DANIEL RADZINSKI

UNBOUNDED SYNTACTIC COPYING IN MANDARIN CHINESE*

0. INTRODUCTION

Ever since Pullum and Gazdar (1982) rebutted earlier claims which argued against the non-context-freeness of natural language on grounds of WEAK GENERATIVE CAPACITY (henceforth wgc), i.e., the capacity to generate stringsets, various new arguments have been published exemplifying non-CFness in NLs. Among those generally assumed to be valid are Huybregts (1984) and Shieber (1985) based on Swiss–German unbounded cross-serial dependencies; Kac (1987) and Kac et al. (1987) based on English *respectively* constructions; and Culy (1985) based on unbounded copying in the vocabulary of Bambara. Additional arguments include those of Bresnan et al. (1982) on grounds of STRONG GENERATIVE CAPACITY (henceforth sgc), i.e., the capacity to generate treesets, and Manaster–Ramer (1987) on grounds of CLASSIFICATORY CAPACITY (henceforth cc), i.e. the capacity to distinguish among sublanguages, both based on Dutch unbounded cross-serial dependencies.

The present paper argues for the classificatory non-CFness of Mandarin Chinese (henceforth MC), the most widely natively-spoken NL, on the basis of a very central and highly productive construction of yes–no questions which have come to be known in the linguistic literature as 'A-not-A' questions. This type of yes–no question may exhibit an unbounded syntactic copying dependency. Huang (1982, p. 281) alluded to the fact that A-not-A questions present difficulties for a CF account of MC. He wrote that if some A-not-A question constructions are base-generated, then "they must be generated by context sensitive rules. . . . But ordinary base rules (in so far as one assumes that they exist) are usually context-free". Nevertheless, no formal argument to this effect was offered. In this paper we attempt precisely that.

* I am grateful to Richard Alpert, Gidi Avrahami, Chris Culy, Joyce Friedman, Jack Hoeksema, Jim Huang, Aravind Joshi, Ross King, Susumu Kuno, Alexis Manaster-Ramer, Philip Miller, Geoff Pullum, Bill Rounds, Walt Savitch, Stu Shieber, Fu Tan, Tom Wasow, John Whitman, Yuru Wu, Yuli Zhou, and two anonymous referees for their many suggestive insights, judgements, and comments on earlier versions of this paper. All errors are, of course, my sole responsibility. Many of the central points in this paper, particularly those of Sections 1–3, are discussed also in Radzinski (1988).

Linguistics and Philosophy **13**: 113–127, 1990.

Section 1 focuses on the relevant data from MC. A formal proof which will be used in later argumentation is presented in Section 2. Section 3 consists of the main argumentation for demonstrating the classificatory non-CFness of MC. It examines implications from wgc, sgc and cc. Section 4 discusses and refutes some possible linguistically grounded arguments which might be raised in order to nullify the results obtained from the previous sections. Finally, Section 5 presents the conclusions of this study.

1. DATA

MC commonly employs two types of yes–no questions. One type, known as Particle Questions, consists of an interrogative particle, typically *ma*, appended to the end of a declarative sentence. The other type, known as A-Not-A Questions, consists, in its most basic form, of appending a subjectless negative declarative sentence to its corresponding full non-negative counterpart. We shall deal below only with the latter type.[1]

A typical A-not-A question is example 3 below. Example (1) is the simple declarative form for "S/he is at home", while example (2) is its simple negative counterpart. The interrogative (3) is formed by appending all but the subject of (2) to the end of (1):[2]

(1) ta zai jia
s/he at home
S/he is at home.

(2) ta bu zai jia
s/he not at home
S/he is not at home.

(3) ta zai jia bu zai jia?
s/he at home not at home
Is s/he at home?

[1] The distribution of these two types of yes-no questions seems to be guided by pragmatic principles. On this matter, see Li and Thompson (1979) and Section 18.6 of Li and Thompson (1981). Examples 1–12 below are taken from Li and Thompson (1981, pp. 535–538). Acceptability judgements for all the data presented in this paper have been verified with native speakers.

