After Government and Binding theory

Edward Stabler, UCLA, draft, February 2010

revised version in J.van Benthem and A. ter Meulen, eds., *Handbook of Logic and Language, 2nd Edition*

The tradition in syntax called Government and Binding (GB) theory has been transformed by Chomsky (1995b) and following work, with a number of rather different, sustained efforts to keep structures and mechanisms to a minimum, avoiding the notion of 'government', for example, and preferring mechanisms that can be motivated by phonological or semantic requirements. The simpler structure of these recent 'Minimalist' proposals has facilitated algebraic and logical studies, situating these proposals with respect to other traditions. Some of these developments are very briefly reviewed here.

1 Theoretical developments

As discussed in Higginbotham (1997, §3.1), GB theory provides constituent structure with the mechanisms of X-bar theory: *projecting* each head of category X to an intermediate phrase X' that may contain a *complement* phrase YP, so that X' = [X,YP], and then projecting X further to a maximal projection XP that may contain a *specifier* phrase ZP, so that XP=[ZP,X']. Following Muysken (1982) and others, Chomsky (1995a, 2007) observes that this theory simply encodes the fact that certain properties of phrases X' and XP are determined by the category of the head X. So rather than propagating a category label to each projection, it suffices to let the head itself be the label. In a complex containing two elements X, Y, the label can be indicated by putting it first in an ordered pair $\langle X, Y\rangle$, commonly written in the Minimalist literature with the set-theoretic notation $\{X, \{X, Y\}\}$. In such a complex, we say that the element or category X 'projects', and that Y is 'maximal' in the sense that it does not project here.[1] (The order of elements in the pair is not generally the same as the order in which the elements are pronounced, as discussed in §1.2 below.)

Chomsky (2007) notes that in GB theory and its immediate antecedents, a structure is built and then successively modified in "five separate cycles: X-bar theory projecting D-structure, overt operations yielding S-structure, covert operations yielding LF, and compositional mappings to [sensory-motor] and [conceptual] interfaces." But in recent proposals, structure is built in a single sequence of operations, with each step simply extending the complexes built by earlier steps. In some proposals, certain steps form phrases which are *phases*, with complements that become available for pronunciation and interpretation and hence unavailable for further structure-manipulation in the syntax (Chomsky, 2004; Uriagereka, 1999). The basic structure building operation, Merge, is usually described with two cases.[2] *External merge* (EM) simply takes two elements X, Y and pairs them to produce $\{X, \{X, Y\}\}$. When Y is already part of X, another similar operation is possible, one which merges a copy of Y individuated by reference to its position in X; this operation is a *movement* of Y, now more often called *internal merge* (IM). That is, an IM step is usually regarded as producing a multidominance structure. Using coindexing to indicate this multidominance, if we write $X[Y_i]$ to indicate that X properly contains an occurrence of Y_i, then IM applies to $X[Y_i]$ and Y_i to produce the pair $\{X[Y_i], \{X[Y_i], Y_i\}\}$.[3]

With this perspective on movement, Chomsky suggests that human languages are designed to balance two pressures, namely, to keep arguments adjacent to predicates, and to explicitly mark discourse and scopal properties. "Language seeks to satisfy the duality in the optimal way, EM serving one function and IM the other," Chomsky (2008, pp.140-1) says, "The correlation is reasonably close, and perhaps would be found to be perfect if we understood enough."

[1] Obviously, if properties of X, Y determine which is the label of the complex they form, then while the notation $\{X, \{X, Y\}\}$ is redundant (Collins, 2002; Chomsky, 1995b, p.243), the notation $\{X, Y\}$ is inexplicit. So it is no surprise that Chomsky (1995b, p.246) also considers a third option, "It is natural, then, to take the label of [the complex] K [formed from α and β] to be not α itself, but rather H(K), a decision that leads to technical simplification. Assuming so, we take K=$\{$H(K),$\{\alpha, \beta\}\}$ where H(K) is the head of $\alpha\ldots$" This introduction uses the most common, explicitly labeled set theoretic pair $\{X, \{X, Y\}\}$, without exploring the many variants of these views in the literature.

