ELSEVIER

Cognitive Science 28 (2004) 699–720

COGNITIVE SCIENCE

http://www.elsevier.com/locate/cogsci

Varieties of crossing dependencies: structure dependence and mild context sensitivity

Edward P. Stabler

UCLA Department of Linguistics, University of California, 3125 Campbell Hall, P.O. Box 951543, Los Angeles, CA 90095-1542, USA

Received 14 July 2003; received in revised form 25 April 2004; accepted 16 May 2004

Abstract

Four different kinds of grammars that can define crossing dependencies in human language are compared here: (i) *context sensitive rewrite* grammars with rules that depend on context, (ii) *matching* grammars with constraints that filter the generative structure of the language, (iii) *copying* grammars which can copy structures of unbounded size, and (iv) *generating* grammars in which crossing dependencies are generated from a finite lexical basis. Context sensitive rewrite grammars are syntactically, semantically and computationally unattractive. Generating grammars have a collection of nice properties that ensure they define only "mildly context sensitive" languages, and Joshi has proposed that human languages have those properties too. But for certain distinctive kinds of crossing dependencies in human languages, copying or matching analyses predominate. Some results relevant to the viability of mildly context sensitive analyses and some open questions are reviewed.

Keywords: Sensitivity; Goal; Language

1. Introduction

One reason for seeking universal properties of human languages is to make sense of the fact that these languages can be learned from the kind of data regularly available to children. With this goal, it is particularly valuable to identify universal properties which can be stated in terms that are plausibly available to the language learner *before the language is learned.*

E-mail address: stabler@ucla.edu (E.P. Stabler).

0364-0213/$ – see front matter
doi:10.1016/j.cogsci.2004.05.002

For example, before a language learner has the notion of a "noun phrase," a "subject," or a "phase," it is hard to see how proposals like the following could be useful:

(1) Extractions from complex noun phrases are blocked. (Ross, 1967)
(2) Subjects can be null unless positive evidence indicates otherwise. (Hyams, 1986; Rizzi, 2000)
(3) A spelled-out phase is inaccessible to further form-altering operations. (Chomsky, 2001a, 2001b)

This point has been noted many times before.[1] There are some universal regularities that do not rely on obscure notions like "noun phrase" or "subject" or "phase," like Zipf's hypothesis about the relation the position of a morpheme in a list from most frequent to least frequent (its "rank") to its frequency in any natural discourse of reasonable length:

(4) An inverse exponential relates rank and frequency of morphemes (Zipf, 1949)

This regularity is based only on the segmentation of the discourse into words or morphemes, but it is also quite uninformative. Very many different kinds of language generators, including the simplest probabilistic monoids, will manifest this property, and so it tells us little about human language or human language learning (Mandelbrot, 1961; Miller & Chomsky, 1963). Here we will consider a different hypothesis:

(MCS hypothesis) Human languages are mildly context sensitive (MCS), (Joshi, 1985)

where a language is MCS if it has (i) limited crossing dependencies, (ii) constant growth, and (iii) polynomial parsing complexity. Of the three parts of the hypothesis, the idea that human languages have tractable, polynomial parsing complexity is probably the most familiar.[2] A language is said to have "constant growth" if there is a bound k, such that whenever two sentences have lengths that differ by more than k, there is a sentence of intermediate length. The intuition here is that sentences are built up by simple combinations of smaller constituents. Having "limited crossing dependencies" is more difficult to define precisely and will be discussed later. Unlike (4) the MCS hypothesis draws a significant structural distinction among languages. There are MCS grammars for context free languages, and for languages with reduplication, but not for all context sensitive languages. For example, the context sensitive language a^{2^n} which has as sentences all sequences of a's with a length that is a (non-negative, integer) power of 2 is not MCS. This is easily seen by noting that it lacks the constant growth property: there is no bound on the difference between the length of one sentence and the next longer one. The MCS hypothesis might seem obvious, or so technical as to be irrelevant to the main concerns of linguists and psycholinguists, but in fact it is right at the center of current controversies. We get a hint of the controversy when we notice that many linguists have adopted grammar frameworks that extend well beyond the MCS languages.[3] But linguistic frameworks differ in so many respects and have so much power, so many degrees of freedom, that it is sometimes tempting to conclude that the differences between them do not really matter either: a good idea in one framework can be translated into any another, can't it? No, this is not the case. There are some real, substantial differences among linguistic proposals, and the MCS hypothesis draws one of the substantial distinctions among them. It is hard to think of any other clear, theory-independent proposal about human language which has this status. This

paper will survey some of the constructions that pose a challenge to the MCS hypothesis and consider where the debate stands today.

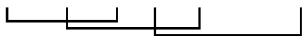
2. Crossing dependencies

It is clear that speakers of human languages are often quite aware of repetition of linguistic expressions of various kinds, and that this can play a role in the grammars of human languages. This involves noticing "crossing dependencies," "crossing correspondences" that cannot be defined with simple context free grammars, and so it is interesting to consider what kinds of mechanisms could capture them. There are various ways to do this, using different sorts of mechanisms, including some that easily extend beyond the MCS limitations. To make the issues clear, we will first provide some very brief historical remarks, and then distinguish four quite different kinds of proposals.

2.1. The historical setting, briefly

If we link each verb with the affix it determines on the following verb in English constructions like the following, it is easy to see that the dependencies are crossing:

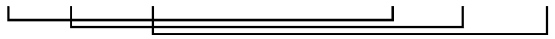
(5) John will have -∅ be -en eat -ing pie



This was the basis for Chomsky's (1956) early claims about the "structure dependence" of human grammars: we would like to have grammars that allow a simple treatment of the discontinuous relations between each auxiliary verb and the affix on the next verb. The extent of dependencies in English auxiliary constructions is bounded, but we can get unbounded crossing dependencies between subjects and verbs in Dutch constructions like the following (Bresnan, Kaplan, Peters, & Zaenen, 1982; Huybregts, 1976):

(6) ... because I Cecilia Henk the hippo saw help feed

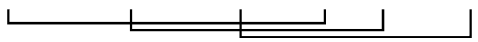
... omdat ik Cecilia Henk de nijlpaarden zag helpen voeren



And in Swiss–German, the relation between verbs and their objects is signaled not only by word order but also by case marking (Shieber, 1985):

(7) ... that we the children Hans the house let help paint

... das mer d'chind em Hans es huus lönd hälfe aastriiche



The linguistic and psychological interest is not simply to generate these word sequences, but to define, as nearly as possible, the very structures recognized by speakers of human languages. We expect that the sequences of words are recognized by their association with syntactic structures that are semantically transparent, structures that include the crossing relations, since these structures provide the basis for interpretation and reasoning about what has been said. The relations indicated by the arcs in (6) and (7) are *semantic* relations, ones that we might expect to be mirrored in the syntax.[4]

2.2. Four grammars for copy dependencies

2.2.1. A context sensitive rewrite grammar

A context sensitive rewrite grammar can be given by a vocabulary V, a set of categories Cat and rules that rewrite strings of vocabulary and category symbols, subject to the requirement that if $s \rightarrow t$ then the length of s is less than or equal to the length of t, except that for a designated "start" category S, the production $S \rightarrow \epsilon$ (where ϵ is the empty string) is allowed if S does not appear on the right side of any rule. Consider the following context sensitive grammar, for example:

$V=\{a, b\}$		$Cat = \{S, L, M, X, Y, Q, R\}$	
$S \rightarrow aSX$	$S \rightarrow bSY$	$S \rightarrow LQ$	$S \rightarrow MR$
$QX \rightarrow XQ$	$RX \rightarrow XR$	$QY \rightarrow YQ$	$RY \rightarrow YR$
$LX \rightarrow LQ$	$MX \rightarrow MQ$	$LY \rightarrow LR$	$MY \rightarrow MR$
$L \rightarrow a$	$M \rightarrow b$	$Q \rightarrow a$	$R \rightarrow b$