[2] It is always possible to optionally insert the disjunction *haishi* – 'or' immediately before the negative marker *bu*. However, since the presence of this item is optional, all of our examples will lack *haishi*.

It is not essential, however, to have the entire verb phrase repeated in both clauses. Thus, in example (4) below the first VP consists only of the verb *zai*, while in example (5) the second VP comprises only this same verb:

(4) ta zai bu zai jia?
s/he at not at home
Is s/he at home?

(5) ta zai jia bu zai?
s/he at home not at
Is s/he at home?

Nevertheless, it is not possible to freely delete just *any part* of the VP. As Li and Thompson (1981, p. 536) mention: "In general, elements forming a semantic unit must be deleted together". Thus examples (7)–(10) below, which are reduced variants of example (6), are all unacceptable A-not-A questions:

(6) ni xihuan ta -de chenshan bu xihuan ta
you like s/he GEN shirt not like s/he
de chenshan?
GEN shirt
Do you like his/her shirt?

(7) *ni xihuan ta -de bu xihuan ta
you like s/he GEN not like s/he
-de chenshan?
GEN shirt

(8) *ni xihuan ta -de chenshan bu xihuan
you like s/he GEN shirt not like
ta -de?
s/he GEN

(9) *ni xihuan chenshan bu xihuan ta
you like shirt not like s/he
-de chenshan?
GEN shirt

(10) *ni xihuan ta -de chenshan bu
you like s/he GEN shirt not
xihuan chenshan?
like shirt

Example (7) results from deleting the noun *chenshan* in the first clause, without deleting the other elements of the noun phrase, namely, the pronominal *ta* and the genitive marker *de*. Example (8) results from the same deletion pattern as in example (7), but within the second clause. Examples (9) and (10) are symmetrical to (7) and (8), respectively, with *ta* and *de* deleted instead of *chenshan*. If, however, an entire object NP is deleted, as in (11) and (12) below, then the resulting question is perfectly acceptable:[3]

(11) ni xihuan bu xihuan ta -de chenshan?
you like not like s/he GEN shirt
Do you like his/her shirt?

(12) ni xihuan ta -de chenshan bu xihuan?
you like s/he GEN shirt not like
Do you like his/her shirt?

Although A-not-A questions lacking fully repeated VPs are used more commonly in MC than those with full repetitions, they do complicate somewhat the search for a construction necessary to raise a non-CFness argument for MC. Therefore, I shall deal hereunder only with A-not-A questions which contain fully repeated VPs.

Generally, for a non-context-freeness argument based on a copying constraint to hold, it is essential that the copied string be unbounded in length. That is, our argument cannot be based on the mere fact that some

---

[3] When the entire object noun phrase of the first clause is deleted, then it is even possible for the deletion to include the second syllable of the verb, as in the following example:

ni xi- bu xihuan ta- de chenshan?
you like- not like s/he GEN shirt
Do you like his/her shirt?

Thus, the deletion might even extend into the lexical, or perhaps phonological, level. Nevertheless, these cases need not distract us from our present purpose. It is worth mentioning that throughout this paper, I am using the terms deletion and insertion only in their pretheoretical and descriptive senses. There is no bias towards any of the many theories that attempt to offer explanatory accounts for the phenomena discussed herein.

finite string is copied within a sentence. This can be handled even by a regular grammar, i.e., one that generates a language that can be accepted by a finite-state automaton (the weakest known device for accepting non-finite languages). It is cardinal that the length of this string be arbitrarily great.[4] Furthermore, it is of equal importance that the identically copied string contain repetitions of at least two distinct items in an arbitrary, yet precise, linear order. Otherwise, a CF grammar could easily handle an identically copied string consisting of only one multiply repeated item.[5] Let us attempt therefore to find such a construction.