[2] Other cases of merge have been proposed. 'Pair Merge' is a a kind of adjunction operation (Chomsky, 2004). 'Parallel Merge' has been proposed to handle coordination and certain other constructions (Citko, 2005). And a 'Sidewards Merge' operation has also been proposed (Nunes, 2001). We leave these aside here.

[3] The multidominance structures usually proposed are like trees except that some elements can be immediately dominated by more than one node; so they are unordered, directed, acyclic, labeled graphs. Kracht (2008) provides a careful analysis.

There are many different proposals about the conditions under which a sequence of Merge steps yields a complete, well-formed derivation, determined in part by features of the lexical items, conditions on the result (*interface conditions*), and various kinds of *least effort* conditions. There are also various different proposals about how the derived expression is pronounced and interpreted, determined in part by the phonetic properties of the lexical elements and the operations that affect them. Obviously, the real content of each particular theory is in these details; we briefly and informally survey a few of them here and then mention some formal assessments.

1.1 Features

In GB theory, Wh-movement is triggered by a (possibly empty) complementizer with a +Wh feature (Higginbotham, 1997, §6). In Minimalism, the triggering head is often called a *probe*, the moving element is called a *goal*, and there are various proposals about the relations among the features that trigger syntactic effects. Chomsky (1995b, p.229) begins with the assumption that features represent requirements which are checked and deleted when the requirement is met. This first assumption is modified almost immediately so that only a proper subset of the features, namely the 'formal,' 'uninterpretable' features, are deleted by checking operations in a successful derivation (Collins, 1997; Chomsky, 1995b, §4.5). Another idea is that certain features, in particular the features of certain functional categories, may be initially unvalued, becoming valued by entering into appropriate structural configurations with other elements (Chomsky, 2008; Hiraiwa, 2005). And some recent work adopts the view that features are never deleted (Chomsky, 2007, p.11). These issues remain unresolved.

There is also the substantive question: what are the syntactic features? What properties of lexical items and of complexes is the derivation sensitive to? Some early work in Minimalism stays close to GB theory, assuming categorial features N (noun), V (verb), A (adjective), P (preposition), T (tense), D (determiner), and so on, with additional features for agreement (often called ϕ features): Person, Number, Gender,..., and for movement: Wh, Case, Focus,.... Recent work has anatomized these features and provided additional structure. Analysis of traditional categorial features in terms of more basic properties ±V, ±N, originally suggested in the GB era (Chomsky, 1970), has been developed further (Baker, 2003). And the features implicated in overt movement of elements to phrase edges (formerly sometimes called 'licensors') seem to have distinctive syntactic and semantic properties (Rizzi, 1997; Rizzi, 2004; Chomsky, 2004; Boeckx, 2008b). Finally, developing earlier ideas about basic clausal requirements encoded in an 'extended projection principle' (Chomsky, 1981; Chomsky, 1982), the special features triggering movement ('EPP') are now often distinguished from the rest (Chomsky, 2000b; Boeckx, 2008a). The presence or absense of these features is similar to the 'strong' or 'weak' features of GB theories.

1.2 Linear order

In GB theory, the linear order of pronounced elements in a syntactic structure is typically assumed to vary from one language to another even at D-structure. For example, English grammar might include (perhaps as part of 'Case Theory') the stipulation that heads precede their complements, and the stipulation that subjects precede verb phrases; these are fundamental parameters of language variation (Koopman, 1983; Travis, 1984; Chomsky, 1982, pp.9-11, for example). Kayne (1994) takes a very different approach. Greenberg (1963) observes that certain constituent orders are rare, across all languages. For example, while the neutral order subject-verb-object

(SVO) is fairly common, OSV, VOS and OVS are very rare. And considering the orders of (1) demonstrative, (2) numeral, (3) adjective, and (4) noun in noun phases, the order 1234 is quite common, but some orders are unattested: 2134, 2143, 2413, 4213, 3124, and 3142. Recent studies confirm these observations (Cinque, 2005; Hawkins, 1983). Kayne proposes that some of these regularities may be due in part to a very simple structural fact: universally, heads take complements on the right and specifiers on their left. If a verb underlyingly takes its object as complement on its right, and its subject as specifier on its left (Koopman and Sportiche, 1991), and if all movement is to specifier position, on the left, then while all orders can still be derived, some orders will require more steps in their derivation than others.[4] Noting that linear order is needed only at the PF interface, Chomsky (1995a, §4.8) also proposes adopting some variant of Kayne's proposal. With this kind of view, the parameters of word

[4]This motivation for underlying SVO order is critiqued by Abels and Neeleman (2006) and Stabler (2010).

order variation in grammar are determined structurally, for example, by the properties of (sometimes empty) functional elements in the lexicon that may trigger movements.