Mateescu and Salomaa (1997) show that this grammar generates the simple non-context-free language $L_{xx} = \{xx | x \in \{a, b\}^+\}$, the language consisting of strings comprising two copies of any non-empty string of a's and b's. This grammar has derivations like the following:

$$\begin{aligned} S &\Rightarrow aSX \Rightarrow aaSXX \Rightarrow aaMRXX \Rightarrow aaMXRX \Rightarrow aaMXXR \\ &\Rightarrow aaMXXb \Rightarrow aaMQXb \Rightarrow aaMXQb \Rightarrow aaMXab \Rightarrow aaMQab \\ &\Rightarrow aabQab \Rightarrow aabaab \end{aligned}$$

It would be nice to present this derivation in some kind of phrase structure tree, but it is not clear how to portray the action of the non-CF rules. We could try something like the one shown in Fig. 1. It is not obvious how to identify the "constituents" in a derivation like this, constituents that could be compositionally interpreted. And context sensitive rewrite grammars can define languages that are intractable in the technical sense that they cannot be recognized in polynomial time (Karp, 1972; Kuroda, 1964), and apparently non-human-like languages like $\{a^n | n \text{ is prime}\}$. Languages like L_{xx} can be defined with rules that apply to constituents, regardless of context, if the constituents have more structure than strings (like trees, or tuples of strings) or if the rules can analyze and test the string components of their arguments. There are many natural classes of grammars of this kind, and some of them define only languages with polynomial recognition problems. We illustrate this more appealing kind of grammar with 3 more examples, all defining L_{xx}. Each of these 3 grammars will be presented in the simple format of Keenan and Stabler (2002, 2003) with a vocabulary V, a set of categories Cat, a finite lexicon Lex built from V and Cat, and generating rules that apply "bottom-up" to build complexes from the lexicon. The language is everything that can be obtained from the lexicon by application of the rules—that is, the language is the closure of the lexicon under the rules.

2.2.2. A 'matching' grammar

$V = \{a, b\}$ $Cat = \{S, T, U\}$ $Lex = \{\langle a, U\rangle, \langle b, U\rangle\}$ $Rule = \{F, G, H\}$

Here we have just two vocabulary elements, three categories, a lexicon containing two categorized strings (both with category U), and three rules for building complex expressions, defined

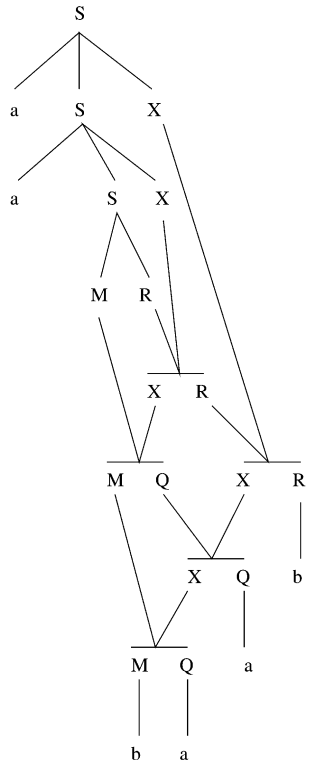


Fig. 1. A context sensitive grammar derivation

as follows:

$$\begin{aligned}
&\text{rule F}: \quad \langle s, \mathrm{T}\rangle\langle t, \mathrm{T}\rangle \mapsto \langle st, \mathrm{S}\rangle \quad \text{only if } s = t\\
&\text{rule G}: \quad \langle s, \mathrm{U}\rangle\langle t, \mathrm{T}\rangle \mapsto \langle st, \mathrm{T}\rangle\\
&\text{rule H}: \quad \langle s, \mathrm{U}\rangle \mapsto \langle s, \mathrm{T}\rangle
\end{aligned}$$

The first rule F says that a string s of category T and a string t of category T can be concatenated to yield the string st of category S, but only on condition that s and t are identical. Notice that this condition applies to strings s and t of any length. Intuitively, that means there is no finite bound on the amount of work required to apply this single rule. The second rule G and third rule H are read similarly but have no special conditions on their application, and so we can derive the string *abaaba* of category *S* as follows. The Montague-style derivation tree in Fig. 2 shows intermediate results fully at every node, with the more common phrase structure tree in Fig. 3. (Both kinds of trees are shown here to set the stage for the following grammars, where the Montague-style tree can be drawn straightforwardly but it is not clear how to draw the phrase structure trees.) It should be clear that this grammar does in fact derive exactly the strings L_{xx} with category S. We will call a grammar with conditions on the identity and non-identity of argument strings like "$s = t$" a matching grammar.

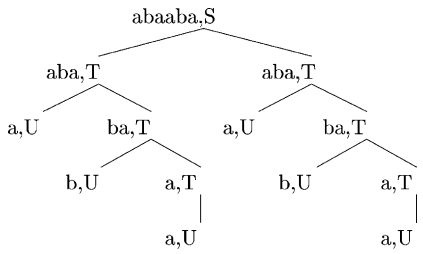


Fig. 2. Matching grammar derivation tree

2.2.3. A 'copying' grammar

Instead of (or in addition to) having discontinuous elements, we can define the same language using rules that copy strings of any length.

$$V = \{a, b\} \qquad \text{Cat} = \{S, T, U\} \qquad \text{Lex} = \{\langle a, U\rangle, \langle b, U\rangle\} \qquad \text{Rule} = \{F, G, H\}$$

Once again we have two vocabulary elements, three categories, and two categorized strings in the lexicon, with structure building rules are defined as follows:

$$\text{rule F}: \quad \langle s, T\rangle \mapsto \langle ss, S\rangle$$
$$\text{rule G}: \quad \langle s, U\rangle\langle t, T\rangle \mapsto \langle st, T\rangle$$
$$\text{rule H}: \quad \langle s, U\rangle \mapsto \langle s, T\rangle$$

Notice that the first rule says that if a string s has category T, then the string ss, the result of concatenating s with another copy of itself, has category S. We will say that grammars with rules like this have copying. Using these three rules, we have derivations like the one shown in Fig. 4, following the standard conventions for showing in full how each element is derived from simpler ones. It is not so clear how to draw a phrase structure tree for this derivation, but Fig. 5 indicates roughly how the copying introduces a number of crossing dependencies in a single step on its leftmost branch. It should be clear that this grammar again derives exactly the strings L_{xx} with category S, like the previous grammars.

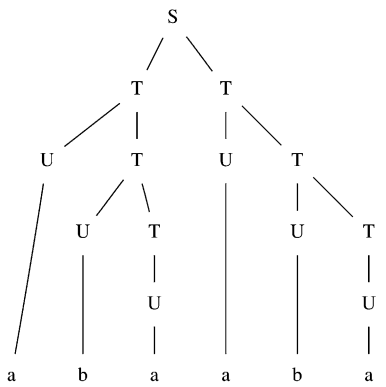


Fig. 3. Matching grammar phrase structure tree

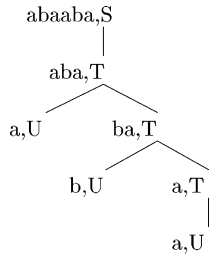


Fig. 4. Copying grammar derivation tree

2.2.4. A 'generating' grammar

We now define a grammar that categorizes some pairs of strings as well as some single strings.

$$V = \{a, b\} \qquad \text{Cat} = \{S, T, U\} \qquad \text{Lex} = \{\langle a, a, U\rangle, \langle b, b, U\rangle\} \qquad \text{Rule} = \{F, G, H\}$$

Again we have three vocabulary elements and three categories, but this time the lexical items are categorized *pairs of strings*. The pair of strings *a*, *a* and the pair of strings *b*, *b* are both lexical items with category U. Then the structure building rules are defined as follows:

$$\text{rule F}: \quad \langle s, t, T\rangle \mapsto \langle st, S\rangle$$
$$\text{rule G}: \quad \langle s, t, U\rangle\langle u, v, T\rangle \mapsto \langle su, tv, T\rangle$$
$$\text{rule H}: \quad \langle s, t, U\rangle \mapsto \langle s, t, T\rangle$$