The most obvious candidate for a string lacking an upper bound on its length within a sentence would be a noun phrase. It is clear that an NP may include an arbitrary large number of relative clauses. Likewise, an NP may consist of an arbitrary large concatenation of smaller conjoined or disjoined NPs. Also, an NP may include an unbounded number of adjectives, some repeated perhaps, which modify its head noun. Let us, therefore, construct some A-not-A questions in MC containing a repeated NP which has one of the aforementioned properties resulting in great arbitrary length, say that of having an unbounded number of adjectives modifying its head noun. Examples (13) and (14) below, exhibit one and two adjectives, respectively, modifying the head noun of the repeated object NP:

(13) ni xihuan geng-da de pingguo bu xihuan geng-da
you like bigger GEN apple not like bigger
de pingguo?
GEN apple
Would you like a bigger apple?

(14) ni xihuan geng-da geng-hao de pingguo bu xihuan
you like bigger nicer GEN apple not like
geng-da geng-hao de pingguo?
bigger nicer GEN apple
Would you like a bigger nicer apple?

---

[4] Nevertheless, claims have recently been made contesting the importance of unbounded length as regards generative capacity arguments. See Rounds et al. (1987) and Savitch (1989).

[5] For elaborations and proofs of the well-known formal claims discussed in this paragraph and throughout this paper, see any introductory textbook on formal languages and automata theory. E.g., Hopcroft and Ullman (1979), Lewis and Papadimitriou (1981), or Salomaa (1973), among others.

Notice from the unacceptability of A-not-A question (15) below, that the NP in the second clause must contain its adjectives in the same exact linear order as they appear in the first clause:

(15) *ni xihuan geng-da geng-hao de pingguo bu xihuan
you like bigger nicer GEN apple not like
geng-hao geng-da de pingguo?
nicer bigger GEN apple

In (15), the repeated string inverts the linear order of its two adjectives, yielding *geng-hao geng-da* rather than *geng-da geng-hao*, as in the first clause. Moreover, as already noticed from examples (7)–(10), it is necessary for the entire NP to be copied, if it is to be copied at all. Thus, if we add one more adjective to the first clause, as in examples (16)–(18) below, we see that *all* the adjectives must be copied in their entirety and in the same linear order:

(16) ni xihuan geng-da geng-hao geng-da de pingguo
you like bigger nicer bigger GEN apple
bu xihuan geng-da geng-hao geng-da de pingguo?
not like bigger nicer bigger GEN apple
Would you like a bigger nicer bigger apple?

(17) *ni xihuan geng-da geng-hao geng-da de pingguo bu
you like bigger nicer bigger GEN apple not
xihuan geng-da de pingguo?
like bigger GEN apple

(18) *ni xihuan geng-da geng-hao geng-da de pingguo bu
you like bigger nicer bigger GEN apple not
xihuan geng-da geng-da geng-hao *de* pingguo?
like bigger bigger nicer GEN apple

Clearly, the unbounded addition of any more adjectives, whether repeated or new, will preserve the properties maintained by the construction containing two or three adjectives, namely, all the adjectives must be copied in their entirety and in the same linear order which appears in the first clause. These facts suffice for us to begin developing a non-CFness argument for MC.

2. Proof

In this section we prove a wgc result which will be of help for the argumentation of Section 3. Let us construct the following regular language R:

R = {ni xihuan geng-da *x* de pingguo bu xihuan geng-da *y* de pingguo?
| *x*, *y* are in (geng-da + geng-hao)*}

This artificially constructed language R comprises all sentences which begin with "ni xihuan geng-da", followed by zero or more instances of either 'geng-da' or 'geng-hao', followed by "de pingguo bu xihuan geng-da", followed by zero or more instances of either 'geng-da' or 'geng-hao', and terminated by the string 'de pingguo?'. This language is regular, i.e., it can be accepted by a finite-state automaton, and it does not require any constraint of identity between its possibly repeated elements. Thus, *x* and *y* need not be identical. Now, let us intersect this regular language R with the language Boolean Chinese (henceforth BC) which comprises all and only the strings which qualify in MC as A-not-A questions. The result of this intersection is the following language L:

