ELSEVIER

Cognitive Science 28 (2004) 669–697

COGNITIVE SCIENCE

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

Restricting grammatical complexity

Robert Frank

Department of Cognitive Science, Johns Hopkins University, 3400 N. Charles Streeet, Baltmore, MD 21218, USA

Received 16 June 2003; received in revised form 10 December 2003; accepted 18 March 2004

Available online 21 July 2004

Abstract

Theories of natural language syntax often characterize grammatical knowledge as a form of abstract computation. This paper argues that such a characterization is correct, and that fundamental properties of grammar can and should be understood in terms of restrictions on the complexity of possible grammatical computation, when defined in terms of generative capacity. More specifically, the paper demonstrates that the computational restrictiveness imposed by Tree Adjoining Grammar provides important insights into the nature of human grammatical knowledge.

Keywords: Grammar; Computation; Tree Adjoining Grammar; Syntax; Generative Capacity; Complexity

1. Introduction

Many theories of natural language syntax characterize human grammatical knowledge in computational terms. In such theories what a speaker knows is a system of abstract computation for constructing the possible sentences of his or her language. I will argue in this paper in favor of such a computational perspective on linguistic knowledge, by showing how important properties of grammar can be derived from computational restrictions. To do this, it is necessary to be precise about what kinds of computational restrictions must be imposed. In Section 2, I review a number of options, ultimately fixing on abstract generative capacity. In Sections 3 and 4, I explore the alignment between the specific restrictions on generative capacity found

E-mail address: rfrank@jhu.edu (R. Frank).

0364-0213/$ – see front matter
doi:10.1016/j.cogsci.2004.03.006

in the Tree Adjoining Grammar (TAG) formalism (Joshi, Levy, & Takahashi, 1975) and the characteristics of human grammar. Specifically, I demonstrate in Section 4.1 that the existence of "island effects" in displacement phenomena follows from the local manner in which TAG requires syntactic dependencies to be encoded (what I refer to as the "Fundamental TAG Hypothesis") and limitations on the combinatory machinery that TAG provides. In Section 4.2, I explore a potentially negative consequence of TAG's limited computational power: the restrictive TAG machinery also apparently predicts the impossibility of a certain type of construction that is present in Hungarian. Closer analysis of this construction, however, reveals a number of curious properties that point to an analysis that is within the limits of the TAG formal system. Rather than being a counter-example, then, this construction provides further evidence for the alignment between the properties of natural language and the computational restrictions imposed by the TAG formalism.

2. Characterizing complexity

The human language capacity provides us with the ability to represent and compute mappings between semantic content and some kind of form (sound, orthography, etc.). Individuals who represent the same or a sufficiently similar mapping are able to share their ideas, a useful trait. Formal approaches to the cognitive science of language start from the assumption that an individual's knowledge of such a form-meaning mapping takes the form of an autonomous mental grammar. Under this conception, a grammar is a means of specifying a set of pairings of forms and meanings. Any theory of human grammar must come to terms with the apparent conflict between the infinite productivity of the form-meaning pairing and the finiteness of our brains. As is familiar to students of cognitive science, the key to the resolution of this conflict came with developments in mathematical logic and the theory of computation during the early part of the twentieth century, which for the first time provided a variety of means for precisely specifying such an infinite set in finite terms.

Over its 50-year history, generative grammar has embraced this formalist perspective, and has been preoccupied with determining the nature of mental grammar. Much of the work toward this end has been empirical, attempting to establish the range and limits of possible grammatical structures both within a single language as well as to identify linguistic universals that restrict variation across languages. Given the commitment to providing a precise characterization of grammar, this empirical work has necessarily gone hand in hand with attempts to identify formal models for expressing these structures and mappings. Indeed we should expect that in an explanatory theory of grammar the very structure of this formal model should play a significant role in restricting grammatical variation. In this, I endorse the view put forward by Gazdar, Klein, Pullum, and Sag (1985, p. 2):

> The most interesting contribution a generative grammar can make to the search for universals of language is specify formal systems that have putative universals as *consequences*, as opposed to merely providing a technical vocabulary in terms of which autonomously stipulated universals can be expressed.

One thread that has united many of the formal models that have been proposed is that they consist of a kind of computation. What I mean by this is that the form-meaning relation is characterized in terms of the input and output of some sequence of processing steps. This can, and indeed has been done in a variety of ways. The model proposed by Chomsky (1965), for instance, characterizes grammar in terms of a process that takes representations of meaning as input and return representations of form as output. In other models, for instance those of Steedman (1996) and Chomsky (1995), grammatical computation takes as input a set of lexical resources and produces representations of meaning and form as output. Whatever its nature, we might expect, following Gazdar et al., that there will be properties of this computation that play an important role in our understanding of grammatical variation. Moreover, as grammatical computation is assumed in the generative paradigm to constitute a speaker's knowledge of her grammar, her *grammatical competence*, understanding the character of this computation is important for cognitive science. Under the assumption that cognition in general is a form of computation, an understanding of the kinds of computation exploited by grammar may be very revealing about the nature of cognition.

In order to address the question of grammatical complexity in a substantive way, we need to be explicit about how we should measure it. One intuitively appealing way to do this is in terms of the effects that a particular conception of grammatical computation has for language use. We might, for instance, measure the complexity of a grammar by determining the complexity of the parsing problem for that grammar, that is, by measuring how much time or space is needed to determine for a given surface form the range of interpretations that the grammar permits. The relevance of complexity on this measure seems clear: since people are apparently able to comprehend language in an effortless fashion, we might expect that grammatical computation should be structured in such a way as to limit the complexity of the parsing problem (or perhaps analogous problems in the domain of production). There are, however, significant obstacles to this view. Note first that there are classes of structures that are well-formed according to the grammar but which the performance system cannot efficiently manipulate, for reasons of memory load say, and which speakers as a result avoid entirely and do not accept—center embedded structures are a case in point (Miller & Chomsky, 1963). It is of course a non-trivial matter to determine whether the acceptability of a certain structure is due to the fact that such a structure is not generated by the grammar or because it imposes unmanageable processing demands (Joshi, Becker, & Rambow, 2000). However, to the degree that the characterization of grammatical well-formedness and processability are qualitatively different, the simplest account of the range of possible sentences will be one that separates restrictions on grammar from restrictions on use (Hale & Reiss, 2000; Miller & Chomsky, 1963), and the properties of that account will not be determined by the exigencies of use. An additional reason for skepticism about the importance of processing-derived measures for grammatical complexity stems from the directional manner in which the grammatical mapping is often stated, say from an underlying meaning say to a surface form. In such a system, there is no grammatical process corresponding to the parsing process, since the latter must search among possible inputs to the grammatical computation. For this reason, it has sometimes been proposed that in comprehending sentences, the language processor initially exploits a variety of computational tricks, external to the grammatical specification of the form-meaning mapping (Townsend & Bever, 2001). If this view is correct, understanding the properties of such tricks would surely tell us little about the character of grammar.[1] Putting aside complexity measures tied to language use, then, how else might we assess grammatical complexity?[2] One possibility exploits the notion of generative capacity, which characterizes the complexity of a class of grammars in terms of the languages it is capable of generating. If we take a grammar as generating a set of strings, the resulting complexity measure is that of *weak generative capacity*. This is to be distinguished from the notion of *strong generative capacity*, where a grammar class is characterized in terms of the sets of structural descriptions it can generate. From the perspective of grammatical theory, where it is crucial that a sentence be assigned the correct structure, strong generative capacity would seem to be the more important notion; we may well want to distinguish between two grammars that are weakly equivalent, in the sense of being able to describe the same sets of sentences, but not strongly equivalent, since they define different sets of structural descriptions. However, it has proven to be difficult to compare different types of grammars along the dimension of the structures that they generate, since these structures are often qualitatively distinct. As we shall see, there is nonetheless considerable mileage to be gotten from the notion of weak generative capacity. In the foundational paper on the topic of generative capacity, Chomsky (1959) defines four classes of grammars and compares them according to the languages they could generate, where language is understood as a set of strings (hence we are dealing with weak generative capacity). These classes of grammars are all systems of string rewriting: given some initial start symbol, grammatical computation consists of rewriting the string according to the rules given by the grammar. Computation continues so long as the string includes at least a single symbol from a distinguished subalphabet of *non-terminal symbols* and concludes once the string contains only *terminal symbols*. The different classes of grammars are distinguished on the basis of the types of rules they permit:

(1) Type 0 (Unrestricted) grammar: rules are of the form $\alpha X\beta \rightarrow \gamma$
Type 1 (Context-sensitive) grammar: rules are of the form $\alpha X\beta \rightarrow \alpha\gamma\beta$, where $\gamma \neq \epsilon$
Type 2 (Context-free) grammar: rules are of the form $X \rightarrow \gamma$
Type 3 (Right linear) grammar: rules are of the form $X \rightarrow wX$ or $X \rightarrow w$

In these definitions X is a single non-terminal symbol, α, β, and γ are strings over the set of non-terminal and terminal symbols, and w is a string over the set of terminal symbols. Chomsky shows that the generative capacity of each of these grammatical classes is strictly increasing: not only can each class of grammars in (1) generate all the languages generable by the classes of grammars listed below it, there are languages which can be generated by grammars of each class that cannot be generated by any of the grammar types appearing below it. That is, there are languages that can be generated by some context-free grammar but not by any right linear grammar, there are languages that are generated by some context-sensitive grammar but not by any context-free grammar, and there are languages that are generated by some unrestricted grammar but not by any context-sensitive grammar.