The first rule says that if a pair of strings *s*, *t* has category T, then the concatenated string *st* has category S. The second rule says that if the pair *s*,*t* has category U and *u*, *v* has category T, then the pair *su*, *tv* has category T. The third rule is read similarly, and so we have derivations which can be depicted with trees like the one shown in Fig. 6, following the standard conventions for showing in full how each element is derived from simpler ones. And again it is not clear how to

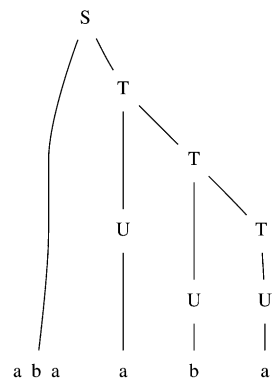


Fig. 5. Copying grammar phrase structure tree

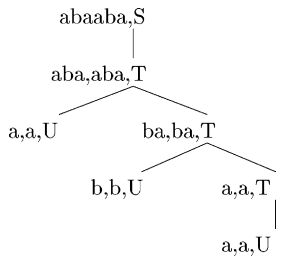


Fig. 6. Generating grammar derivation tree

draw a corresponding phrase structure tree, but the one in Fig. 7 indicate show the discontinuous pairs of elements get assembled in the derivation. It should be clear that this grammar does in fact derive exactly the strings L_{xx} with category S, like the previous grammars. Grammars like this where there is no matching or copying, but only tuples of (possibly discontinuous) elements, will be called a generating grammar.

2.3. First comparisons

The differences between the last three grammars in the previous section may seem insignificant, but their differences matter. As noted already, the matching and copying grammars have single steps that require a test or copy of unbounded size, while the generating grammar is more like a standard rewrite grammar. In fact, the generating grammar is an instance of a "multiple context free grammar" (Nakanishi, Takada, & Seki, 1997; Seki, Takashi, Mamoru, & Tadao, 1991) which is similar to (and equivalent in expressive power to) set-local multiple component tree adjoining grammars (Weir, 1988) and to a simple kind of 'minimalist grammar' (Harkema, 2001; Michaelis, 2001). These grammars are known to define a mathematically natural and well-behaved class of MCS languages.[5] We will use the simple generating grammar of Section 2.2.4 as a representative of this range of similar MCS grammatical formalisms. If we add the ability to copy strings, as the copying grammar does, it becomes easy to generate languages

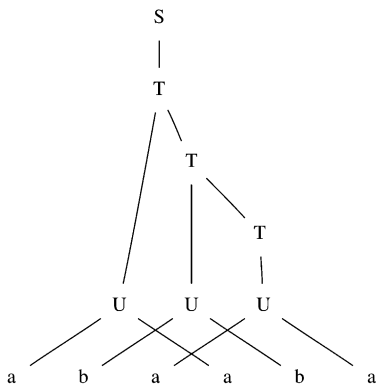


Fig. 7. Generating grammar derivation tree

which violate the constant growth bound. For example, the non-MCS language a^{2^n} is defined by the grammar with lexicon $\{\langle a,S\rangle\}$ and the following simple copying rule:

$$\text{rule F}: \quad \langle s, S\rangle \mapsto \langle ss, S\rangle$$

And obviously if a "matching grammar" is allowed to apply any string testing function, it is possible to generate any language at all. In sum, of the four kinds of grammar considered here (cs-rewrite, matching, copying and generating), the generating grammars appear to involve the most restrictive mechanisms, defining only MCS languages, and so they may be preferred for this reason unless relevant considerations support one of the other options. Here the focus will be on the nature of the relation between syntactic structure and semantic values.[6]

3. Limited cross-serial dependencies and islands

As noted above, our example generating grammar is an instance of a "multiple context free grammar" (MCFG) equivalent in expressive power (and structurally similar) to set-local multiple component tree adjoining grammars, to minimalist grammars, and many other similar formalisms. Each MCFG has rewrite rules for categorized k-tuples of expressions, where k is bounded to some finite value. For example, the generating grammar above derives some constituents with pairs of strings ($k = 2$), where these strings can be discontinuous in the final string: no rule can take apart any substring that has already been derived, but the rules can "stretch apart" the separate components of a constituent. These generating grammars are restricted in the sense that they can only define languages with tractable, polynomial recognition problems. Furthermore, MCFGs are limited in the cross-serial dependencies they can define, even though they apparently do not enforce any conditions like those mentioned in (1–3), since nothing in the formalism itself says anything about subjects, noun phrases or phases. But the bound on tuple size guarantees that every language will have "island constraints" of some kind, in the sense that no operation in the grammars can exceed the bound k on the number of tuples. This restriction will emerge in different ways in different grammars. For example, when $k = 2$ we could construct a grammar of relative clauses where one component is used to hold a relative pronoun that is "moving" from its canonical "gap" position to its surface position. Since there is no way to add another "moving element," a condition like (1) could follow. In fact, subjacency-like restrictions emerge in tree adjoining grammars in just this way (Frank, 2002; Kroch, 1987, 1989). It is plausible that various versions of (3) could also emerge similarly. So the bound k on the number of elements "waiting assembly" imposes a restriction that is welcome in linguistic theory, since various kinds of island constraints are so common in human languages.

4. Some controversial cases

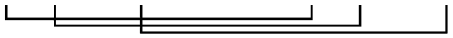
It is not difficult to provide generating grammars for the constructions mentioned in the introduction. For example, tree adjoining grammars have been proposed for the English auxiliary, for Dutch crossing dependencies, and for Swiss–German-like crossing dependencies (Abeillé & Rambow, 2000; Joshi, 1990, for example). But the situation is more difficult for A-not-A questions, VP ellipsis, and certain other constructions.[7]

4.1. A-not-A questions

A certain kind of yes/no question in Mandarin Chinese appears to involve copying a verbal constituent that we will call a verb phrase (VP). (The question of whether the copied element in the following examples is a VP, vP, TP or some other related verbal phrase need not be decided here.) Consider, for example, the following yes/no question, where (*) indicates a position where a pause is very awkward,

(8) Zhangsan like play basketball not like play basketball

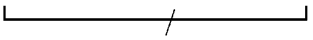
Zhangsan ai da lanqiu (*,) bu ai da lanqiu



'Does Zhangsan like to play basketball?'

The perfect copy correspondence is required in this construction, as we see by the contrast with the following where there is a non-identical element, with a required pause (a clause break) indicated by (*) and the structure is not interpreted as a yes/no question:

(9) Zhangsan like play basketball not like play volleyball
Zhangsan ai da lanqiu *(,) bu ai da paiqiu

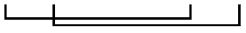


'Zhangsan likes to play basketball, not to play volleyball'

However, the copied element does not need to be the whole VP. The following examples are fine yes/no questions, synonymous with (8) above:

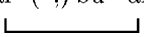
(10) Zhangsan like play not like play basketball

Zhangsan ai da (*,) bu ai da lanqiu



(11) Zhangsan like not like play basketball

Zhangsan ai (*,) bu ai da lanqiu



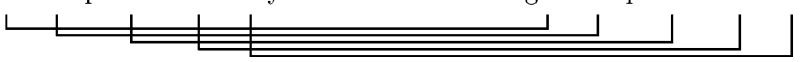
Huang (1991), and Radzinski (1990) and others observe that these constructions apparently involve copy-type dependencies of unbounded complexity (since the complexity of the copied VP is not subject to any fixed, finite bound), and hence this construction is beyond the weak expressive power of a context free grammar. Huang (1991) proposes that some A-not-A constructions are formed by a kind of reduplication process (copying), while others involve deletions that can be performed only under an identity condition (matching). Interestingly, no generating analysis of A-not-A questions has ever been proposed, to my knowledge, but only matching and copying analyses (Dai, 1990; Ernst, 1994; Huang, 1991; Li, 1992, 2002; McCawley, 1994; Wu, 1997; Zhang, 1997).Why is this? One obvious idea is this: while the crossing relations in (6) and (7) are semantically complex, with each dependency corresponding to a distinct pair of semantically related elements, in A-not-A, the whole set of crossing dependencies corresponds to a single simple semantic contribution, namely, the indication that the proposition is being questioned. Many languages have productive copy constructions, like A-not-A in Chinese: the X-or-no-X construction in English (Manaster-Ramer),[8] contrastive focus reduplication in