L = {ni xihuan geng-da *x* de pingguo bu xihuan geng-da *x* de pingguo?
| *x* is in (geng-da + geng-hao)*}

L seems superficially very similar to R, except for the fact that it contains an additional requirement of identity between the two series of repeated items. That is, instead of independent strings *x* and *y*, L has *x* and *x*. Hence, L is a copying language, sometimes referred to in the literature as a *ww* or *xx* language, since the string *x* is repeated in its exact size and linear order. Copying languages of this sort are not CF. Pushdown automata, which can accept any weakly CFL, cannot both keep count of the size of *x* and the precise linear order of its elements, which are more than one item (*geng-da* and *geng-hao*). L is therefore non-CF. Yet, the intersection of a regular language, such as R, with a CFL *always* yields a CFL due to well-known closure properties of CFLs. Since L is not weakly CF, then neither is BC, else L would also have to be CF. Thus, BC is not weakly CF.

3. Argument

In the previous section, we proved that BC is not weakly CF. However, if we wish to make a claim about its proper superset MC then we must

proceed one step further. We cannot simply maintain that MC is not weakly CF because it contains a non-CF proper subset BC. This would surely be a fallacy. For example, the non-CFL $\{xx \mid x \text{ is in } (a+b)^*\}$ is a proper subset of the language $\{xy \mid x, y \text{ are in } (a+b)^*\}$, which is regular, a fortiori CF. In order to prove that MC is weakly non-CF using an intersection, we must intersect a regular language with the whole of MC (or at least with the language generated by the syntax of MC). If the result of intersecting R with MC yields L, then we will have proved that MC is weakly non-CF.

Regretably, this is not the case. The intersection of R with MC yields R itself, a regular language, hence a CFL. This is due to the following fact about MC: any two VPs can be juxtaposed so as to form a coordinate phrase.[6] Thus, all strings which are in R, including examples (15), (17) and (18), are also in MC, though not as A-not-A questions but rather as contrastive-like coordinations. (In other words, these sentences are aptly starred only under an A-not-A interrogative interpretation). Example (17), for instance, can be interpreted as something, albeit odd but clearly grammatical, akin to "You like a bigger, nicer, bigger apple and/but don't like a bigger apple". Therefore, even though these sentences are not acceptable A-not-A questions, they qualify as grammatical strings in MC. Hence for MC as a whole, there is no copying constraint; substrings x and y of R do not have to be equal in content, linear order and length in order to be also in MC, even though they have to meet such a requirement in order for their superstring to be interpreted as an A-not-A question in MC. This, however, completely undermines any possibility of claiming that MC is weakly non-CF on the basis of Boolean A-not-A constructions.[7]

Having failed to demonstrate the non-CFness of MC on wgc grounds, we shall now move on and present an argument based on sgc. Namely, we can argue that although all the strings of MC might still be generated by a CF grammar, no CFG can generate their correct structural descriptions, i.e., trees. Bresnan et al. (1982) follow precisely this line of argument for Dutch claiming that due to its cross-serial dependencies, Dutch

---

[6] This most crucial piece of information has been pointed out to me by Alexis Manaster-Ramer as well as by two anonymous referees.

[7] It remains to be determined whether taking intonational information into account would permit here the formulation of a non-CFness argument on grounds of wgc. Encoding intonational information into the strings of the regular set might lead to an overt, i.e., string-based, differentiation between A-not-A cases and others. Likewise, orthographic information may also play a similar role. However, such intonational and orthographic distinctions would appear to be most likely optional and, hence, of little value in establishing a wgc proof based on the closure of CFLs under intersection with RLs.

cannot be described by a CFG in a fashion that would assign the correct structural descriptions to the terminal strings of the language. Under the reasonable assumption that the assignment of correct structural descriptions to MC must be part of any grammar of MC, A-not-A questions should then, as intuitively expected, have structural representations that differ from those of homophonous strings which are not A-not-A questions, e.g., contrastive coordinations.[8] These structural representations would be a direct reflection of the properties of A-not-A questions. Since A-not-A questions, when viewed as a language (BC), are not weakly CF, as shown in Section 2, then a fortiori they are not strongly CF. Their trees, encoding a copying dependency, cannot be derived solely by CF rules. This would render MC strongly non-CF.