Consider, for instance, the language consisting of a string of *a*s followed by a string of *b*s. This language is generated by a right linear grammar with the following four rules:

1. $S \rightarrow aS$
2. $S \rightarrow aB$
3. $B \rightarrow bB$
4. $B \rightarrow b$

To generate a string with this grammar, we begin with the symbol S, successively rewriting according to one of these rules until we cannot rewrite any longer (the number above the arrow reflects the rewriting rule used at the stage in the computation):

$$S \overset{1}{\Rightarrow} aS \overset{1}{\Rightarrow} aaS \overset{2}{\Rightarrow} aaaB \overset{3}{\Rightarrow} aaabB \overset{4}{\Rightarrow} aaabb$$

Such a derivation can be written for a string with any number of *a*s and *b*s. Observe, though, that the *a*s and *b*s in such a string are generated separately from one another, and the grammar does not encode any dependencies among them. However, if we want to ensure that the number of *a*s and *b*s in the string are the same, so that our grammar generates exactly the language $\{a^n b^n | n \in N\}$, right linear grammars are no longer sufficient. We can, however, write a context-free grammar that will generate this language:

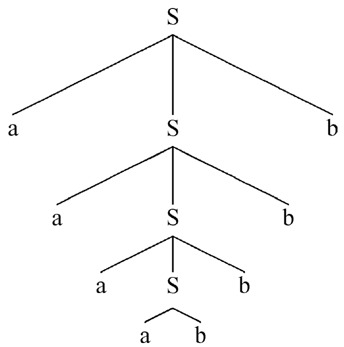
1. $S \rightarrow aSb$
2. $S \rightarrow ab$

Generating strings with this grammar involves the simultaneous addition of *a*s and *b*s to the right and left of the center of the string, and consequently, they will remain equal in number.

$$S \overset{1}{\Rightarrow} aSb \overset{1}{\Rightarrow} aaSbb \overset{1}{\Rightarrow} aaaSbbb \overset{2}{\Rightarrow} aaaabbbb$$

In this derivation, there are indeed dependencies between the *a*s and *b*s: as just noted these elements are generated in a nested fashion, so that the first *a* is dependent on the last *b*, the second *a* on the penultimate *b*, and so on. This is especially clear when we represent this derivational history in terms of the following tree structure:

(2)



Each node in this *derivation tree* corresponds to a point in the rewriting sequence just given. If two elements are sisters at some level of the tree, this implies that they are derivationally dependent on one another, having been introduced at the same point in the derivation. One might imagine that dependencies could work differently. Suppose for instance that we wanted to generate strings where the dependencies worked in a "crossing" manner, so that the first element from the left part of the string was associated with the first element from the right part of the string. Of course, the strings of such a language would be equivalent, what is known as *weak equivalence*, though their derivations and dependencies would not be, so that they are not *strongly equivalent*. There are languages whose strings require that they show crossing dependencies. One example is the language whose members consist of two concatenated copies of a single string of *a*s and *b*s, i.e., $\{ww|w \in \{a, b\}^*\}$. Crossing dependencies, whether evident in the string language or not, cannot be generated by a context-free grammar, but can by a context-sensitive grammar.[3] To finish the picture, there are, as I said above, languages that are generable by unrestricted grammars, though not by context-sensitive grammars. It turns out that simple examples are not especially easy to come by, and indeed most often are the product of somewhat complex constructions. Nonetheless, we can conclude that the classes of grammars in (1) indeed form a hierarchy in terms of generative capacity. This hierarchy has come to be known as the *Chomsky hierarchy*.

What makes the generative capacity distinctions within the Chomsky hierarchy especially interesting as a probe into the complexity of grammatical computation is the correspondence that has been discovered between these grammar classes and the type of abstract computational device, or *automaton*, that is needed to accept the same languages.[4] The class of abstract computers with finite memory, so-called *finite state automata*, accept precisely the same class of languages as right linear grammars. If we add the infinite memory capacity provided by a push-down (last in-first out) stack, resulting in a *push-down automaton*. we get a class of automata which accept precisely the same class of languages generated by context-free grammars. To accept the context-sensitive languages, we require a still more complex sort of computer, involving a memory tape that can be read from and written to in any order, but whose length is bounded as a linear function of the length of the input string, a *linear bounded automaton*. Finally, unrestricted languages exhaust the power of a *Turing machine*, the most general notion of computation available following Church's thesis. Understanding the generative capacity of some conception of grammar, then, tells us about the nature of the abstract device that mental computation exploits.

What good is a characterization of the generative capacity of a certain conception of grammar? One motivation is so that we can compare it with the sorts of complexity witnessed in natural language. For if we can find natural languages that exhibit dependencies requiring a certain degree of grammatical complexity, we can be certain that a grammatical framework lacking this sort of power is inadequate. Arguments to this effect date back to Chomsky (1957), and have been oft-repeated. Chomsky and Miller (1963), for instance, note that the English sentence in (3) exhibits dependencies (among subscripted elements) that have the same sort of nesting observed in (2).

(3) Anyone$_1$ who feels that if$_2$ so-many$_3$ more$_4$ students$_5$ whom we$_6$ haven't$_6$ actually admitted are$_5$ sitting in on the course than$_4$ ones we have that$_3$ the room had to be changed, then$_2$ probably auditors will have to be excluded, is$_1$ likely to agree that the curriculum needs revision.

Since right linear grammars/finite state automata are insufficient for the formal case, Chomsky argues that so too are they insufficiently powerful to characterize natural language.[5] More recently, Shieber (1985), Culy (1985), and Huybregts (1985) have shown that even context-free grammars/push-down automata are insufficiently powerful to capture the regularities of natural language. These arguments derive from grammatical structures like those seen in Swiss German:

(4)

Jan säit das mer d'chind em Hans es huus lönd hälfe aastriiche
John said that we the children-acc Hans-dat the house-acc let help paint

'John said that we let the children help Hans paint the house.'

In this sentence, noun phrase objects are related to the verbs which determine their morphological case in a crossing fashion, a kind of dependency that is not generable with context-free grammars. Following the Chomsky hierarchy to its next rung, we are left with the possibilities that natural languages are in the class of context-sensitive languages, unrestricted languages, or neither.

The central goal of generative linguistics is the construction of restrictive theories of grammar. Yet, if we are hoping to use computational restrictiveness in the guise of generative capacity as a means of limiting the range of possible grammars, it appears that the Chomsky hierarchy has led us to a dead end. If natural language grammars are not context-free, the only remaining candidates for possible grammatical classes are systems so powerful as to be essentially unrestrictive. This need not be the end of the story, however. If one could provide computational characterizations for restrictive language classes that fell between those provided by context-free and context-sensitive grammars, but which were sufficiently powerful to deal with the range of natural language structures, we might be in the position to derive properties of natural language grammar. Joshi (1985) hypothesizes that such a class of languages, which he calls *mildly context-sensitive*, might be defined in terms of the following three properties:

(5) a. Generation of limited crossing dependencies (like those in (4))
b. Constant growth property: for every mildly context-sensitive language L, there is a bound k such for every sentence $w \in L$, there is another sentence $w' \in L$ such that $|w| < |w'| < |w| + k$.
c. Polynominal time parsing complexity

In turn, Weir (1988) has defined an infinite hierarchy of language classes, defined by a hierarchy of grammar classes called Linear Context Free Rewriting Systems (LCFRS), all mildly context-sensitive. The lowest rung in Weir's hierarchy has proven to be an especially interesting language class. As with the language classes in the Chomsky hierarchy, this class is not only defined by a particular type of rewrite system, but can also be characterized in terms of an automaton: the *Embedded Push-down Automaton* or EPDA (Vijay-Shanker, 1987; Vijay-Shanker & Joshi, 1985). This automaton is a variant of the push-down automaton whose memory consists of a stack of stacks rather than a simple stack of symbols, where symbols may be added and removed not only from the top of the (top) stack, but also from the top of stacks immediately below the top stack. Strikingly, it has turned out that these languages are also characterized by a variety of formal systems, including Combinatory Categorial Grammars, Head Grammars, Linear Indexed Grammars and Tree Adjoining Grammars (Joshi, Vijay-Shanker, & Weir, 1991). The fact that so many different grammar formalisms converge on the same degree of formal power suggests that there is something about this degree of complexity that fits naturally with linguistic description. If this is correct, we should minimally expect that this class is sufficiently powerful to characterize the structures of natural language. But even more interestingly, we should expect to find empirical consequences of the limited formal complexity of this class.

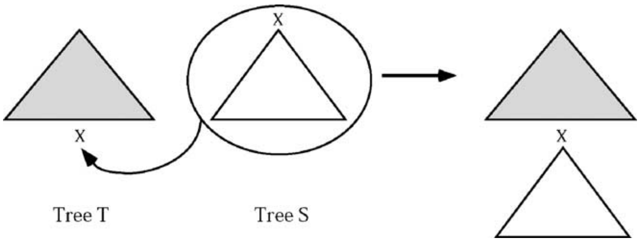
In the remainder of this paper, I will attempt to draw out consequences of precisely this sort. I will focus my investigation specifically on the grammatical consequences of adopting one specific mildly context-sensitive grammatical formalism, Tree Adjoining Grammar (TAG), as part of a grammatical theory. Though my arguments I make will make specific reference to the properties of TAG, it is my expectation that they will carry over in some form to other formalisms with the same formal power. One should keep in mind, however, that while the weak generative capacity of these formalisms is equivalent, it is much less clear to what degree they are equivalent in terms of the derivations and dependencies that they generate, that is, their strong generative capacity. It would be surprising if complexity considerations concerning strong generative capacity were not also relevant to understanding the nature of grammatical computation, though at present it has proven difficult to characterize strong generative capacity in a sufficiently abstract fashion (Miller, 1999).