English (Ghomeshi, Jackendoff, Rosen, & Russell, 2004), whole word reduplication in the African language Bambara (Culy, 1985), partial reduplication in the Australian language Mangarayi (Kurisu & Sanders, 1999), the American language Yaqui (Harley & Amarillas, 2003), the Austronesian language Manam (Buckley, 1997), and very many others. In all these cases, the semantic contribution of the copy is simple in the sense that it is not dependent on the semantic values of the elements of the copied constituent. That is, the copying is "atomic", "morphemic" in this sense. In the Dutch and Swiss–German constructions above, on the other hand, the dependencies are not constrained to relate identical elements, and each dependency corresponds to a different instance of the semantic is an argument of relation. For those cases, a generating grammar with crossing dependencies derived in some number of steps is natural. But when the related elements are identical and the semantic contribution of the whole copied sequence is simple, it is perhaps more natural to use a copying or matching operation, so that a simple syntactic operation makes a simple semantic contribution (even when the copied element is complex). This idea could be expressed in a hypothesis like this, motivated by our assumptions about the semantic values of derived constituents: (H1) Generating grammars do not provide a good model for mono-morphemic copy constructions. If H1 is correct, and if the copy correspondences are enforced by the syntax, then it appears we need copying or matching mechanisms that will threaten the MCS hypothesis. We will not attempt to resolve the matter here, but only observe that the lack of generating analyses of copy constructions in the linguistic literature seems not to be a mere accident of linguistic fashion. The analyses of A-not-A and the many similar constructions in syntax, morphology and phonology all have this common trait, distinguishing them from the familiar Dutch and Swiss–German constructions: the apparently complex copied material makes a simple semantic contribution. It is tempting here to consider also a stronger hypothesis: Isn't the simplicity of the semantic contribution of the copy in these constructions predictable, since, roughly, a copy introduces no new semantic element except the copying itself? So maybe we should propose that in all languages, every sequence of crossing dependencies among elements that are required to be identical (modulo regular phonological processes) is morphemic, semantically simple. But this would be a mistake, or at least highly suspect, as shown in the following section.

4.2. VP ellipsis

English and many other languages seem to allow a VP to be missing or unpronounced in certain cases, and some theories propose that the missing VP is actually computed—either it is derived and then "deleted," or it is "reconstructed" somehow. Here the unpronounced VP is shown crossed out:[9]

(12) He didn't give the presents to Mary but I think Bill did ~~give~~ ~~the~~ ~~presents~~ ~~to~~ ~~Mary~~



Perhaps the strongest arguments for the view that the missing VP is really there even though it is not pronounced come from examples where constraints on grammatical form seem to apply to the missing VP. One such case is provided by Haïk (1987) and Fiengo and May (1995), who suggest that the contrast in acceptability between the forms (13) and (14) with VP ellipsis is

analogous to the contrast between (15) and (16). In these examples we put e at ellipsis sites, *t* at empty question word positions, and we use subscript *i* to indicate coreference:

(13) Dulles suspected everyone who Angleton believed that Philby did *e*.
(14) * Dulles suspected everyone who Angleton wondered why Philby did *e*.
(15) Dulles suspected everyone who$_i$ Angleton believed that Philby suspected t_i.
(16) * Dulles suspected everyone who$_i$ Angleton wondered why Philby suspected t_i.

Similarly, Fiengo and May (1995) suggest that there is a contrast in acceptability between the VP ellipsis constructions (17) and (18) analogous to the contrast between (19) and (20):

(17) This is the book which$_i$ Max read t_i before knowing that Lucy did *e*.
(18) * This is the book which$_i$ Max read t_i before hearing the claim that Lucy did *e*.
(19) This is the book which$_i$ Max read t_i before knowing that Lucy read t_i.
(20) * This is the book which$_i$ Max read t_i before hearing the claim that Lucy read t_i.

Other kinds of arguments have been offered too, based not on similarities but on subtle differences between the semantic values of the elided and corresponding non-elided forms (Kennedy, 2003, for example). These constructions pose a challenge for linguists who have argued that the missing VP does not have any hidden syntactic structure (Dalrymple, Shieb, & Pereira, 1991; Hardt, 1993), and so some recent proposals have been mixed, assuming that an elided VP structure is computed in some but not all cases (Kehler, 2002; Lappin, 1996). There is a similar range of assumptions in psycholinguistic studies of the comprehension and acquisition of these structures (Foley, del Prado, Barbier, & Lust, 2003; Shapiro & Hestvik, 1995; Tanenhaus & Carlson, 1990, for example) We will not try to resolve the issues here. The relevant point is that various kinds of copying and matching theories of VP ellipsis are very much alive, and they pose a threat to the MCS hypothesis. It is interesting to note that while there are matching and copying theories of syntactic deletion (or reconstruction) in VP ellipsis, there are no MCS generating accounts, to my knowledge. Why is this? We did not present any deletion rule in the generating grammar of Section 2.2.4, but there is no obstacle to it. The class of MCFGs mentioned above allows rules which fail to copy one or more of the string components of an argument into a string component of a value. And the absence of generating accounts of VP ellipsis is not predicted by H1, since the contribution of the elided VP is semantically complex; to a first approximation, each element of the elided makes the same semantic contribution it would make if it were actually there (even though much of the cited literature explores the extent to which this fails to be quite true). So what explains the lack of MCS generating accounts for VP ellipsis? Again, there seems to be a reason for this fact. In these constructions, it seems implausible that the copied phrase is syntactically related to the phrase it copies, at least if one restricts attention to the immediate syntactic relations among constituents in a simple clause. We could tentatively try to formulate this idea more generally:

(H2) Generating grammars do not provide a good model for copy constructions in which the pairwise corresponding elements are neither lexical items nor related by selection.

In the simple grammar of Section 2.2.4 the corresponding pairs were discontinuous lexical items, and the corresponding pairs in (5–7) are plausibly related directly by “selection”: one

is a syntactic argument of the other. It could be that generating accounts could be extended to certain other kinds of crossing dependencies, but the basic idea is clear: the copy relation in VP ellipsis (if there is one) relates elements that we would not expect to be common arguments of any syntactic rule.

4.3. An alternative perspective: H1 and H2 rejected

We have seen that for certain types of crossing dependencies, all the well-known proposals involve non-MCS copying or matching type explanations, or else they claim the relevant phenomena are non-syntactic or non-grammatical in some sense. H1 might be the implicit assumption behind the lack of generating analyses of Chinese A-not-A questions and related constructions. And H2 might be the implicit assumption behind the lack of generating analyses of VP ellipsis and related things, even among those linguists who think that a syntactic account is needed. But if H1 and H2 are correct, then there are crossing correspondences that either need to be excluded as non-syntactic or non-grammatical, or else they need to be defined by copying or matching mechanisms that threaten the MCS hypothesis.[10] So let's assess these these challenges to the MCS hypothesis more carefully. It could be that lack of generating analyses for these phenomena does not really represent a consensus in support of H1 and H2; this gap in the literature could be simply an accident, or it might have some other explanation. If H1 and H2 are actually unsupported, and especially if they can be rejected on independent empirical grounds, then the threat to the MCS hypothesis is removed. We will briefly sketch just one position of this sort, not to resolve the controversial issues but to make clear that the prospects for MCS hypothesis still look reasonable, if not quite empirically secure. Consider H1 first: why should we believe it? The natural idea, hinted at earlier, is that semantically simple expressions should be syntactically simple. If this were true, the correspondence between basic syntactic constituents and basic semantic values would be very neat. But human languages are not like this, as we see immediately from the abundance of idioms in all languages. Idioms are syntactically complex expressions with meanings that are not calculated in the usual way from the meanings of their parts. Rather, the meaning of an idiom must just be learned: idioms, like simple morphemes, are 'semantic atoms'. For example, in English, the phrase *keep tabs on*, meaning something like *pay attention to*, is clearly syntactically complex, as we can see for example from the fact that the object of the preposition can vary, it contains a verb which gets inflected in the usual way for third person present or past tense, and (at least for some speakers) it can even participate in passivization:

(21) He keeps tabs on me
(22) He kept tabs on you.
(23) Tabs were kept on all suspicious persons.