With Kayneian assumptions, a moved element Y_i will be pronounced before $X[Y_i]$ in the pair $\{X[Y_i], \{X[Y_i], Y_i\}\}$, at the 'left edge' because it is in a specifier position. As for the lower occurrence of Y_i in $X[Y_i]$, called the 'trace' position in GB theory, usually it is not pronounced at all. But in certain cases, it seems that the trace is interpreted as if it were in its original position (Higginbotham, 1997, §6), and in certain 'partial movement' and 'overt copying' constructions, a moved element (or parts of it) is apparently pronounced more than once (Bošković and Nunes, 2007), as in the Vata (1a) from Koopman (1983), the Yoruba (1b) from Kobele (2006), the German (1c) from McDaniel (2004), and the Portuguese (1d) from Martins (2007):

(1) a. **li** à **li**-da zué saká
eat we eat-PAST yesterday rice
'We ATE rice yesterday'

b. Ri-**ra** **adie** ti Jimo **ra** **adie**
buying chicken REL Jimo buy chicken
'the fact that Jimo bought chicken'

c. **Wen** glaubt Hans **wen** Jakob gesehen hat?
whom thinks Hans whome Jakob seen has
'Who does Hans think Jakob saw?'

d. **Sabes** se/que ele vem à festa, **sabes**
know-2SG whether/that he comes to-the party know-2SG
'You do know whether he is coming to the party'

There has been some controversy about whether VP ellipsis, sluicing and other constructions also, at least sometimes, involve deletion of a full syntactic copy (Dalrymple, Shieber, and Pereira, 1991; Fiengo and May, 1995). Notice for example that, in at least some English dialects, pronunciation of the parenthesized phrase is fairly natural in (2a), to overtly express what has been ellided, while it is very unnatural with the non-copies in (2b):

(2) a. John **went to the store**, and Mary did too (**go to the store**)

b. John went to the store, and Mary did too (?go out, ?buy groceries).

Empirical and formal studies of these constructions are ongoing (Merchant, 2003; Kehler, 2002; Johnson, 2008).

1.3 Least effort and locality for IM and Agree

In GB theory, it is assumed that maximal projections and heads can both move, subject to certain locality conditions. Heads can only move to the 'closest' head position in a certain sense, and maximal projections cannot move across more than one 'bounding node' or 'barrier'. In early Minimalist proposals, there are various proposals of a similar nature: the 'shortest move constraint' (Chomsky, 1995b, §3.2) and the 'minimal link condition' (Chomsky, 1995b, §4.5), etc. More recent work introduces 'phases' which like the 'bounding nodes' of GB theory provide absolute bounds of a certain sort (Chomsky, 2001; Chomsky, 2000a). These proposals and other alternatives are surveyed and compared to GB theory in many places (Boeckx, 2008c; Boeckx and Grohmann, 2007; Bošković, 2007; Hornstein, Lasnik, and Uriagereka, 2007; Rizzi, 1997), but it remains unclear how movement should be bounded.

As noted in §1.1, IM is triggered by a certain kind of correspondence between the features of a licensing head or 'probe' and an element that needs to be licensed, a 'goal'. One idea is that a probe is a 'functional' element with an unvalued feature; it seeks a goal with a matching feature; the goal assigns a value to the probe; and IM occurs if the probe has a certain additional property (e.g. an 'EPP' feature). It is conceivable that the first steps of identifying a matching pair and assigning a feature value could occur without movement, an operation called *Agree* (Chomsky, 2008; Chomsky, 2007; Chomsky, 2000a). In GB theory, agreement was often supposed to be a reflex of local specifier-head relations (Kayne, 1989; Sportiche, 1998), but many recent proposals assume that agreement is a long-distance, movement-like relation (Chomsky, 2000b), as in the following Hindi example from

Boeckx (2004) in which the matrix verb *chaah* agrees not with its subject but with the embedded object

(3) Vivek-ne [kitaab parh-nii] chaah-ii
Vivek-ERG book.F read-INF.F want-PFV.V
'Vivek wants to read the book'

In such approaches, Agree is often assumed to have different locality conditions from IM, attributed to the fact that it does not move any material with phonetic properties, but simply assigns values to features.