As a final introductory point before I move on to the main point of the paper, I want to stress the importance of keeping two notions of grammatical complexity distinct: the characterization of the type of automaton needed to recognize the elements generated by the grammar and the properties of an algorithm that recognizes the same set of elements. For any type of automaton, there is an upper bound on the time complexity of an algorithm that accepts the strings accepted by any automaton of that type. For finite state automata, the process is linear in the size of the input, while for push-down automata it is cubic in the size of the input, i.e., $O(n^3)$ for a sentence of length n. Nonetheless, there are languages, $\{a^n b^n | n \in N\}$ for example, which require a push-down automaton, i.e., they are properly context-free, but which can nonetheless be accepted by an algorithm whose time complexity is linear: one only need to count the number of *a*s and *b*s and check that these are the same. For this reason, the generative capacity metric has at times been maligned as being insufficiently sensitive to subtle differences in grammatical complexity. An alternative approach, taken by Barton, Berwick, and Ristad (1987), uses the time complexity for the problem of recognizing the strings from some grammar. This question recalls the use of parsing complexity discussed above, but Barton et al. propose their measure as a way of assessing the abstract complexity of grammatical computation, without appeal to problems of language processing. Note, however, that it is an empirical question as to which of these measures of grammatical complexity turns out to be relevant for the purposes of constraining grammar in an empirically supported fashion. In the case of TAG, a considerable variety of arguments have been adduced to in support of the claim that TAG's limited generative capacity does indeed constrain grammars in ways that rule out unattested linguistic systems. These arguments are discussed extensively in Frank (2002) and I briefly review some of them below. However, so far as I am aware, it remains to be shown that limiting the complexity for the recognition problem, to linear time say, produces any linguistically significant restrictions on grammars. We are left then with generative capacity as the only empirically significant metric for grammatical complexity.

3. An introduction to TAG

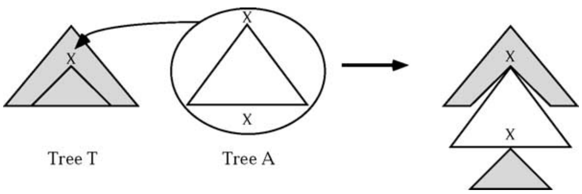
In contrast to the string rewriting systems of the Chomsky hierarchy, TAG (Joshi, 1985; Joshi et al., 1975) is a system of *tree rewriting*. Rather than beginning with a start symbol, the derivation begins with a piece of tree structure, called an *initial tree*, that is specified as part of the grammar. The derivation proceeds by replacing individual nodes in this tree with one of the small pieces of tree structure, the *elementary trees*, which constitute the rules of the grammar. In turn, nodes in this more complex structure can be rewritten, until the derivation concludes. Rewriting in a TAG can take place in one of two modes. The first of these, *substitution*, involves the rewriting of a non-terminal node at the frontier of one elementary tree as another elementary tree with the requirement that rewritten node must have the same label as the root of the elementary tree that rewrites it. This is depicted schematically in (6).

(6)



The second mode of tree rewriting, and the one from which TAG derives its name, is *adjoining*. In this operation, a non-terminal node anywhere within an elementary tree is rewritten as another elementary tree, called an *auxiliary tree*. This operation is depicted schematically in (7).

(7)



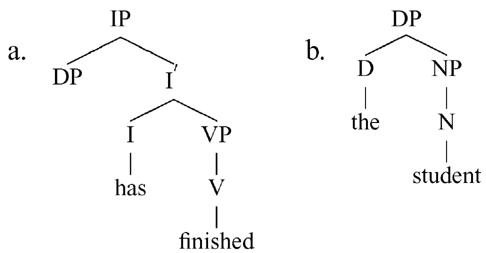
Note that in order for this latter operation to be well-defined, auxiliary trees must include a distinguished frontier node, called the *foot node*, which will, following adjoining, dominate the structure that was dominated by adjoining site prior to adjoining. In order to guarantee that local structural relations among nodes in the original elementary tree are preserved, including for instance parent-child or sisterhood, the foot node of an auxiliary tree must be labeled identically to the root.

The presentation of TAG I have given to this point conceives of elementary trees as the rules of the grammar that specify how nodes can be rewritten, in either the substitution or adjoining modes. TAG is typically presented in a different fashion, where adjoining and substitution are operations for tree combination, so that TAG derivation involves the successive combination of elementary trees using these two operations rather than the iterated rewriting of elementary trees. These two conceptions are essentially equivalent. There is, however, one important property of TAG derivations that follows directly from the rewriting conception that is not immediately obvious on the combinatory conception. This property concerns the source of recursive structures that can be used during adjoining. If we think of adjoining as an operation involving recursive structures, the question arises of whether we can compose two non-recursive elementary trees to produce a recursive tree, which can in turn be used as an auxiliary tree. In contrast, under a rewriting conception of the grammar, it is the elementary trees, qua grammar rules, that license each step of rewriting during a derivation. As a result, any instance of adjoining must necessarily involve an auxiliary tree among those specified by the grammar. To preserve the equivalence between the two conceptions, we require that each derivational step of adjoining or substitution must relate a node of one tree and an elementary tree that is substituted or adjoined at that node.

As a tree rewriting system, the output of a TAG derivation is itself a tree structure. This is unlike the situation with string rewriting systems where the tree typically associated with a sentence is its derivation tree. The history of a TAG derivation can also be represented as a derivation tree, in which each node corresponds to an elementary tree T whose children represent elementary trees that have themselves substituted or adjoined into T. While the tree that results from the TAG derivation determines the phonological form of the sentence, it has been argued by Kallmeyer and Joshi (2003) that it is the derivation structure that is the object that feeds semantic interpretation.

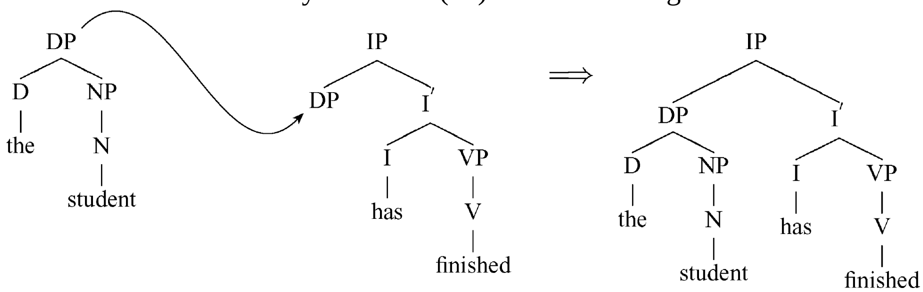
While TAG provides a set of combinatory operations, it does not provide all that we need to produce a grammatical theory. Specifically, any linguistic application of TAG will need to make substantive (as opposed to merely formal) claims about the nature of elementary trees. For example, a TAG-based grammatical theory must specify what information is encoded in the structure of the trees, what the range and content of category labels is, how much structure can be contained within a single elementary, etc. There have been a variety of proposals regarding these questions (Abeillé, 1993; Carroll, Nicolov, Shaumyan, Smets, & Weir, 2000; Frank, 2002; Hegarty, 1993; Kasper, Kiefer, Netter, & Vijay-Shanker, 1995; Kroch, 1989; Joshi & Kulick, 1997). In my work on the topic and in the rest of this paper, I assume that elementary trees are built around a single lexical element, that is, a semantically contentful word like a noun, verb or adjective. I assume that elementary trees localize the expression of predicate-argument relations, so that, for instance, the elementary tree headed by a verb includes the syntactic expression of all of the verb's arguments, a slot into which the argument can be inserted. In addition, I assume that each elementary tree may include the grammatical elements associated with the lexical head of the elementary tree, what Grimshaw (1991) calls the extended projection. So, a noun-headed elementary tree may include a determiner and some prepositions, a verb-headed elementary tree may include auxiliaries and complementizers. Examples of such trees are given in (8).

(8)



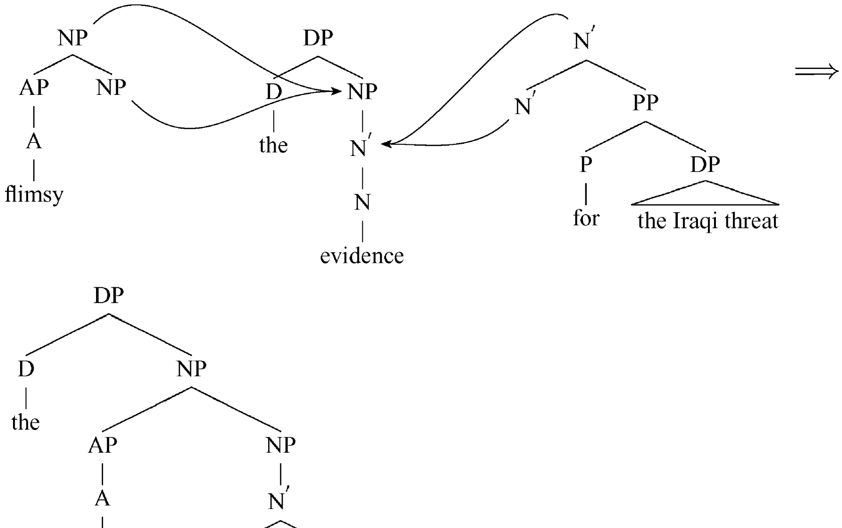
With these assumptions in place, the grammatical role played by the TAG operations now comes into focus. Substitution can be used, for instance, to fill the argument slots present in lexically-headed elementary trees. The subject of the verb *finished* in (8a), for instance, is filled by the *student*-headed elementary tree from (8b) in the following derivation:

(9)



The linguistic function of the adjoining operation is slightly different from that of substitution as a result of the recursive nature of auxiliary trees. Specifically, adjoining has a natural application to any structure in which there is structural recursion. One such case is modification. This is seen in the two instances of adjoining shown in (10).

