay the terms 'preposition' and 'coverb' are used interchangeably.

4 I shall assume that the phenomenon of lexical reduplication is not directly implicated here. There are good reasons for so doing. Firstly lexical reduplication does not create strings greater than four syllables long. (Cf Chao (1968:198ff) and Lu (1980:637ff) for general discussion.) Secondly the processes involved in lexical reduplication are clearly not of a unitary nature, as they differ according to part of speech, and are highly idiosyncratic in their applicability.

5 Mai in this sentence does mean 'sell'. It is only tonally distinct from the verb meaning 'buy' that we have already come across.

6 There is one linguistically warranted assumption that must be made, namely that there is a finite number of subcategorization frames. That done, the arithmetic is simple. Chinese has a finite number of syllables, the figure being roughly 2200; verbs in Chinese are either mono- or bisyllabic and the verbal complex may only consist of a verb and a finite number of 'directional complements', the latter drawn from a closed class. The only other source of variation is the possibility of verbs having multiple subcategorization frames, and this variation under the above assumption is limited. As all of the terms in the calculation are finite and non-zero, the total figure for distinct Chinese verbs must itself be finite. The figure is of course large, of the order of one per few hundred Mandarin speakers, but that does not compromise the argument.

7 Alternatively, we might view the subscript i as representing the value of a feature, say [LEXEME i]; each distinct verb is associated with a unique integer and a verb may only be introduced under the category symbol V[LEXEME n] if that verb is associated with the integer n. This would permit the use of a single rule for verb conjunction, but of course leaves us with a feature that has a very large number of values.

8 The argument above also goes through for other constructions that appear to require identity between syllables, such as the lian and the 'concessive' constructions, should we wish to use it (cf DeFrancis (1966:256) and Chao (1968:693). My attention was drawn to these constructions by Steve Harlow). The same is true for the forms of the A-not-A construction that are found with 'potential resultative constructions', as described in eg Tewksbury and Fenn (1967:164, example A1) and Cartier (1972).

9 The difference is clearest in the case of the coverbs ba and bei:

i) Lisi ba Zhangsan da le
Lisi CV Zhangsan hit ASP
'Lisi hit Zhangsan'

ii) Lisi bei Zhangsan da le
Lisi CV Zhangsan hit ASP
'Lisi was hit by Zhangsan'

While these two sentences share the same constituents, a typical GPSG analysis would assign different translation rules to them. In both cases, the object is displaced from its canonical position after the verb, and so some mechanism is necessary to ensure that the arguments to the verb are correctly distributed. In the case of ii), a treatment like that offered by Gazdar (1982:161f) for dealing with English passives is implicated; using lambda abstraction over an NP-type variable inserted into the VP, we may obtain a function which takes the translation of the grammatical subject NP and returns a truth value, with the NP in question properly construed as the object of the verb.

REFERENCES

Aho, A V and J D Ullman (1972) The theory of Parsing, Translation, and Compiling, Volume 1: Parsing. Englewood Cliffs: Prentice-Hall.

Bresnan, J W, R Kaplan, S Peters and A Zaenen (1982). 'Cross-serial dependencies in Dutch', Linguistic Inquiry, 13.

Calder, J H R (1984) The problem of reduplication in restrictive theories of grammar. Unpublished MSc thesis, University of Edinburgh.

Cartier, A (1972) Les verbes résultatifs en Chinois Moderne. Paris: Librairie Klincsieck.

Chao Yuen Ren (1968) A Grammar of Modern Chinese. Berkeley: University of California.

Chao Yuen Ren (1976) Aspects of Chinese Sociolinguistics. Stanford: Stanford University Press.

Fenn, H C and M G Tewksbury (1967) Speak Mandarin. Yale.

Gazdar, G J M (1982) 'Phrase structure grammar', in P Jacobson and G K Pullum (eds) The Nature of Syntactic Representation. Dordrecht: D Reidel.

Gazdar, G J M, E H Klein, G K Pullum and I A Sag (1985) Generalized Phrase Structure Grammar. Oxford: Blackwell.

Gazdar, G J M and G K Pullum (1982) Generalized Phrase Structure Grammar: a theoretical synopsis. Bloomington: Indiana University Linguistics Club.

Higginbotham, J R (1984) 'English is not a context-free language', Linguistic Inquiry, 15, 225-234.

Hopcroft, J E and J D Ullman (1969) Formal Languages and their Relation to Automata. Reading, Mass: Addison-Wesley.

Hopcroft, J E and J D Ullman (1979) Introduction to automata theory, languages and computation. Reading, Mass: Addison-Wesley.

Huang, Chu-Ren (1983) Reduplication Reanalysed: a fragment of Chinese PS rules. Cornell University, ms.

Huang, J C-T (1982a) 'Move wh in a language without wh movement', The Linguistic Review, 1, No.4, 369-416.

Huang, J C-T (1982b) Logical Relations in Chinese and the Theory of Grammar. PhD dissertation, MIT Cambridge.

Huybregts, M A C (1976) 'Overlapping dependencies in Dutch', Utrecht Working Papers in Linguistics, 1, 24-65.

Li, C N and S A Thompson (1973) 'Coverbs in Mandarin Chinese: verbs or prepositions?', Journal of Chinese Linguistics, 2.3.

Li, C N and S A Thompson (1981) A functional reference grammar of Mandarin Chinese. Berkeley: UCal Press.

Lu Shuxiang (1980) Xiandai Hanyu Babai Ci. Peking: Shangwu Yinshuguan.

Miller, G A and N A Chomsky (1963) 'Finitary models of language users', in Luce, Bush and Galanter (eds) Handbook of Mathematical Psychology. New York: John Wiley.

Pullum, G K (1984) 'Syntactic and semantic parsability', Proceedings of the 22nd Annual Meeting for the Association of Computational Linguists, Coling84, 112-122.

Pullum, G K and Gazdar, G J M (1982) 'Natural Languages and Context-Free Languages', Linguistics and Philosophy, 4, 471-504.

Rand, E (1969) The Syntax of Mandarin Imperatives. Berkeley: University of California Press.

Simon, H F (1958) 'Some remarks on the structure of the verbal complex in standard Chinese', Bulletin of the School of Oriental and African Studies 21.

Wang, W S-Y (1964) 'Some syntactic rules for Mandarin', Proceedings of the IXth Congress of Linguists.

Wang, W S-Y (1965) 'Two aspect markers in Mandarin', Language, 41.3.

Zhang, Shengyu (1979) 'Chaoyang Fangyan de Chongdie shi (Reduplicative forms in the Chaoyang dialect)', Zhongguo Yuwen, 2, 106-114