There are many different kinds of examples like this, complexes that are semantically idiosyncratic Sag, Baldwin, Bond, Copestake, and Flickinger (2002), and these have important implications for how we understand expressions (Abeillé & Schabes, 1989; Gibbs, Nayak, & Cutting, 1989, for example). But here we need only observe that the existence of many uncontroversial cases of expressions that are syntactically complex and semantically simple removes any temptation to suppose that semantic simplicity implies syntactic simplicity.

Without this supposition, is there any reason to believe H1? There is another more subtle point to be made against H1. While many linguists can accept the idea that a complex expression might be syntactically analyzed or generated and then assigned a meaning as a whole, some will still be reluctant to imagine that the there is a mechanism in the grammar which allows the construction of one verb phrase to control the step-by-step construction of another, as generating approaches to A-not-A or VP ellipsis require.[11] 'Copying grammars', in contrast, extend a structure by copying some part of it in one step, with no bound on the size of the copied material. But it is interesting to notice that linguists are also inclined to suppose that the copied material has structure, just like the source phrase does, a structure that has consequences for prosodic properties (roughly: the tone and stress in its pronunciation), binding options, and so on. While we can of course stipulate that not only a string of unbounded size but in fact a whole structure of some kind is copied, this raises the question: Why should the copied material have the structure of the source, even when it is produced in an entirely different way? The generating accounts need no stipulation of this sort, since the source and the copy are both generated step by step, in the same or very similar ways, with structures that reflect that fact. So again, this consideration favors the generating accounts. The initial plausibility H1 seems to give way on closer inspection, and I find no persuasive evidence of any kind for it. Turning to H2 and VP ellipsis, some additional issues come up. We are familiar with the fact that a predicate can be arbitrarily far from an argument that it selects, when that argument has "moved" or been "stretched away" by adjunctions:

(24) I know *which book* Mary said you wanted me to tell Bill to buy *t*.

Here *which book* is the object selected by the deepest predicate *buy* even though it appears far away from that predicate. Recent proposals involving 'remnant movement' and 'sideward movement' aim to significantly extend earlier ideas of the kinds of discontinuous dependencies that should be countenanced here (Koopman & Szabolcsi, 1998; Hornstein, 1999; Nunes, 2001), but still there are severe constraints on the relations between between positions, as we see from the unacceptability of things like

(25) *I know both *which book* John laughs and Mary likes *t*.
(26) *I know *which book* you believed the claim that Bill bought *t*.
(27) **Which book* did you like *t*? Did Bill buy *t*?

The prospects for generative theories of relations subject to such constraints look very good. Considering the relation between a VP source and an associated ellipsis site, on the other hand, matters seem rather different. The VP source obviously is not 'selected' in the ellipsis site, or anything like that, and furthermore, while there are constraints on the relative structural positions of a VP and associated ellipsis site, these constraints seem—at least on the face of it—unlike the constraints on the relation between a wh-phrase and the position where it is selected. For example, the VP source and the ellipsis site do not even need to be in the same sentence, and there can be multiple antecedents appearing in different orders than the ellipsis sites, as in these examples from corpora studies discussed by Klein and Stainton-Ellis (1989) and Gardent (1993):

(28) a. I promised myself I *wouldn't go to Manchester*$_i$ unless I first *opened a big stack of mail*$_j$. I didn't e$_j$, so I didn't e$_i$.
b. If you *work hard, make the right choices and keep your nose clean*$_i$, you *get ahead*$_j$. If you don't e$_i$, you don't e$_j$.
c. I was *really thin*$_i$ then, and I tried some ski-pants that *looked really good on me*$_j$, and I *should have bought them*$_k$. But I didn't e$_k$, and now I'm not e$_i$, so they wouldn't e$_j$.

This does not look like anything that clausal syntax should extend to cover.[12] Nevertheless, there are some cases where not only the meaning of the source VP but its structure seem to constrain the interpretation of the ellipsis site. In addition to the slightly exotic contrasts among (13–20) above, there are simple cases like (29b) in which subscripting is used to indicate not the VP source and ellipsis but coreference between the nominals *Bill* and *he*:

(29) a. The lawyer *defended Bill*, and I did too *e*.
b. *The lawyer *defended Bill*$_i$, and he$_i$ did too *e*.
c.*He$_i$ defended Bill$_i$.

(29b) is not interpreted with the indicated coreference relations, any more than (29c) is, because of the structural constraints on the configuration of pronoun and antecedent. Noting examples like these, Kehler (2002) makes a persuasive case for the view that some but not all discourse functions of clauses require a structural similarity, and so the listener tries to reconstruct the source phrase at some but not all ellipsis sites. In particular, in cases like (29a), the point of the utterance seems to be to say that two events are similar or parallel, which sets up a problematic correspondence in (29b) between the second subject and the parallel elided pronoun position. In examples like (28), on the other hand, the point of the sentences is to identify cause-effect relationships, and so no structural parallelism is sought. Kehler proposes that syntactic effects are found, and expected, in clauses that are used to indicate parallelism, while they are typically not found, and not expected, in accounts of cause-effect relations like those of (28). If we suppose this is true, then we should consider the option of letting a generating theory of discourse provide a model for syntactic reconstruction, when it occurs. Notice that since the source VP is sometimes in the same sentence and sometimes not, a single grammar projecting structure from lexical items to clauses to discourses is a natural option to pursue here. And there are reasons going beyond the apparent structure sensitivity of interpretation in 'parallel' clauses to think that this may be the right kind of account.[13] For example, Johnson (2001) notices that there is a contrast between acceptable VP ellipsis constructions like those in (30) and the much less acceptable ones in (31):

(30) a. He would *eat rutabagas*, and Holly would *e* too.
b. He has *eaten rutabagas*, and Holly has *e* too.
c. He is *eating rutabagas*, but Holly's not *e*.
d. He wants to *eat rutabagas*, and Holly wants to *e* too.

1. *He *would eat rutabagas*, and Holly *e* too.
b. *He is being *discussed widely*, and Holly is being *e* too.
c. *I consider Sue *intelligent*, but she considers Holly not *e*.
d. *John made Bill *eat rutabagas*, and Holly made me *e* too.

And he observes that this parallels to a surprising extent a contrast between possible sites of 'VP topicalization'—something a grammar of clause construction is certainly expected to account for. For example, the constructions in (32) are marginal, but at least for some speakers, those in (33) are much more so:

(32) a. *eat rutabagas*, Holly would *t*.
b. *eaten rutabagas*, Holly has *t*.
c. *eating rutabagas*, Holly's not *t*.
d. *eat rutabagas*, Holly wants to *t*.

(33) a. **would eat rutabagas*, Holly *t*.
b. **discussed widely*, Holly is being *t*.
c. **intelligent*, I consider Holly not *t*.
d. **eat rutabagas*, Holly made me *t*.

Johnson argues that this is not a coincidence, and that the two processes must have some common underlying elements: for example, it may be that every ellided VP is topicalized first. A generating theory that defines both clausal and discourse structure could provide an analysis of this kind, with a significant MCS component for structural reconstruction when it occurs, relying on non-syntactic and probably non-MCS mechanisms to establish the meanings of ellided VPs in cause-effect reports not requiring structural parallelism. This perspective extends grammatical mechanisms much further than (H2) contemplates.