(10)



These derivations illustrate two important properties that underlie all of the various conceptions of TAG-based grammatical theory. First, elementary trees constitute an extended domain of locality, that is, they include more structure that the domain typically described by a single rewrite rule in the grammars of the Chomsky hierarchy. The linguistic importance behind this idea lies in the following hypothesis about TAG-based grammatical description (Frank, 2002, p. 22):

(11) The fundamental TAG hypothesis: Every syntactic dependency is expressed locally within a single elementary tree.[6]

This hypothesis goes some way toward answering the question of what can constitute an elementary tree. What it leaves open, though, is what counts as a syntactic dependency and how such dependencies are best expressed, and these are the sources of divergence among different proposals. As mentioned above, I have assumed in my work that the notion of syntactic dependency invoked here includes the relation between a lexical head and its thematic arguments, even when such arguments have been displaced from their canonical positions, and the relation holding between a lexical head and its associated functional elements. These are illustrated in the trees given above: these elementary trees localize the dependencies between the lexical verbs and their arguments, as well as the dependencies between determiners and nouns, auxiliary elements and verbs, etc.

The fundamental TAG hypothesis is plausible only in the face of the other property that unifies all TAG-based grammatical theories, namely that syntactic dependencies become local only once recursion is factored from derived syntactic structures. This factoring of recursion is illustrated in (10): the determiner-noun dependency in the derived structure is locally expressible in the elementary tree only because the recursive modification structure is introduced by the adjoining operation. An operation analogous to adjoining that is capable of factoring out recursion from syntactic structure is missing from other lexicalized grammar formalisms such as Combinatory Categorial Grammar (Steedman, 1996), Head-Driven Phrase Structure Grammar (Pollard & Sag, 1994), and Lexical Functional Grammar (Bresnan, 2001). And it is precisely this absence that causes the fundamental TAG hypothesis to be a distinctive property of TAG, imposing substantive empirical constraints on possible grammatical analyses. In Section 4.2, I will explore the effects of one such constraint to derive distinctive, and what appear to be correct, predictions about Hungarian focus constructions.

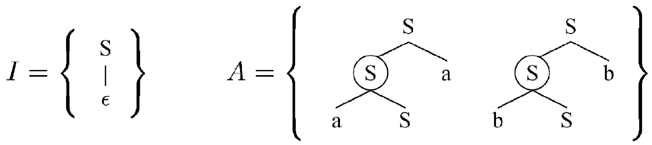
Crucially, the fundamental TAG hypothesis is limited to dependencies that are syntactic in nature. It imposes no analogous restriction, then, on the realization of *semantic* dependencies, such as that holding between a quantifier and a bound variable, or between coreferent elements. Such dependencies are well known to have rather different properties from syntactic dependencies, particularly with respect to the types of constraints discussed in Section 4.1 below. Therefore, it is an empirically justifiable move to avoid imposing on semantic dependencies the restrictions that TAG imposes on syntactic dependencies. Of course, semantic dependencies are not completely unconstrained and it is an important question how such dependencies can be established and their properties explained in the context of a TAG derivation. This topic has received attention recently (see, for example, Kallmeyer & Joshi, 2003; Romero, 2002), but many important issues remain open.

4. Formal restrictiveness and linguistic explanation

In Section 1, I mentioned the fact that the generative capacity of TAG, both in terms of the sets of strings and structures that it can generate, goes beyond that of context-free grammars. Thus, both of the following non-context-free languages can be generated by a TAG:

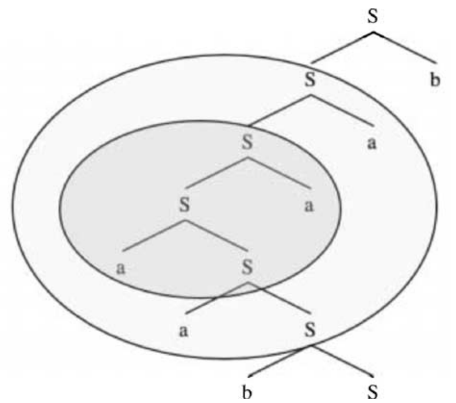
(12) $L_1 = \{ww | w \in a, b^*\}$
$L_2 = \{a^n b^n c^n d^n \ | n \in N\}$

To generate L_1, for instance, the requisite TAG consists of the following set of elementary trees, with initial tree I and auxiliary trees in A:



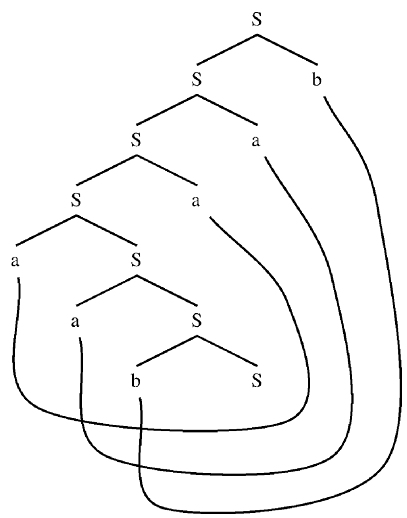
Derivations in this grammar proceed by adjoining an instance of one of the auxiliary trees into another at the circled S node, and then adjoining this result into yet another auxiliary tree.[7] As we do this, we produce a complex auxiliary tree with an identical string of either side of the spine that connects the root and foot nodes.

(13)



To finish the derivation, we adjoin this complex auxiliary that is derived into the S node of the initial tree, producing distinct sentences from L_1 depending on which auxiliary trees were adjoined and in which order. As shown in (14), the resulting trees exhibit the kind of cross-serial dependencies that are not generable by context-free grammars.

(14)



With this ability to capture such crossing dependencies comes the possibility of providing grammars for non-context-free structures, such as those seen in Swiss German.[8] The additional generative power that TAG provides beyond context-free grammars is, however, extremely limited. For instance, even though both of the following languages are easily generated by context-sensitive grammars, neither is generable by a TAG.

(15) $L_3 = \{www | w \in a, b^*\}$
$L_4 = \{a^n b^n c^n d^n e^n | n \in N\}$

Because of this very limited extension into the realm of the context-sensitive, the class of languages accepted by TAGs has been called the *mildly context-sensitive* languages. At present, it appears that there are no natural languages which require more generative power than that provided by TAG (though see, Becker, Rambow, & Niv, 1992; Bleam, 2000; Joshi et al., 2000; Kulick, 2000; Miller, 1991 for objections and refutations). In the absence of further evidence, then, it is plausible to assume that TAG is the most restrictive formalism capable of generating all natural languages, and as such constitutes a plausible candidate for the formal basis of human grammatical competence.

Yet, we might wonder what kind of role the restrictive complexity embodied in the TAG formalism will actually play in grammatical theory. In the remainder of this section, I will try to demonstrate the importance of TAG's limited complexity in two ways. First, I will argue that we can parlay this restrictiveness into substantive limitations on grammatical structures, concerning the locality of syntactic movement, in a way which accords with the empirical data. Secondly, I will attempt to show that the formalism limits the range of analytic options for certain grammatical constructions in empirically supported ways, focusing on the properties of Hungarian focus constructions. To the degree that these arguments are successful, we have support for the hypothesis that some portion of grammatical competence is best characterized in terms of computational restrictiveness.

4.1. Wh-questions and islands

As is well known, natural languages exhibit constructions in which elements appear displaced from their canonical positions. One example of such a construction is English questions. When an object is questioned, the wh-pronoun that replaces the object appears at the front of the sentence rather than in its usual post-verbal position.

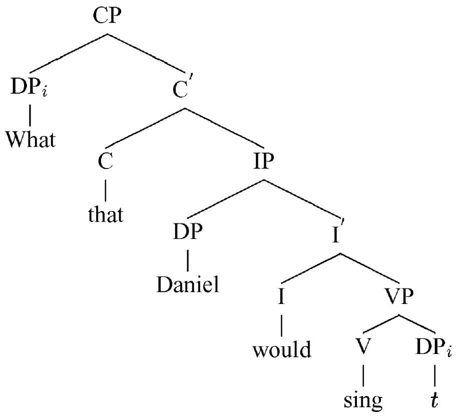
(16) a. What did Daniel sing?
b. Daniel sang that song.

This displacement of a wh-object can occur over unbounded distances. That is, by embedding the wh-element inside of a more complex sentence, the distance between the surface and canonical positions of the wh-object can grow without bound. Examples like the following illustrate this fact:

(17) a. What did you hear that Daniel would sing?
b. What did you hear that Gabriel thought that Daniel would sing?
c. What did you hear that Gabriel thought that I hoped that Daniel would sing?