Nevertheless, arguments based on sgc are subject to numerous disadvantages.[9] In particular, what *are* the 'correct' structural descriptions for A-not-A questions? 'Correct' structural descriptions tend to be theory-dependent. Although most linguistic theories would probably assign trees of non-CF nature to A-not-A questions, some theory might somehow manage to be adequate and still assign trees to that construction that are of a strictly CF nature. It is, again, theory-dependent. This leaves a lot to be desired of sgc arguments.

We now move on to consider cc. Here the situation appears to be far

---

[8] An anonymous referee has suggested that the difference between an instance of an A-not-A question and a (homophonous) contrastive coordination may very well be based solely on pragmatic considerations. According to this view, both constructions would have the same syntactic category and constituent stucture and both would have the same denotation of a doubleton set of states-of-affairs. Yet, the pragmatic force associated with each construction would differ: the one for an A-not-A question would be: "I want you to tell me which of these two states-of-affairs is actual", while the one for a contrastive coordination would be: "Here are two actual states-of-affairs: note the contrast".

Though there is little doubt that the pragmatic forces mentioned in this view are indeed the ones associated with the respective constructions and that these constructions do denote some doubleton set of states-of-affairs, this position appears to overlook an important fact: While the two clauses of a contrastive coordination may be any two VPs, the A-not-A ones MUST involve identity (subject to the negation marker *bu* commencing the second VP). This is not a pragmatic fact. There is no pragmatic reason why a force of "I want you to tell me which of these two states-of-affairs is actual" must be limited to complementary states-of-affairs manifested by means of syntactic copying-cum-negation. In other words, there must be something beyond mere pragmatic force that limits A-not-A questions to the structural copying form they exhibit, but does not limit contrastive coordinations to such a form. Thus, the distinction between the two constructions cannot be based solely on pragmatic considerations devoid of any reference to structure. Pragmatic considerations may help in resolving ambiguity, but not in accounting for a copying constraint in a particular construction.

[9] For an elaborate discussion of this issue, see Section 3 of Manaster–Ramer (1987).

more promising than in wgc and sgc. Classificatory capacity is preliminarily defined by Manaster-Ramer (1987, p. 238) as "the measure of a formalism's ability [to] classify a set of strings (and substrings) and specify which ones are like which other ones". More specifically, this is a criterion used in determining whether some grammar (of a particular Chomsky-type, for example) can both generate an object lg. and at the same time characterize the constructions, or sublanguages, of this object lg. In other words, can the grammar generate a lg. in a way that the distinctions among its internal constructions are merely a consequence of the generation itself? A CFG, for example, can do so via its nonterminals: Let some nonterminal in the grammar ultimately be rewritten only as a string belonging to some particular construction. E.g., if passive sentences are to be considered a construction in some NL L, then let the CFG G(L) include a member PASS in its set of nonterminal symbols which yields in one or more steps all and only passive sentences of L. We say then that G(L) *classifies* passives.[10]

What we would wish to test then is whether a CFG for MC can classify A-not-A questions. Namely, can the set of nonterminals of a CFG G(MC) include a symbol ANOTA which yields in one or more steps all and only A-not-A questions? The answer to this last question is clearly in the negative, since the set of strings comprising the construction of A-not-A questions, referred to as the sublanguage BC, has been shown in Section 2 to be non-CF. ANOTA provenly cannot yield all and only A-not-A questions, i.e., the sublanguage BC, since these are weakly non-CF while ANOTA is a nonterminal in a CFG. Thus G(MC), or any other arbitrary CFG for MC, cannot classify A-not-A questions.