5. Conclusions

It is clear that the MCS hypothesis draws a distinction among grammars that cuts right through several quite different controversies in linguistic theory. And it has been argued here that the controversies are substantial, not whims of theoretical fashion or notational fancies. We noticed that the MCS hypothesis seems to be threatened by a couple of different kinds of phenomena. Previous discussions of these phenomena invoke non-MCS mechanisms, sometimes with the proviso that they are really non-syntactic or non-grammatical, so that the 'syntax' or 'grammar' could remain MCS. Here, we formulated two non-MCS hypotheses H1 and H2 in a first effort to characterize this apparent consensus, but even a brief assessment shows that these hypotheses are not well supported. That is, although these hypotheses seem consistent with what we find in the literature, on closer examination we discover that the hypotheses are rather dubious. Putting the matter more directly: although the literature seems to lack any MCS proposals for A-not-A questions and for structural effects in VP ellipsis, the prospects for such proposals look reasonable. Consequently, the apparent non-MCS consensus should not be taken seriously until matters are investigated further. Pushing towards a resolution of these issues will reveal a better picture of human linguistic abilities, and help establish the status of MCS. Previous discussions of the MCS hypothesis have also led to some slightly weaker hypotheses which we did not consider here. Though weaker, these claims still have real bite, still conflicting with many proposals in the field. In the first place, the grammars that result from the addition of unbounded copying to MCFGs have been studied. Seki et al. (1991) call

these grammars "parallel" MCFGs (pMCFGs). They still have polynomial parsing complexity, and if the components of expressions are used to hold "moving" elements, then we would still expect "island effects" to emerge in the same way (as mentioned briefly in Section 3 above). The pMCFGs cannot define all polynomially recognizable sets, and the polynomially recognizable languages have a remarkably simple generative definition, so another slightly weaker but still appealing position would be to keep just this part of the MCS hypothesis. Groenink (1997, 1997a) and Kracht (1998, 2003) consider this and some other related weakenings of the MCS hypothesis. There is a natural tendency for linguists to help themselves to whatever mechanisms seem adequate and appropriate to the task at hand, but human languages are vast and complex, so it becomes easy to lose track of the big picture: when are two proposed mechanisms really the same; when is one mechanism really an instance of another; what kinds of capabilities are really needed? The MCS hypothesis and the research it has fostered provides a good first step towards a more general and parsimonious perspective. Recent work even shows that some interesting, infinite subsets of the MCS class (and probably closely related classes too) can be identified from "texts" of positive examples (Christian & Bonato, 2001; Kanazawa, 1998; Kobele, Collier, Taylor, & Stabler, 2002), suggesting new and more feasible perspectives on the human language learning problem.

Notes

1. For example, consider the following passage: "In the general case of theory construction, the primitive basis can be selected in any number of ways, ... But in the case of UG [universal grammar], other considerations enter. The primitive basis must meet a condition of epistemological priority. That is, . . . we want the primitives to be concepts that can plausibly be assumed to provide a preliminary, prelinguistic analysis of a reasonable selection of presented data, that is, to provide the data that are mapped by the language faculty to a grammar . . . . It would, for example, be reasonable to suppose that such concepts as "precedes" or "is voiced" enter into the primitive basis, and perhaps such notions as "agent-of-action" if one believes, say, that the human conceptual system permits an analysis of events in these terms independently of acquired language. But it would be unreasonable to incorporate, for example, such notions as "subject of a sentence" or other grammatical notions, since it is unreasonable to suppose that these notions can be directly applied to linguistically unanalyzed data. (Chomsky, 1981, p. 10)"
2. The notion of polynomial recognizability is discussed in any standard introduction to formal languages and computing, such as Hopcroft, Motwani, and Ullman (2000) or Lewis and Papadimitriou (1981).
3. For example, there is a long tradition of results showing that many prominent linguistic frameworks do not respect the restriction to polynomial parsability: Peters and Ritchie (1973) showed that a "standard" transformational grammar can define undecidable sets; the results of Johnson (1988), Trautwein (1995), Torenvliet and Trautwein (1995) show the undecidability of grammars which allow unrestricted unification. Barton, Edward, Robert Berwick, and Ristad, (1987) provide a useful discussion of the sources of

intractability in these results. In contrast to these frameworks, tree adjoining grammars (Abeille & Rambow, 2000,Joshi & Schabes, 1997), combinatory categorial grammars (Steedman, 1996),VijayShanker & Weir, 1994, and simple 'minimalist' grammars (Stabler, 1997; Michaelis, 1998; Harkema, 2000) are MCS.

4. Keenan and Stabler (2002, 2003) provide a rigorous definition of what it means to say that a semantic relation is coded in the syntax, and show that while some semantic relations, like entailment, are never coded in the syntax of human languages, others like *is an argument of*, or *is a possible antecedent of*, (arguably) are. The relations indicated by the arcs in (6) and (7) are *is an argument of* relations, which are often coded by *sisterhood* or some other very simple syntactic relation.
5. The languages defined by multiple context free grammars, (set-local) multiple component tree adjoining grammars and minimalist grammars are mathematically natural in the sense that they are semi-linear and they define an "abstract family of languages" (AFL) (Seki et al., 1991; Weir, 1988).
6. One finds more abstract criteria for linguistic theory proposed too, sometimes with rather obscure motivations. For example, Chomsky (2001a) suggests that it is desirable to have grammars that are "failure proof" in the sense that every derivation extends to a complete one. The significance of this proposal clearly depends on the definition of a derivation. Do all three grammars presented above have this property? Well, perhaps the matching grammar loses out, since in the other grammars, it is possible to extend any constituent of category T to an S, while in the matching grammar it is possible to build two constituents of type T with different strings $s \neq t$ and there is no way to extend these constituents to a complete derivation of S that includes them both. Chomsky does not offer any argument for this criterion, but says that while it may be too strong,"something like it has motivated much recent work, which has sought to eliminate comparison of derivations, backtracking and look-ahead, and "non-local" operations generally." The move toward "failure proof" grammars apparently conflicts with the earlier (and even now very well entrenched) idea that derivations are filtered by constraints. Indeed, some linguists regard the existence of (filtering) constraints as the main discovery of recent linguistic theory (Baker, 2001, Section 4.2 for example). Chomsky (2001a) also proposes some other properties that linguistic theories should "optimally" have: (Inclusiveness) the derivation introduces no new elements but only rearranges lexical material, and (one cycle) the mapping from phonological to semantic representations is accomplished in "one cycle," one pass through the structure. Again the motivations are obscure, but all three of our example grammars (arguably) have these latter properties.
7. The analysis of Chinese number names in Radzinski (1991) poses a problem for the MCS hypothesis, as does the analysis of the case marking system of Old Georgian in Michaelis and Kracht (1997). These constructions are peripheral in Chinese and Georgian, and the linguistic status of the data in these examples is debated, and so we will leave them aside here in favor of several other much more common constructions: A-not-A questions in Mandarin Chinese and VP-ellipsis in English.
8. This was pointed out in some of Manaster-Ramer's unpublished lectures, which are cited in Gazdar and Pullum (1985) and in Manaster-Ramer (1986).

9. On the analysis of Li (2002), examples like (10) and (11) above have already shown VP ellipsis, but since in those examples the verb itself appears while its object is missing, Li's proposal involves complexities we do not need to get into here.
10. The existence of non-MCS copying or matching mechanisms is assumed by theories in generative syntax that treat syntactic movement as a complex of copying and deletion processes, and other traditions in syntax have mechanisms of similar power. The grounds for having such expressive power are more theory-internal than the issues considered here, and so are excluded from this study.
11. Remember again that while the example 'generating grammar' given above implements this by constructing the two copies 'at once', the same effect can be implemented in various other ways. For example, some grammars have a kind of control mechanism to allow early steps to condition later steps. The results of Seki et al. (1991) and Vijay-Shanker and Weir (1994) show how many different formalisms can encode these kinds of our dependencies.
12. Even worse, there are cases where the interpretation of a single VP ellipsis site is determined in some way by two different antecedents, as in this example from Webber (1978):(34) Wendy is eager to *sail around the world* and Bruce is eager to *climb Kilamanjaro*, but neither of them can e because money is too tight. Here the ellided e is interpreted as something like the disjunction of the two antecedent VPs.
13. Webber (2004) briefly mentions some possible difficulties for single grammars defining both clause and discourse structures. Those difficulties do not look insuperable to me, but a more careful exploration of whether the theory of discourse really should be 'independent' in some sense beyond the scope of this paper.