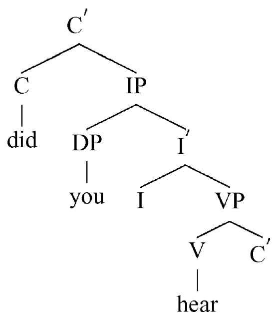
Recall that the fundamental TAG hypothesis requires that all syntactic dependencies be expressed within the local context of an elementary tree. The existence of such unbounded dependencies, then, would appear to pose a problem for TAG. However, by exploiting the adjoining operation, we can provide a natural account of examples like those in (17) (Frank, 2002; Kroch, 1987, 1989). Under this account, the dependency between the wh-element and its base position (or alternatively the verb with which it is semantically associated) is established within a single elementary tree, in accordance with the fundamental TAG hypothesis:

(18)

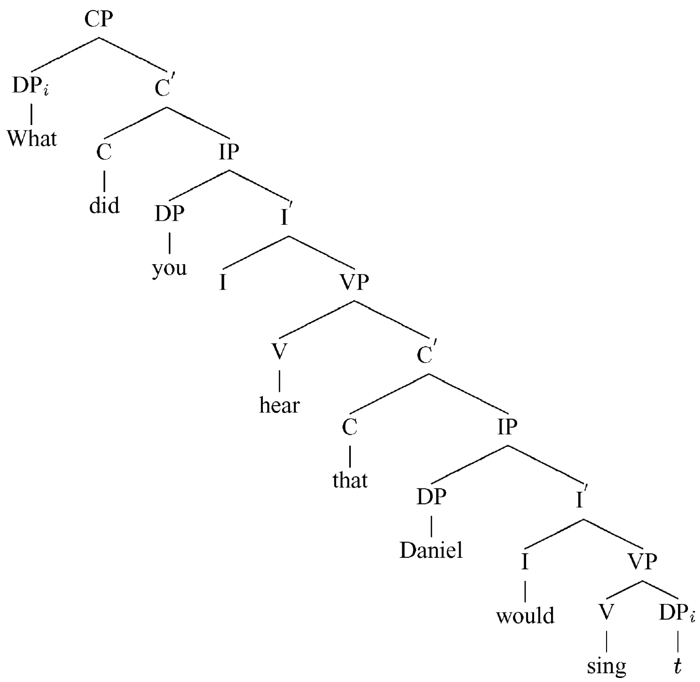


Given such a structure, we can accomplish the effect of dislocating the wh-phrase into higher clauses by exploiting the adjoining operation. Specifically, by adjoining the auxiliary tree in (19) to the C′ node in (18), we produce the sentence in (17a).[9]

(19)



(20)



Further instances of embedding, like those seen in (17), can be derived in an analogous fashion with additional instances of adjoining of such C′-recursive auxiliary trees.

One of the major discoveries of generative linguistics is that there are significant limitations on such unbounded dependencies. Specifically, there are a variety of structures, dubbed "islands" by Ross (1967), which do not permit elements whose canonical positions fall within them to move beyond their limits. Two examples of such islands are relative clauses and verb phrase-modifying prepositional phrases.

(21) a. Gabriel met the musician who would sing that song.
b. *What did Gabriel meet the musician who would sing?

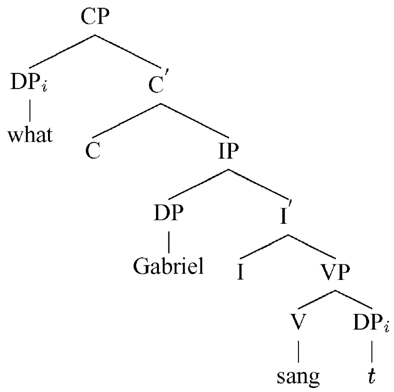
(22) a. Daniel laughed after Gabriel sang that song.
b. *What did Daniel laugh after Gabriel sang?

Substantial effort has been devoted to the characterization of the structures which constitute islands. Initial efforts in this direction consisted of catalogs of constructions that constituted islands. Subsequent work attempted to provide a general conditions on the syntactic movement operation (or its analog in other theories) that accounted for the distinction between island and non-island constructions. Yet, the very existence of such conditions raises a puzzle: In a theory in which dislocation is the result of a process that applies to already constructed syntactic structure, why should syntactic movement be restricted at all? One can easily imagine a grammatical system, differing only in the absence of such a locality condition. Of course, this would have the unfortunate empirical effect of eliminating the kinds of contrasts we see in (21) and (22). Nonetheless, it appears that there is something fundamental about the locality of dislocation:

so far as I am aware, every language that has been studied having English-type wh-questions shows precisely the same kinds of contrasts. Why then does human language work this way?

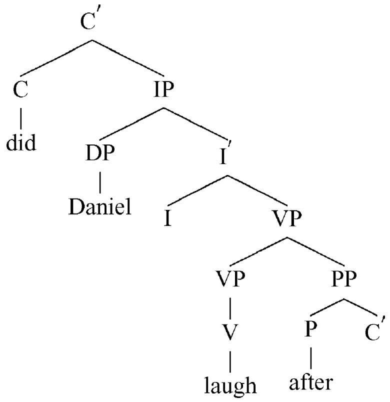
Strikingly, the TAG analysis of wh-movement just sketched provides an answer to this question. Specifically, as Kroch (1987) first pointed out, given certain assumptions concerning the nature of TAG elementary trees, the existence of island effects and of the subjacency condition in particular follow as corollaries of the TAG derivational system and need not be stated explicitly. To see how, consider the ungrammatical example (22b). Recall that the TAG analysis of wh-movement I have sketched requires that a wh-element be generated in the same elementary tree as the verb of which it is an argument. As a result, the wh-element must be generated within the adjunct clause headed by the verb *sing*.[10]

(23)



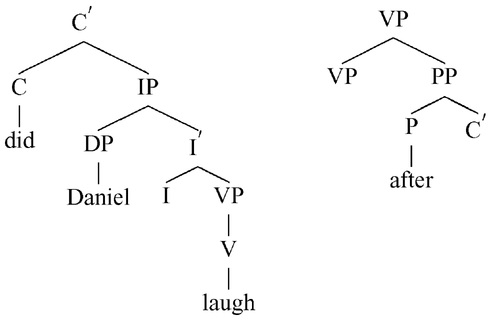
Now in order to front the wh-phrase to the front of the sentence, we must adjoin material into this structure, separating *what* from the rest of its clause. Assuming we adjoin to C′ as before, such a structure would be the following:

(24)



This structure meets the formal requirements for being an auxiliary tree: it has root and foot nodes with identical labels. Yet, this structure does not constitute a well-formed elementary tree from the perspective of our linguistic assumptions: it contains the syntactic structure licensed by two distinct lexical elements, the verb *laugh* and the preposition *after*. Each of these elements will separately head its own elementary tree:

(25)



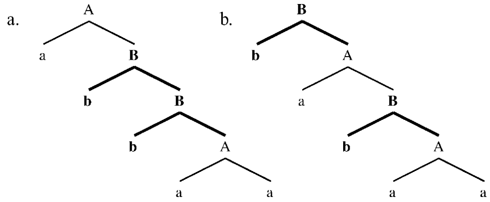
So, why can't we use these two separate trees to derive (22b), by combining them to produce a complex auxiliary tree that is then adjoined into (23)? Although it is certainly the case that these trees can combine by adjoining the VP-recursive *after*-headed auxiliary tree into the *laugh* tree, the result is not itself an auxiliary tree, in spite of the fact that it has the structure shown in (24). As mentioned above, adjoining, thought of as rewriting a node by a tree, is licensed by the basic set of auxiliary elementary trees given in the grammar.[11] Since there is no possible derivation for the structure, it is ruled out by the grammar. It can be shown that many other cases of island effects follow similarly (again see Frank, 2002; Kroch, 1987, 1989). Especially interesting is the result that those cases of islands for which there is cross-linguistic variation can be understood as resulting from independently necessary differences in what set of elementary trees different grammars permit. We see then that the TAG-based account of movement, when coupled with certain assumptions about the nature of elementary trees, enables us to understand why movement should show island effects, what constructions they will occur in, and where they will permit cross-linguistic variation. All of this follows without any explicit condition on syntactic movement, but instead from the way in which TAG combines local syntactic domains with the adjoining operation. The close alignment between the empirical patterns and the results of the formal system provide support, then, for the hypothesis that the restrictiveness of TAG constitutes part of human grammar.

4.2. Interleaving dependencies and derivational options

We have just seen that TAG enables us to generate cases of long-distance dependencies by the adjoining-in of intermediate material, and that the restrictiveness of the TAG machinery for structure composition derives some of the locality properties of such dependencies. One needs to remain vigilant in the face of such restrictiveness, however. For if our grammatical machinery prevents the derivation of ill-formed structures, we need to be sure that it does not similarly block structures which we want to permit.

The TAG derivation of long-distance dependencies yields the simple prediction that an element that is apparently dislocated from a lower clause to a higher one should always appear in a position at the periphery of the higher one. The reason for this is straightforward: the effects of interclausal movement result from the insertion of an elementary tree representing the higher clause between the dislocated element and the remainder of its clause via the adjoining operation. Because of the geometry of this combination, there is no way for elements of the higher clause to precede the dislocated element. This is shown schematically in the following pair of trees:

(26)



In these trees, nodes labeled B represent material that is part of the adjoined matrix clause's elementary tree, while nodes labeled A represent material that is part of the subordinate clause's elementary trees (into which adjoining takes place). The TAG combinatory operations lead us to predict that clauses should intermingle as in the structure on the left, but not the one on the right.

This prediction seems correct for English. Wh-elements moved out of a lower clause always appear preceding all of the material from the higher clause.[12] However, if we look a bit further afield, we see that other languages permit dependencies of this sort. In Hungarian, we see that both wh-pronouns and focused elements that are moved from one clause to another end up in precisely the sort of position the TAG analysis of dislocation we have been pursuing predicts should not be possible, namely interleaved between elements of the main clause.

(27) a. **János** kinek **akarta**, hogy adjuk a dijat? (A. Szabolsci, p.c.)
John to whom wanted that we give the prize

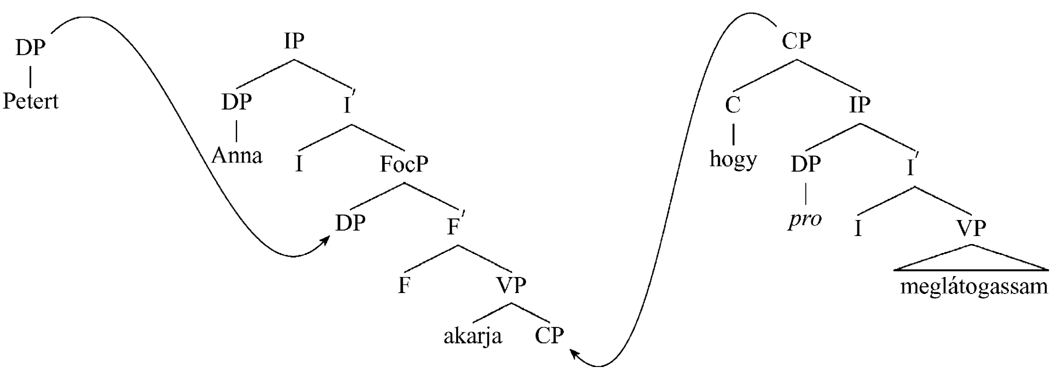
'Who did John want us to give the prize?'

b. **Anna** PETERT **akarja** hogy meglátogassam. (Kenesei, 1994, p. 317)
Ann Peter-acc wants that I visit

'It's Peter that Ann wants me to visit.'