An objection then might be raised that perhaps A-not-A questions are not a construction, or sublanguage of MC. But how could A-not-A questions be anything other than a separate construction in MC? Their interpretation as Boolean questions does not strictly follow from the interpretation of their subparts. Furthermore, they differ most intuitively from other coordination-based strings which might happen to be similar to them in their superficial form. Thus, on cc grounds with the minimal consent that A-not-A strings have derivations that differ nontrivially from those

[10] There undoubtedly exist alternative means for defining classification within CFGs. E.g., rather than using a single nonterminal as a basis, classification can be done via a set of nonterminals. Likewise, it can be based on the existence of a production rule that applies in all and only the derivations of a particular construction. However, these would be alternative formalizations of a single intuitive notion that do not differ much in essence. Defining classification in formalisms other than CFG is clearly possible, but beyond the scope of this discussion.

of other constructions and on wgc grounds proving that A-not-A strings as a sublanguage of MC are non-CF, we reach the conclusion that MC is not classificatorily CF.

Notice that the cc argument above is not subject to the disadvantages of sgc arguments. As Manaster-Ramer (1987, p. 241) points out: 'Unlike strong generative capacity, classificatory capacity does not involve the specific derivations (structures) considered correct by a particular linguistic school, but merely the sameness or difference of derivations (structures) of particular sublanguages". Indeed, we have not assumed here any specific structure. All we have done is consider A-not-A questions to be a construction in MC and this construction was previously shown to be non-CF on wgc grounds. This renders the MC language non-CF on cc grounds because a CFG will not be capable of classifying A-not-A questions, i.e., generating them as a separate subcomponent of MC (by means of some unique nonterminal) distinct from other constructions in the language.

To conclude this section, we summarize its findings. Although it was proved in Section 2 that MC's A-not-A questions viewed as a formal language BC are not weakly CF, MC as a whole remains weakly CF. If we assume reasonable and intuitive tree-structures for A-not-A questions, we can argue that MC is not strongly CF. However, these structures would still need to be theory-dependent, allowing for the formulation of numerous alternative tree-structures, some which might retain strong CFness. Considering A-not-A questions a construction within MC, as any serious grammar would be expected to do, together with the result from Section 2 regarding the weak non-CFness of these questions, yields a classificatory capacity result that MC, as a whole, is not CF, since no CFG can classify A-not-A questions as a separate sublanguage of MC.[11]

4. Discussion

Let us now consider some possible empirically-based linguistic arguments which might be raised in order to refute our conclusion that MC is not classificatorily CF. One might argue that the data, which crucially bear on the non-CFness argument, are evidently correct, yet acceptability judgements do not coincide with grammaticality judgements. In other

[11] The fact that some linguistic theories, such as those rooted in a Chomskyan principles and parameters approach, view constructions as epiphenomena does not affect this last argument. Any theory of grammar will need to account somehow for the essential difference between A-not-A questions, standard coordination, contrastive coordination, etc., in MC no matter what status it gives to constructions, whether primitive or not.

words, sentences which are ruled out as A-not-A questions are done so for some extra-syntactic reason, making it hard to reach any conclusions about the syntactic component of MC. Assuming this to be true, what then could this "extra-syntactic reason" possibly be? Let us list the options:

(i) *Lexical (or Morphological)*: This seems to be impossible, since we are crucially dealing with the repetition of full NPs (not to mention VPs) which, almost by definition, are formed in the syntax, rather than smaller nouns, for example, which are formed in the lexicon.

(ii) *Discourse-based*: In order to seriously entertain such a possibility, one would have to show that the conditions governing the acceptability of an A-not-A question, or lack thereof, are of an extra-sentential nature. This, however, seems impossible, since these A-not-A constructions can only be sentential. Any copying outside of a sentential boundary does not constitute an A-not-A question, nor any other type of yes–no question. Hence, such copying would be merely optional and completely irrelevant to our issue at hand.