Acknowledgments

This paper is dedicated to Aravind Joshi, who has provided so many inspiring examples of clear, fundamental, empirically assessable linguistic research, and who has generously encouraged so many others. A version of this paper was presented to the 2002 UCLA Mathematical Linguistics seminar and to the 2003 Annual Meeting of the Cognitive Science Society. Thanks to Ying Lin, Greg Kobele, Marcus Kracht, Edward Keenan and Dominique Sportiche for valuable suggestions, and to Ying Lin, for examples (8–11).

References

Abeillé, Anne and Owen Rambow. 2000. Tree adjoining grammar: An overview. In *Tree adjoining grammars: Formalisms, linguistic analysis, and processing* (pp. 1–68). Stanford, CA: CSLI Publications.

Abeillé, Anne and Yves Schabes (1989). Parsing idioms in lexicalized TAGs. In *Proceedings of the European association for computational linguistics* (pp. 161–165).

Baker, Mark C. (2001). Syntax. In Mark Aronoff, Janie Rees-Miller (Eds.), *The Handbook of Linguistics* (pp. 265–294). Oxford: Blackwell.

Barton, G. Edward, Robert C. Berwick, & Eric Sven Ristad (1987). *Computational complexity and natural language*. Cambridge, MA: MIT Press.

Bresnan, J., Kaplan, R. M., Peters, S., & Zaenen, A. (1982). Cross-serial dependencies in Dutch. *Linguistic Inquiry*, *13*(4), 613–635.

Buckley, Eugene (1997). Integrity and correspondence in Manam double reduplication. In *Proceedings of the North Eastern linguistic society, NELS 27*.

Chomsky, N. (1956). Three models for the description of language. *IRE transactions on information theory, IT-2*, 113–124.

Chomsky, Noam (1981). *Lectures on government and binding*. Dordrecht: Foris.

Chomsky, Noam (2001a). Beyond explanatory adequacy. *MIT occasional papers in linguistics* (p. 20).

Chomsky, Noam (2001b). Derivation by phase. In Michael Kenstowicz (Ed.), *Ken Hale: A life in language*. Cambridge, MA: MIT Press.

Culy, C. (1985). The complexity of the vocabulary of Bambara. *Linguistics and Philosophy*, *8*(3), 345–352.

Dai, J. Z.-L. (1990). Some issues on A-not-A questions in Chinese. *Journal of Chinese Linguistics*, *18*(2), 285–371.

Dalrymple, M., Shieber, S., & Pereira, F. (1991). Ellipsis and higher order unification. *Linguistics and Philosophy*, *14*, 399–452.

Ernst, T. (1994). Conditions on Chinese A-not-A questions. *Journal of East Asian Linguistics*, *3*, 241–264.

Fiengo, R., May, R. (1995). *Indices and identity*. Cambridge, MA: MIT Press.

Foley, C., del Prado, Z. N., Barbier, I., & Lust, B. (2003). Knowledge of variable binding in VP ellipsis. *Syntax*, *6*, 52–83.

Frank, R. 2002. *Phrase structure composition and syntactic dependencies*. Cambridge, MA: MIT Press.

Gardent, C., 1993. A unification-based approach to multiple VP ellipsis resolution. In *Proceedings of the sixth European association for computational linguistics, EACL* (pp. 139–148).

Gazdar, G., Geoffrey K. Pullum (1985). Computationally relevant properties of natural languages and their grammars. *New generation computing* (Vol. 3, pp. 273–306). In Walter J. Savitch, Emmon Bach, William Marsh, & Gila Safran-Naveh (Eds.), *The formal complexity of natural language*, Reidel, 1987.

Ghomeshi, J., Jackendoff, R., Rosen, N., & Russell, K. (2004). Contrastive focus reduplication in English. *Natural Language and Linguistic Theory*, *22*, 307–357.

Gibbs, R. W., Nayak, N. P., & Cutting, C. (1989). How to kick the bucket and not decompose: Analyzability and idiom processing. *Journal of Memory and Language*, *28*, 576–593.

Groenink, A. (1997). Mild context-sensitivity and tuple-based generalizations of context-free grammar. *Linguistics and Philosophy*, *20*, 607–636.

Haïk, I. (1989). Bound variables that need to be. *Linguistic Inquiry*, *11*, 503–530.

Hardt, Daniel (1993). *Verb phrase ellipsis: Form, meaning and processing*. Ph.D. Thesis, University of Pennsylvania.

Harkema, Henk (2000). A recognizer for minimalist grammars. In *Proceedings of the sixth International Workshop on parsing technologies, IWPT'2000*.

Harkema, Henk (2001). A characterization of minimalist languages. In *Proceedings of the logical aspects of computational linguistics, LACL'01*, Port-aux-Rocs, Le Croisic, France.

Harley, Heidi & Maria Amarillas (2003). Reduplication multiplication in Yaqui: Meaning x form. In Luis M. Barragan, Jason D. Haugen (Eds.), *Studies in Uto-Aztecan*, MIT Working Papers on endangered and less familiar languages (p. 5). Cambridge, MA: MIT Press.

Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman (2000). *Introduction to automata theory, languages and computation* (2nd ed.). Reading, MA: Addison-Wesley.

Hornstein, N. (1999). Movement and control. *Linguistic Inquiry*, *30*, 69–96.

Huang, Cheng-Teh James (1991). Modularity and Chinese A-not-A questions. In C. Georgopoulos, R. Ishihara (Eds.), *Interdisciplinary approaches to language: essays in honor of S.-Y. Kuroda*. Dordrecht: Kluwer.

Huybregts, M.A.C. (1976). Overlapping dependencies in Dutch. Technical report, University of Utrecht. Utrecht Working Papers in Linguistics.

Hyams, Nina (1986). *Language Acquisition and the Theory of Parameters*. Reidel, Boston.

Johnson, Kyle (2001). What VP ellipsis can do, and what it can't, but not why. In Mark Baltin, Chris Collins (Eds.), *The handbook of contemporary syntactic theory* (pp. 439–479). Oxford: Blackwell.

Johnson, Mark (1988). *Attribute value logic and the theory of grammar*. Number 16 in CSLI Lecture Notes Series. Chicago: CSLI Publications.

Joshi, Aravind (1985). How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen & A. Zwicky (Eds.), *Natural language processing: Theoretical, computational and psychological perspectives* (pp. 206–250). NY: Cambridge University Press.

Joshi, A. K. (1990). Processing crossed and nested dependencies: An automaton perspective on the psycholinguistic results. *Language and Cognitive Processes*, *5*(1), 1–27.

Joshi, Aravind K., & Yves Schabes (1997). Tree-adjoining grammars. In G. Rozenberg, A. Salomaa (Eds.), *Handbook of formal languages*, (Vol. 3, pp. 69–124) Beyond words. NY: Springer.

Kanazawa, Makoto (1998). *Learnable classes of categorial grammmars*. Stanford, CA: CSLI Publications.

Karp, R.M. (1972). Reducibility among combinatorial problems. In R.E. Miller, J.W. Thatcher (Eds.), *Complexity of Computer Computations* (pp. 85–103). NY: Plenum Press.

Keenan, Edward L., & Edward P. Stabler. 2002. Syntactic invariants. In Ann Copestake, Kees Vermeulen (Eds.), *Algebras, Diagrams and Decisions in Language, Logic and Information*, Stanford, CA: CSLI Publications.

Keenan, Edward L.,& Edward P. Stabler. 2003. *Bare grammar*. Stanford, CA: CSLI Publications.

Kehler, Andrew (2002). *Coherence, reference, and the theory of grammar*. Stanford, CA: CSLI Publications.

Kennedy, Christopher (2003). Ellipsis and syntactic representation. In Kerstin Schwabe, Susanne Winkler (Eds.), *The interfaces: deriving and interpreting omitted structures*. Amsterdam: John Benjamins.