One might take the existence of such constructions to argue against TAG as the right formalism for expressing natural language syntax. We need not jump to such a conclusion, however. In fact, there is a TAG derivation for these examples, just not one along which generates these dependencies along the lines discussed for English wh-question in the previous section. To generate the Hungarian examples, the position for the moved wh- or focused element would not be part of the lower clause's elementary tree, but rather the one of the matrix clause. An elementary tree representing the embedded clause with a phonologically null object (represented here as *pro*) could then be substituted into this structure.

(28)



Under the analysis depicted here, the relationship between the focused element in the main clause is not syntactically mediated. Instead I assume that it is established semantically, via a mechanism like the one implicated in establishing coreference between the main clause and embedded clause's subjects in sentences like the following:

(29) John tried [PRO to attend the rally]

Since the fundamental TAG hypothesis requires only that *syntactic* dependencies be expressed within individual elementary trees, the proposed analysis avoids running afoul of this hypothesis.

One potential problem for the analysis depicted in (27) is that it treats the Hungarian focus construction in a fundamentally different fashion from English wh-questions: one is derived from movement and the other involves a semantically mediated control-like dependency. In contrast, under the analyses of these constructions given in a transformational framework (cf., É. Kiss, 1987, inter alia), the English and Hungarian cases are treated as formally parallel, the only difference being the position in the matrix clause to which movement takes place. In fact, there is certain evidence that the Hungarian cases ought to be treated via syntactic movement as well: this construction exhibits certain properties that are characteristic of movement but not control, such as unboundedness (30a) and sensitivity to islands (30b).

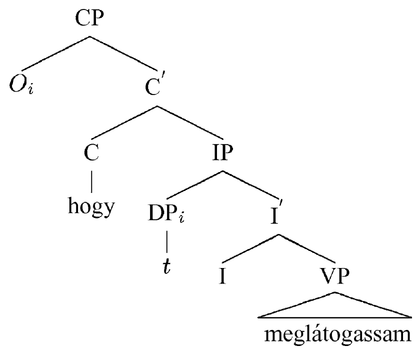
(30) a. János kit$_i$ gondol, hogy Mária 'szeretne, hogy meg hívjunk t_i? (É. Kiss, 1987, p. 126)
John whom thinks that Mary would like that perf we invite

'Who does John think that Mary would like us to invite?'

b. *János JULISKÁT$_i$ hallotam a hírt hogy elveszi feleségül t_i (É. Kiss, 1995, p. 224)
John Julie-acc heard-1sg the news that takes as wife

'As for John, it's Julie who I heard the news that John will marry.'

One way of reconciling these facts with the analysis given in (28) is to introduce movement of the an empty operator from the "base position" of the focused element to the front of its clause.

(31) 

As in (28), this tree can be substituted into the complement of a higher verb's elementary tree containing a focused element. From this position, the control dependency between the focused element and the moved empty operator is a local one, paralleling what is seen in control constructions such as (29). The unboundedness of the focus dependency can be derived by adjoining other trees to the C′ node in (31), as we did with wh-questions in the previous section.

And because of the parallelism between this derivation and the one we used for wh-questions, island effects for focus movement are derived just as before.

It is worth noting that there is nothing preventing a transformational framework from adopting the analysis for Hungarian that TAG forces upon us, in which the focused element is generated in its higher position. In fact, there are a number of other constructions involving apparent cases of displacement, but where it has been argued that the "displaced" element is generated in its surface position with movement of an empty operator from the gap position. One such case is the phenomenon of "*tough* movement," seen in (32a):

(32) a. This government is tough for me to support.
b. It is tough for me to support this government.

The synonymy of the pair in (32) gives intuitive appeal to the idea that these sentences are related via movement: specifically that (32a) is derived from the structure undelying (32b) via movement of the object *this government* from the lower object position to the matrix subject. Like wh-movement, this phenomenon is unbounded and shows sensitivity to islands:

(33) This government is tough to want to continue to support *e*

(34) a. *This government is tough to find someone who supports *e*.
b. *The massacre was tough to laugh after witnessing.

However, following Chomsky (1977), syntacticians have widely rejected this "direct dependency" analysis in favor of the kind of analysis I just sketched for the focus movement construction: the "tough-moved" subject is generated in the main clause, and is semantically linked to an empty question word that has been moved from the object position to the front of the infinitival complement of *tough*:

(35) This government is tough [O_i for me to support t_i]

There are a number of reasons for this move, stemming from a lack of total parallelism between the displaced and undisplaced versions of the tough construction as seen in (32). Note for instance that when a pronoun is in the base object position, it shows accusative (object) case, whereas it shows nominative (subject) case when it appears in the higher subject position.

(36) a. It is tough to get to know *them*.
b. *They* are tough to get to know.

Such a contrast is not characteristic of wh-questions, where the displaced wh-phrase retains the case marking it receives in its base position. This effect is only weakly visible in English, given the marked status of the form *whom*, but is robustly visible in languages with richer systems of case marking. This is illustrated in the following German examples:

(37) a. Wer/*Wen glaubst du, hat seinen Freund gesehen?
who-nom/whom-acc think has his friend seen

'Who do you think saw his friend?'

b. *Wer/Wen glaubst du, hat der Lehrer gesehen?
who-nom/whom-acc think you has the teacher seen

'Whom do you think the teacher saw?'

This contrast can be explained directly if the wh-phrase in questions is displaced from the position of the gap, therefore retaining the properties it had in that position, while the displaced subject in *tough* constructions never appeared in this position.

Another lack of parallelism between the displaced and undisplaced versions of the *tough* constructions concerns an interpretive contrast. Consider the following pair of examples (from Epstein, 1989):

(38) a. It is tough to talk to many people.
b. Many people are tough to talk to.

For (38a), there are two possible interpretations: this sentence can mean either (i) that there are many specific people, say Fred, Wilma, Barney, Betty and all of their friends and neighbors, who have the property of being difficult to talk to, or (ii) it is a difficult task to talk to many people, regardless of who these people are. This interpretive contrast can be represented by a difference in the scopal orderings for the quantifier *many* and the predicate *tough* in a predicate logicrepresentation. With *many* having scope outside of *tough* (i.e., there exist many x for which [tough(talk-to(x))]) we get the first interpretation. With the other scopal ordering where *many* takes scope inside of *tough* (i.e., tough(there exist many x for which [talk-to(x)])), we get the second interpretation. Observe now that in sentences in which the object undergoes tough movement, like (38a), only the first of these interpretations is possible, namely wide scope for *many*. Note, however, that other instances of displacement, specifically wh-movement, do not induce this interpretive effect.

(39) How many people is it tough to talk to?

In (39), where the quantifier *many* is part of a fronted wh-phrase, both interpretations are possible: when *many* has wide scope (i.e., what is the x such that there exist many x for which [tough(talk-to(x))]), the sentence is asking about the number of the set of people to whom it is difficult to talk. When *many* takes narrow scope, (i.e., what is the x such that tough(there exist many x for which [talk-to(x)])), the sentence is interpreted as asking about the size of a group which is difficult to talk to. Once again, we can explain the contrast between wh-questions and tough movement by assuming that the former but not the latter involves displacement from the position of the apparent gap in the object position, assuming the existence of a process like quantifier lowering (May, 1985) under which quantificational elements like *many* can be interpreted in their pre-displacement positions. If the apparently displaced subject in *tough* constructions never occupied a position in the lower clause, quantifier lowering will be of no use, and we will be stuck with the wide scope interpretation.

We see, then, that when faced with an apparent instance of dislocation, the formally unrestricted machinery of transformational grammar (Peters and Ritchie, 1973) admits analyses involving either movement or base generated dislocation. Choosing which of these is involved in any particular case requires careful empirical study, and would be expected in the transformational framework to vary from case to case. In contast, for cases of dislocation in which we see the kind of clausal interleaving illustrated in (26b), TAG admits only the possibility of base generation. Such restrictiveness on the part of TAG is potentially advantageous, assuming it leads us to choose the correct analysis, since it allows us to eliminate alternatives without explicit stipulation. Thus, if it turns out that base generation analysis is correct for the Hungarian dislocations we have been examining, it provides evidence that the limitations TAG imposes on grammatical analysis mesh well with the properties of natural language, suggesting once again that the formal restrictions embodied in TAG do indeed constitute part of the human language faculty.

So what *is* the right analysis for Hungarian? As a first step toward finding the answer, let us see how this construction fares with respect to distinctive properties of *tough* movement we have been discussing. Turn first to the issue of case marking. Subjects of Hungarian sentences are typically marked with nominative case. However, as É. Kiss (1987) notes, when a subject from an embedded clause is focus moved to a higher clause, it surfaces in accusative case. In (40a), for instance, the embedded subject is nominative, while in (40b) the same subject, which has undergone focus movement to the matrix clause, appears in accusative case.[13]

(40) a. János akarja, hogy **az ilösebbik fia** orvos legyen.
John wants that the elder son-nom doctor becomes

‘John wants his elder son to become a doctor.’

b. János **az ilösebbik fiát** akarja, hogy orvos legyen (É. Kiss, 1987, p. 143)
John the elder son son-acc wants that doctor becomes

‘It is his elder son that John wants to become a doctor.’