(iii) *Semantic*: This would mean that we are dealing with semantic copying, rather than syntactic copying. Nevertheless, such a conjecture would fail the test of empirical adequacy. This conjecture would predict that the second clause of an A-not-A question could contain an element not lexically identical to its corresponding element in the first clause, but rather merely synonymous to it (cf. English *fat* vs. *obese*). It appears to be the case, however, that this is not so. Examples (19) and (21) below, which involve 'strict' verb copying are clearly acceptable A-not-A questions, while this is not the case concerning their respective 'sloppy' counterparts 20 and 22:

(19) ni ai ta bu ai ta?
you love s/he not love s/he
Do you love him/her?

(20) *ni ai ta bu xihuan ta?
you love s/he not like s/he

(21) ni zhidao ta bu zhidao ta?
you *know s/he not know s/he*
Do you know him/her?

(22) *ni zhidao ta bu renshi ta?
you know s/he not recognize s/he

Likewise, substituting an overt pronoun for a full NP, even under the intention of a coreferential interpretation, yields an unacceptable A-not-A question. Strict string copying must take place, as in (23), and not referential identity copying, as in (24), below:

(23) ni ai Zhangsan bu ai Zhangsan?
you love Zhangsan not love Zhangsan
Do you love Zhangsan?

(24) *ni ai Zhangsan bu ai ta?
you love Zhangsan not love him

Nevertheless, one might argue that synonymous verbs, as well as overt pronouns and their antecedents, are never genuine cases of *full* semantic identity, therefore their use in A-not-A constructions yields unacceptable judgements. A requirement for full semantic identity would therefore automatically account for the unacceptability of examples (15) and (18), since the representation of adjectives in variant linear order, might affect the width of scope for each adjective in the NP. Again, this argument seems unfounded, since when we construct an A-not-A question involving, for instance, the copying of a conjoined non-temporal and non-quantified NP, we still require strict string copying. Thus, in example (25) below, 'melons and apples' is substituted for 'apples and melons'', yet yielding an unacceptable result if interpreted as an A-not-A question (cf. example (26)):

(25) *ni xihuan pingguo gua bu xihuan gua pingguo?
you like apple melon not like melon apple

(26) ni xihuan pingguo gua bu xihuan pingguo gua?
you like apple melon not like apple melon
Do you like apples and melons?

But generally 'apples and melons' and 'melons and apples', with no quantifiers attached, do not differ in meaning. So here again, it is not semantic identity that is required for repetition, but rather pure syntactic string copying.

Finally, it might be argued that since A-not-A questions actually present a VP followed by its negation, whatever copying there may be is simply a consequence of encoding the negation of a VP by means of *bu* preceding an instance of that same VP. Namely, rather than there being a syntactic

copying constraint, there is a semantic constraint that requires the negation of a preceding predicate. Since such negation is manifested via a VP followed by the negation marker *bu*, which in turn is followed by another-instance of the same VP, we indeed end up with a copied VP, but merely as a consequence of the negation and not due to a copying constraint. Yet, if what is at stake here is merely the negation of some predicate, then we would expect a pair of antonymous verbs to be able to do the same job generally done by two instances of the same verb. In other words, rather than having some verb *v* followed by *bu* followed by *v* again, we would have *v* followed by a verb *u*, which is an antonym of *v*, and *u* and *v* have no common morpheme. However, such a structure cannot be an A-not-A question. While example (27) below is a perfectly fine A-not-A question, example (28) is not:

(27) men kai bu kai?
door open not open
Is the door open?

(28) *men kai guan?
door open closed

Example (28) cannot be interpreted as "Is the door open?", even though this would be expected if semantic negation were the essential characteristic of the construction (since 'open' and 'closed' are antonyms). Therefore, semantic negation cannot be what is required here, but rather syntactic copying, as claimed previously.