Klein, Ewan, & Kate Stainton-Ellis (1989). A note on multiple VP ellipsis. Technical Report EUCCS/RP-30, Centre for Cognitive Science, University of Edinburgh.

Kobele, Gregory M., Travis Collier, Charles Taylor, and Edward Stabler (2002). Learning mirror theory. In *Proceedings of the sixth International workshop on tree adjoining grammars and related frameworks*.

Koopman, Hilda, & Anna Szabolcsi (1998). *Verbal complexes*. UCLA. in press.

Kracht, Marcus (1998). Strict compositionality and literal movement grammars. In *Proceedings of the Logical Aspects of Computational Linguistics, LACL'98* (pp. 126–142). NY: Springer.

Kracht, Marcus (2003). *Mathematics of Language*. Mouton de Gruyter, Berlin. in press.

Kroch, Anthony (1987). Unbounded dependencies and subjacency in a tree adjoining grammar. In Alexis Manaster-Ramer (Ed.), *The mathematics of language* (pp. 143–172). John Benjamins.

Kroch, Anthony (1989). Asymmetries in long distance extraction in a tree adjoining grammar. In Mark Baltin, Anthony Kroch (Eds.), *Alternative conceptions of phrase structure* (pp. 66–98). Chicago University Press.

Kurisu, K., & Sanders, N. (1999). Infixal nominal reduplication in Mangarayi. *Phonology at Santa Cruz*, *6*, 47–56.

Kuroda, S.-Y. (1964). Classes of languages and linear-bounded automata. *Information and Control*, *7*, 207–223.

Lappin, Shalom (1996). The interpretation of ellipsis. In Shalom Lappin (Ed.), *The handbook of contemporary semantic theory*. Oxford: Blackwell.

Lewis, H.R., & C.H. Papadimitriou (1981). *Elements of the theory of computation*. Englewood Cliffs, New Jersey: Prentice-Hall.

Li, Hui-Ju Grace (2002). *Ellipsis Constructions in Chinese*. Ph.D. Thesis, University of Southern California.

Li, Yen-hui Audrey (1992). *Order and constituency in mandarin Chinese*. Boston: Kluwer.

Manaster-Ramer, Alexis (1986). Copying in natural languages, context freeness, and queue grammars. In *Proceedings of the 1986 meeting of the association for computational linguistics*.

Mandelbrot, Benoit (1961). On the theory of word frequencies and on related Markovian models of discourse. In Roman Jakobson (Ed.), *Structure of Language* in its Mathematical Aspect, *Proceedings of the 12th symposium in applied mathematics* (pp. 190–219). Rhode Island: American Mathematical Society, Providence.

Mateescu, Alexandru and Arto Salomaa (1997). Aspects of classical language theory. In G. Rozenberg, A. Salomaa (Eds.), *Handbook of Formal Languages* (Vol. 1, pp. 175–251) Word, language, grammar. NY: Springer.

McCawley, J. D. (1994). Remarks on the syntax of Mandarin yes-no questions. *Journal of East Asian Linguistics*, *3*, 179–194.

Michaelis, Jens (1998). Derivational minimalism is mildly context-sensitive. In *Proceedings of the logical aspects of computational linguistics, LACL'98*, NY: Springer.

Michaelis, Jens (2001). Transforming linear context free rewriting systems into minimalist grammars. In Philippe de Groote, Glyn Morrill, Christian Retor rsquo;e (Eds.), *Logical aspects of computational linguistics* (pp. 228–244). Lecture Notes in Artificial Intelligence, No. 2099, NY: Springer.

Michaelis, Jens and Marcus Kracht (1997). Semilinearity as a syntactic invariant. In Christian Retoré (Ed.), *Logical aspects of computational linguistics* (pp. 37–40). NY: Springer-Verlag (Lecture Notes in Computer Science 1328).

Miller, George A., & Noam Chomsky (1963). Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (Eds.), *Handbook of mathematical psychology*, (Vol. II, pp. 419–492). Wiley, NY.

Nakanishi, Ryuichi, Keita Takada, and Hiroyuki Seki (1997). An efficient recognition algorithm for multiple context free languages. In *Proceedings of the fifth Meeting on Mathematics of Language, MOL5*.

Nunes, J. (2001). Sideward movement. *Linguistic Inquiry*, *32*, 303–344.

Peters, P. S., & Ritchie, R. W. (1973). On the generative power of transformational grammar. *Information Sciences*, *6*, 49–83.

Radzinski, D. (1990). Unbounded syntactic copying in Mandarin Chinese. *Linguistics and Philosophy*, *13*, 113–127.

Radzinski, D. (1991). Chinese number names, tree adjoining languages and mild context sensitivity. *Computational Linguistics*, *17*, 277–300.

Retoré, Christian, & Roberto Bonato (2001). Learning rigid Lambek grammars and minimalist grammars from structured sentences. In L. Popel rsquo;insk rsquo;y, M. Nepil (Eds.), *Proceedings of the third learning language in Logic Workshop, LLL3* (pp. 23–34). Brno, Czech Republic. Faculty of Informatics, Masaryk University. Technical report FIMU-RS-2001-08.

Rizzi, Luigi (2000). Early null subjects and root null subjects. In Luigi Rizzi (Ed.), *Comparative syntax and language acquisition* (pp. 299–319). NY: Routledge.

Ross, John R. (1967). *Constraints on Variables in Syntax*. Ph.D. Thesis, Massachusetts Institute of Technology.

Sag, Ivan, Timothy Baldwin, Francis Bond, Ann Copestake, and Dan Flickinger (2002). Multiword expressions: A pain in the neck for NLP. In *Proceedings of the third International Conference on Intelligent Text Processing and Computational Linguistics (CICLing-2002)* (pp. 1–15). Mexico City, Mexico.

Seki, H., Takashi, M., Mamoru, F., & Tadao, K. (1991). On multiple context-free grammars. *Theoretical Computer Science*, *88*, 191–229.

Shapiro, L. P., & Hestvik, A. (1995). On-line comprehension of VP-ellipsis: Syntactic reconstruction and semantic influence. *Journal of Psycholinguistic Research*, *24*, 517–532.

Shieber, S. M. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, *8*(3), 333–344.

Stabler, Edward P. (1997). Derivational minimalism. In Christian Retoré (Ed.), *Logical Aspects of Computational Linguistics* (pp. 68–95) (Lecture Notes in Computer Science 1328). NY: Springer-Verlag.

Steedman, Mark J. (1996). *Surface structure and interpretation*. Cambridge, MA: MIT Press.

Tanenhaus, M. K., & Carlson, G. N. (1990). Comprehension of deep and surface anaphora. *Language and Cognitive Processes*, *5*, 257–280.

Torenvliet, Leen, & Marten Trautwein (1995). A note on the complexity of restricted attribute-value grammars. In *Proceedings of computational linguistics in the Netherlands* (pp. 165–180). CLIN5.

Trautwein, Marten (1995). The complexity of structure-sharing in unification-based grammars. In *Proceedings of computational linguistics in the Netherlands, CLIN5*, pages 165–180.

Vijay-Shanker, K., & Weir, D. (1994). The equivalence of four extensions of context free grammar formalisms. *Mathematical Systems Theory*, *27*, 511–545.

Webber, Bonnie (1978). *A Formal Approach to Discourse Anaphora*. Ph.D. Thesis, Harvard University.

Webber, Bonnie (2004). D-LTAG: Extending lexicalized TAG to discourse. *Cognitive Science*, this issue.

Weir, David (1988). *Characterizing mildly context-sensitive grammar formalisms*. Ph.D. Thesis, University of Pennsylvania, Philadelphia.

Wu, Jianxin (1997). A model-theoretic approach to A-not-A questions in Chinese. In *Proceedings of the 21st Penn Linguistics Colloquium*. Philadelphia.

Zhang, Z.-S. (1997). Focus, presupposition and the formation of A-not-A questions in Chinese. *Journal of Chinese Linguistics*, *25*(2), 463–477.

Zipf, George K (1949). *Human behavior and the principle of least effort: An introduction to human ecology*. Boston: Houghton-Mifflin.