This is similar to the pattern seen in the *tough* construction, where we see a change in case under movement. As in that case, I will take it to be an indication of the lack of a direct syntactic dependency between the surface position of the focused element and position of the gap in the lower clause.

Consider next the effects of focus movement on scopal interpretation.

(41) a. János nem szeretné, ha el jönne sok ember. (A. Szabolcsi, p.c.)
John not would like if away came many people-nom

‘John would not like it if there were many people who came.’

b. János sok embert nem szeretne, ha el jönne. (É. Kiss, 1987, p. 126)
John many people-acc not would like if away came

‘There are many people such that John would not like it if they came.’

For the first example in this pair, the quantifier *sok* ‘most’ remains in the lower clause, and the sentence has a predominant interpretation in which the quantifier takes narrow scope with respect to the matrix predicate (as indicated by the gloss), though an interpretation where the quantifier takes wide scope is apparently possible as well, at least marginally. In contrast, as indicated in the gloss, the second example, in which the quantifier has been focus moved to the matrix clause, only permits a wide scope interpretation for the quantifier. Once again, this is the same pattern we have seen in the *tough* construction, where the displaced element cannot undergo quantifier lowering to yield a narrow scope interpretation. Once again, as this pattern is analogous to that seen in the *tough* construction, I am led to a similar conclusion: that the apparently displaced element in this construction is in fact generated in the matrix clause.

Of course, considerably more empirical work into the syntax of focus movement needs to be done to confirm the conclusions I have reached here. Nonetheless, if what I have said is on the right track, it suggests that by restricting the computational power of the grammar in the manner imposed by TAG, our analytic options are restricted in empirically desirable ways.

5. Conclusion

To sum up, I have argued in this paper that the study of linguistic competence can benefit from attention to the computational properties of the formal systems in which grammars are stated. I have suggested that the notion of generative capacity, in virtue of its linkage with abstract notions of computation, provides a useful metric for assessing the computational complexity of a grammar. More specifically, I hope to have convinced the reader not only that the mildly context-sensitive class is sufficiently powerful to encompass the range of regularities seen in natural language syntax, but also that the assumption that human grammars are expressed in terms of TAG, a formalism with mildly context-sensitive generative power, derives significant empirical benefits, providing an explanation for the existence and nature of locality effects on movement, and also usefully constrains the range of analytic possibilities for apparent dislocations.

The discussion in this paper has focused on characterizing a speaker's grammatical knowledge. Yet, if the restriction to mildly context-sensitive systems and to TAG in particular is truly a fundamental aspect of grammatical (and indeed mental) computation, we might expect to see its reflections beyond this domain, in the properties of the systems of language use, in which grammatical knowledge is embedded, as well as in the process of grammatical acquisition. Work in these areas is on-going, but there are already some suggestive results in the domain of language production (Ferreira, 2000; Frank & Badecker, 2001), comprehension (Joshi, 1990; Kim, Srinivas, & Trueswell, 2002; Rambow & Joshi, 1994), and acquisition (Frank, 1998, 2000; Joshi, 1989). As this sort of evidence accumulates, we can gain confidence in the importance of abstract notions of computational architecture and complexity in characterizing cognition.

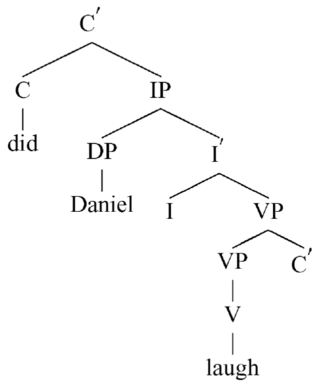
Notes

1. It is of course tempting to try to link the computations implicated in grammatical models to the computations needed during language processing. Early attempts to do this, in the context of the Chomsky's (1965) model, led to the Derivational Theory of Complexity: the proposal that sentences whose derivation involves more computations are more difficult to understand than sentences whose derivation involves fewer computations. This idea, however, met with only limited success (Fodor, Bever, & Garrett, 1974). Under more recent conceptions of grammar, it may be possible to resurrect this idea (Bresnan, 1982; Phillips, 1996; Steedman, 1989). In such a context, the arguments I make in the text are weakened somewhat. Even so, performance is sure to be affected by a variety of extra-grammatical factors, such as memory load or lexical and conceptual salience (Caplan & Waters, 1999; Gibson, 1998; Lewis, 1996).
2. Chomsky (1998), Collins (1997) and Frampton and Gutmann (2002), among others have explored measures of grammatical complexity having to do with the complexity of executing the algorithm that characterizes the grammar. For Chomsky, for instance,

this algorithm begins with a set of lexical items and must produce output representations of meaning and surface form, where part of this process involves a certain kind of optimization. Different characterizations of this optimization process result rather different levels of time complexity for executing this algorithm. One potential pitfall with this measure is that it does not easily allow us to compare different conceptions of grammar, as they can compute different functions, as we have seen. Additionally, it is far from simple to determine what if any empirical consequences result from this kind of complexity.

3. In the interest of space I won't give the grammar here, but see Hopcroft and Ullman (1979) or Partee, ter Meulen, and Wall (1993) for further discussion.
4. More recently, another way of characterizing generative capacity has emerged, in terms of the descriptive complexity: what kind of logical theory is needed to define a certain set of strings or structures. See Rogers (2003) for discussion and recent results.
5. Phonological structure appears to be more limited in its complexity than syntactic structure. In this domain, it appears that finite state power may be sufficient (Johnson, 1972; Karttunen, 1993).
6. In the multi-component extension to TAG (Weir, 1988), syntactic dependencies would be localized to derivationally linked sets of elementary trees rather than individual trees (Bleam, 2000).
7. Part of the TAG formalism, which we do not discuss here, includes restrictions specified on the nodes of the elementary trees concerning which nodes can adjoin. Specifically, in order for the TAG formalism to be formally well-behaved, in the sense of constituting an abstract family of languages (see Salomaa, 1973), it is necessary to allow at least the possibility of blocking adjoining at some node. The linguistic role of such adjoining constraints is explored in Vijay-Shanker and Joshi (1988) and Frank (2002).
8. Indeed, Kroch and Santorini (1991) provide linguistic arguments that the analysis that TAG provides is the right one in that it successfully predicts the attested range of mixed orderings seen in a variety of West Germanic languages and dialects.
9. See Frank (2002) for discussion of the shape of these elementary trees.
10. If we suppose instead that the preposition *after* is the complementizer of the clause headed by the verb *sang*, the derivation is also ruled out, though for slightly different reasons. In this case, the derivation would require the following auxiliary tree to adjoin between the fronted wh-element *what* and the rest of its clause *after Gabriel sang*:

(i)



Under the assumption that elementary trees provide only slots for the arguments (and not modifiers) of a lexical head, this structure is ill-formed: *laugh* is an intransitive verb, and as such does not license the C′ node for the temporal modifier.

11. If we do not adopt the rewriting perspective on the adjoining operation, this derivation is excluded by what Frank (2002) calls the Markovian restriction on TAG derivations: all combinatory operations must be licensed by the properties of the elementary trees that are being combined. As a result, adjoining cannot take place at some node labeled L if there is no L-recursive elementary auxiliary tree involved in the adjoining. In the case discussed in the text, this restriction blocks adjoining to C′ since there is no C′-recursive auxiliary tree.
12. One apparent counter-example occurs in English constructions like the following, where the subject that is raised from the lower clause appears interleaved between the matrix auxiliary and main verbs.
 (i) **Did** Gabriel **seem** to have finished his homework?
 This problem is much more limited in scope than the one posed by the Hungarian cases discussed in the text, and admits a range of solutions that do not extend the power of TAG (Frank, Kulick, & Vijay-Shanker, 2000; Kulick, 2000).
13. Gervain (2002) discusses the fact that some Hungarian speakers permit the nominative case of the subject to be retained under focus movement. For the present, I leave open the proper analysis of the focus movement construction for such speakers.

Acknowledgments

It is a great pleasure to dedicate this paper to Aravind Joshi, whose work and whose humanity have been and continue to be a source of inspiration. For comments and suggestions and for help with the Hungarian data, I wish to thank Judit Gervain, Istvan Kenesei, Aniko Liptak, Mark Steedman, Anna Szabolcsi, and an anonymous *Cognitive Science* reviewer

References

Abeillé, A. (1993). *Les Nouvelles Syntaxes*. Paris: Armand Colin.

Barton, G. E., Berwick, R. C., & Ristad, E. S. (1987). *Computational complexity and natural language*. Cambridge, MA: MIT Press.

Becker, T., Rambow, O., & Niv, M. (1992). *The derivational generative power of scrambling is beyond LCFRS* (Technical report IRCS 92-38). Institute for Research in Cognitive Science, University of Pennsylvania.

Bleam, T. (2000). In Abeillé & Rambow (Eds.), *Clitic climbing and the power of tree adjoining grammar* (pp. 193–220).

Bresnan, J. (Ed.). (1982). *The mental representation of grammatical relations*. Cambridge, MA: MIT Press.

Bresnan, J. (2001). *Lexical-functional syntax*. Oxford: Blackwell.

Caplan, D., & Waters, G. S. (1999). Verbal working memory and sentence comprehension. *Behavioral and Brain Sciences*, *22*(1), 77–126.

Carroll, J., Nicolov, N., Shaumyan, O., Smets, M., & Weir, D. (2000). Engineering a wide-coverage lexicalized grammar. In *Proceedings of the 5th International Workshop on Tree Adjoining Grammars and Related Frameworks*, Paris, pp. 55–60.

Chomsky, N. (1957). *Syntactic structures*. Mouton: The Hague.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control, 2*.

Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, N. (1977). On wh-movement. In P. Culicover, T. Wasow, & A. Akmajian (Eds.), *Formal syntax* (pp. 71–132). New York: Academic Press.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (1998). Some observations on economy in generative grammar. In P. Barbosa, D. Fox, P. Hagstrom, M. McGinnis, & D. Pesetsky (Eds.), *Is the best good enough? Optimality and competition in syntax* (pp. 115–127). Cambridge, MA: MIT Press and MIT Working Papers in Linguistics.

Chomsky, N., & Miller, G. (1963). In R. D. Luce, R.R. Bush, E. Galanter (Eds.), *Introduction to the formal analysis of natural languages*. Handbook of mathematical psychology. Vol. 2, 269–321. New York: John Wiley & Sons.

Collins, C. (1997). *Local economy*. Cambridge, MA: MIT Press.

Culy, C. (1985). The complexity of the vocabulary of Bambara. *Linguistics and Philosophy*, *8*, 345–351.

Epstein, S. D. (1989). Quantification in null operator constructions. *Linguistic Inquiry*, *20*, 647–658.

Ferreira, F. (2000). Syntax in language production: An approach using tree-adjoining grammars. In L. Wheeldon (Ed.) *Aspects of language production* (pp. 291–330). Philadelphia, PA: Psychology Press.

Fodor, J. A., Bever, T., & Garrett, M. F. (1974). *The psychology of language: An introduction to psycholinguistics and generative grammar*. New York: McGraw-Hill.

Frampton, J., & Gutmann, S. (2002). Crash-proof syntax. In S. D. Epstein & T. D. Seely (Eds.) *Derivation and explanation in the Minimalist Program*. Oxford: Blackwell.

Frank, R. (1998). Structural complexity and the time course of grammatical development. *Cognition*, *66*, 249–301.

Frank, R. (2000). In A. Abeillé, & O., Rambow (Eds.), From regular to context-free to mildly context-sensitive tree-rewriting systems: The path of child language acquisition. Tree Adjoining Grammars: Formalisms, Linguistic analysis and processing (pp. 101–120). Stanford, CA: CSLI.

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

Frank, R., & Badecker, W. (2001). *Modeling incremental syntactic encoding with tree-adjoining grammar (iTAG): How grammar constrains production and how production constrains grammar*. Paper presented at the 14th Annual CUNY Conference on Human Sentence Processing, University of Pennsylvania, Philadelphia.

Frank, R., Kulick, S., & Vijay-Shanker, K. (2000). Monotonic c-command: A new perspective on tree adjoining grammar. *Grammars*, *3*(2/3), 151–173.

Gazdar, G., Klein, E., Pullum, G.K., & Sag, I. (1985). *Generalized phrase structure grammar*. Harvard University Press, Cambridge, MA.

Gervain, J. (2002). *Linguistic methodlogy and microvariation in language: The case of operator-raising in Hungarian*. Master's thesis, University of Szeged.

Gibson, E. (1998). Locality of syntactic dependencies. *Cognition*, *68*, 1–76.

Grimshaw, J. (1991). *Extended projection*. Manuscript, Brandeis University.

Hale, M., & Reiss, C. (2000). Substance abuse and "disfunctionalism": Current trends in phonology. *Linguistic Inquiry*, *31*(1), 157–169.

Hegarty, M. (1993). *Deriving clausal structure in tree adjoining grammar*. Manuscript, University of Pennsylvania.

Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to Automata theory, languages, and computation*. Reading, MA: Addison Wesley.

Huybregts, R. (1985). The weak inadequacy of context-free phrase structure grammars. In G. de Haan, M. W. Trommelen, & Zonneveld (Eds.), *Van Periferie naar Kern* (pp. 81–99). Dordrecht: Foris.

Johnson, C. D. (1972). *Formal aspects of phonological description*. The Hague: Mouton.

Joshi, A. K. (1985). How much context-sensitivity is required to provide reasonable structural descriptions: Tree adjoining grammars. In D. Dowty, L. Karttunen, & A. Zwicky (Eds.), *Natural language parsing: Psychological, computational and theoretical perspectives* (pp. 206–250). Cambridge: Cambridge University Press.

Joshi, A. K. (1989). A possible mathematical specification of degree-0 or degree-0 plus a little learnability. *Behavioral and Brain Sciences, 12*.

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, A. K., Becker, T., & Rambow, O. (2000). In A. Abeillé, & O. Rambow (Eds.), *Complexity of scrambling: A new twist on the competence-performance distinction*. Tree Adjoining Grammars: Formalisms, linguistic analysis and processing (pp. 167–181). Stanford, CA: CSLI.

Joshi, A. K., & Kulick, S. (1997). Partial proof trees as building blocks for a categorial grammar. *Linguistics and Philosophy*, *20*, 637–667.

Joshi, A. K., Levy, L., & Takahashi, M., (1975). Tree adjunct grammars. *Journal of the Computer and System Sciences, 10*(1), 136–163.

Joshi, A. K., Vijay-Shanker, K., & Weir, D. (1991). The convergence of mildly context-sensitive grammatical formalisms. In P. Sells, S. Shieber, & T. Wasow (Eds.), *Foundational issues in natural language processing*. Cambridge, MA: MIT Press.

Kallmeyer, L., & Joshi, A. K. (2003). Factoring predicate argument and scope semantics: Underspecified semantics with LTAG. *Research on Language and Computation, 1*, 3–58.

Karttunen, L. (1993). Finite-state constraints. In J. Goldsmith (Ed.), *The last phonological rule* (pp. 173–194). Chicago, IL: University of Chicago Press.

Kasper, R., Kiefer, B., Netter, K., & Vijay-Shanker, K. (1995). Compilation of HPSG to TAG. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics*, Cambridge, MA, pp. 92–99.

Kenesei, I. (1994). Subordinate clauses. In F. Kiefer & K. É. Kiss (Eds.), *The syntactic structure of Hungarian* (pp. 275–354). San Diego: Academic Press.

Kim, A. E., Srinivas, B., & Trueswell, J. C. (2002). A computational model of the grammatical aspects of word recognition as supertagging. In S. Stevenson & P. Merlo (Eds.), *The lexical basis of sentence processing: Formal, computational and experimental issues* (pp. 109–136). John Benjamins.

É. Kiss, K. (1987). *Configurationality in Hungarian*. Dordrecht: D. Reidel.

É. Kiss, K. (1995). NP movement, operator movement, and scrambling in Hungarian. In *Discourse configurational languages*. New York: Oxford University Press.

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

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

Kroch, A. & Santorini, B. (1991). The derived constituent structure of the West Germanic verb raising construction. In R. Freidin (Ed.), *Principles and parameters in comparative grammar* (pp. 269–338). Cambridge, MA: MIT Press.

Kulick, S. (2000). *Constraining non-local dependencies in Tree Adjoining Grammar: Computational and linguistic perspectives*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Lewis, R. L. (1996). Interference in short-term memory: The magical number two (or three) in sentence processing. *Journal of Psycholinguistic Research*, *25*, 93–115.

May, R. (1985). *Logical form*. Cambridge, MA: MIT Press.

Miller, G., & Chomsky, N. (1963). In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Finitary models of language users*. Handbook of mathematical psychology. Vol. 2, pp. 419–491. New York: John Wiley & Sons.

Miller, P. (1991). Scandinavian extraction phenomena revisited: Weak and strong generative capacity. *Linguistics and Philosophy*, *14*, 101–113.

Miller, P. (1999). *Strong generative capacity: The semantics of linguistic formalisms*. Stanford, CA: CSLI.

Partee, B., ter Meulen, A., & Wall, R. E. (1993). *Mathematical methods in linguistics*. Dordrecht: Kluwer.

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

Phillips, C. (1996). *Order and structure*. Ph.D. thesis, MIT, Cambridge, MA.

Pollard, C., & Sag, I. (1994). *Head-driven phrase structure grammar*. Stanford, CA: CSLI.

Rambow, O., & Joshi, A. K. (1994). A processing model for free word order languages. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 267–301). Hillsdale, NJ: Lawrence Erlbaum Associates.

Rogers, J. (2003). Syntactic structures as multi-dimensional trees. *Research on Language and Computation*, *1*, 265–305.

Romero, M. (2002). Quantification over situation variables in LTAG: Some constraints. In *Proceedings of the Sixth International Workshop on Tree Adjoining Grammars and Related Frameworks (TAG+6)*, Venice, pp. 108–117. Available at http://ling.upenn.edu/sanguesa/papers/TAG+6-finalpaper.pdf.

Ross, J. R. (1967). *Constraints on variables in syntax*. Ph.D. thesis, MIT, Cambridge, MA (published as *Infinite syntax*, Norwood, NJ: Ablex, 1986).

Salomaa, A. (1973). *Formal languages*. New York: Academic Press.

Shieber, S. (1985). Evidence against the context-freeness of natural language. *Linguistics and Philosophy*, *8*, 333–343.

Steedman, M. (1989). Grammar, interpretation and processing from the lexicon. In W. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 463–504). Cambridge, MA: MIT Press.

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

Townsend, D. J. & Bever, T. G. (2001). *Sentence comprehension: The integration of habits and rules*. Cambridge, MA: MIT Press.

Vijay-Shanker, K. (1987). *A study of Tree Adjoining Grammars*. Ph.D. thesis, University of Pennsylvania, Philadelphia, PA.

Vijay-Shanker, K., & Joshi, A. K. (1985). Some computational properties of tree adjoining grammar. In *Proceedings of the 23rd Annual Meeting of the Association for Computational Linguistics*, Chicago, pp. 82–93.

Vijay-Shanker, K., & Joshi, A. K. (1988). Feature structure based tree adjoining grammars. In *Proceedings of the 12th International Conference on Computational Linguistics*, Budapest, pp. 714–719.

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