5. Conclusions

While most claims against the context-freeness of natural language tend to be based on constructions which are generally marginal in their nature, this paper has argued that Mandarin Chinese is not CF on the basis of a very central and highly productive construction of A-not-A questions. MC is currently the most widely natively-spoken NL and yes–no questions, such as MC's A-not-A type, could be anything but marginal or peripheral within an NL.[12] The argument made use of a weak generative capacity proof that A-not-A questions, when viewed as a sublanguage, are not

[12] It is interesting to note that the first conclusion of Moravcsik (1971, p. 180), a study within a relatively abstract contrastive framework, states that "Yes-no questions have a number of semantic and syntactic properties which appear [universally DR] explainable by an underlying structure which includes a disjunction of an affirmative sentence and its negative counterpart".

CF. It then combined this formal result with a classificatory capacity argument in order to extend the claim of non-CFness to the superlanguage MC. The assumption that non-CFness is a non-central phenomenon in NL appears now to be far less plausible.

REFERENCES

Bresnan, J., R. Kaplan, S. Peters and A. Zaenen: 1982, 'Cross-serial Dependencies in Dutch', *Linguistic Inquiry* **13**, 613–635.

Culy, C.: 1985, 'The Complexity of the Vocabulary of Bambara', *Linguistics and Philosophy* **8**, 345–351.

Hopcroft, J. and J. Ullman: 1979, *Introduction to Automata Theory, Languages, and Computation*, Addison-Wesley, Reading, MA.

Huang, C-T. J.: 1982, *Logical Relations in Chinese and the Theory of Grammar*, PhD Thesis, MIT.

Huybregts, R.: 1984, 'The Weak Inadequacy of Context-Free Phrase Structure Grammars', in G. de Haan, M. Trommelen and W. Zonneveld (eds.), *Van Periferie naar Kern*, pp. 81–99, Foris, Dordrecht.

Kac, M.: 1987, 'Surface Transitivity, *Respectively* Coordination and Context-Freeness', *Natural Language and Linguistic Theory* **5**, 441–452.

Kac, M., A. Manaster-Ramer and W. Rounds: 1987, 'Simultaneous-Distributive Coordination and Context-Freeness', *Computational Linguistics* **13**, 25–30.

Lewis, H. and C. Papadimitriou: 1981, *Elements of the Theory of Computation*, Prentice-Hall, Englewood Cliffs, NJ.

Li, C. and S. Thompson: 1979, 'The Pragmatics of Two Types of Yes-No Questions in Mandarin and its Universal Implications', *CLS* **15**, 197–206.

Li, C. and S. Thompson: 1981, *Mandarin Chinese: A Functional Reference Grammar*, University of California Press, Berkeley, CA.

Manaster-Ramer, A.: 1987, 'Dutch as a Formal Language', *Linguistics and Philosophy* **10**, 221–246.

Moravcsik, E.: 1971, 'Some Cross-Linguistic Generalizations about Yes-No Questions and their Answers', in *Working Papers on Language Universals* **7**, 45–193, Stanford University.

Pullum, G. and G. Gazdar: 1982, 'Natural Languages and Context-Free Languages', *Linguistics and Philosophy* **4**, 471–504.

Radzinski, D.: 1988, 'Mandarin Chinese and Context-Freeness', in *Proceedings of the International Computer Science Conference '88 Artificial Intelligence: Theory and Applications*, pp. 335–342, Hong-Kong.

Rounds, W., A. Manaster-Ramer, and J. Friedman: 1987, 'Finding Natural Languages a Home in Formal Language Theory', in A. Manaster-Ramer (ed.), *Mathematics of Language*, pp. 349–359, John Benjamins, Amsterdam.

Salomaa, A.: 1973, *Formal Languages*, Academic Press, NY.

Savitch, W.: 1989, 'Infinity is in the Eye of the Beholder', Technical Report #C589–152, UCSD. (to appear in C. Georgopolous and R. Ishihara (eds.), *Interdisciplinary Approaches to Language: Essays in honor of S.-Y. Kuroda*, Kluwer, Dordrecht).

Shieber, S.: 1985, 'Evidence against the Context-Freeness of Natural Language', *Linguistics and Philosophy* **8**, 333–343.

Dept. of Linguistics
Harvard University
Grays Hall Basement
Cambridge, MA 02138
U.S.A.