1.4 Head movement

In GB theory, phrasal movement is distinguished from head movement. The two operations seem not only to displace different kinds of elements, but they seem to respect different locality requirements (Koopman, 1983; Travis, 1984), and unlike phrasal movement, head movement seems to have no semantic consequences. This perspective has been challenged on a number of fronts. Brody (2000) proposes that at least some head movement is a reflex of syntactic structure. Chomsky suggests that this operation might operate at the phonetic interface, with stricter locality conditions and no semantic effects for that reason (Chomsky, 1995b; Chomsky, 2000b; Boeckx and Stjepanovic, 2001). But recent work suggests that head movement actually does have semantic effects (Matushansky, 2006; Roberts, 2006). Furthermore, comparative and diachronic studies suggest that head movement and phrasal movement are closely related (den Besten and Edmondson, 1983; Kroch and Santorini, 1991). Koopman and Szabolcsi (2000) propose that many apparent head movements are really instances of 'remnant movement'. A remnant movement is movement of a a phrase from which something has already been extracted. When a phrase moves after all of its specifiers and complements (if any) have been extracted, this phrasal movement will look like head movement, if it is appropriately bounded. In GB analyses, remnant movement was usually blocked by some version of the Proper Binding Constraint (PBC), which requires that a moved phrase always c-commands its trace (Fiengo, 1977; Lasnik and Saito, 1994).[5] But the PBC blocks a number of seemingly well-supported, early analyses in English, German, Nweh and many other languages (den Besten and Webelhuth, 1990; Nkemnji, 1995):

(4) a. [$_{\text{VP}}$Criticized by his boss t_1]$_2$ John$_1$ has never been t_2.
b. [$_{\text{AP}}$How likely [t_1 to win]]$_2$ is$_3$ John$_1$ t_3 t_2?
c. [$_{\text{VP}}$ t_1 Gelesen]$_2$ hat [das Buch]$_1$ [keiner t_2].
read has the book noone
d. njikem a keʕ [te t_i akendɔŋ]$_j$ pfɛt$_i$ t_j
he Agr P1 neg plaintains eat

These and other analyses finally toppled the PBC (Müller, 1998; Abels, 2007), allowing new analyses like Kayne's (1998, p.134) structure (5a), and the treatment of Hungarian verbal complexes proposed by Koopman and Szabolcsi (2000, p.62) in (5b):

(5) a. John [$_{\text{VP}}$reads t_1]$_2$ [no novels]$_1$ t_2.
b. Nem akartam kezdeni [szét szedni t_1]$_2$ a [rádiót]$_1$] t_2
not wanted-1sg begin-inf apart take-inf the radio-acc

With this kind of remnant movement, the empirical arguments for head movement can be reassessed. The proper treatment of what GB theory calls 'head movement' relations remains an open question.

2 Algebraic analyses

Although the quick survey of recent work in the previous section might seem to suggest that everything is in flux in Minimalist theory, the relative simplicity of recent proposals has allowed mathematical analyses that reveal a remarkable consensus, not just among various minimalist proposals (Thm. 2), but also between these and other grammatical traditions (Thm. 1). From this perspective, the many

[5]This requirement is discussed, but not named, in Higginbotham (1997, §5.2).

changes and controversies reviewed in the previous section involve matters of detail, relatively small adjustments in a framework that is fairly stable and fairly simple. That is, many of the adjustments at the center of controversies now, adjustments directed towards providing the most insightful perspective, are not affecting the broad mathematical and computational properties of grammar. One exception, discussed below, concerns the fundamental properties of movement, IM.

As just reviewed, GB derivations build a basic tree ('D-structure') which is then repeatedly altered, but Minimalist proposals are much simpler. Certain operations simply apply to construct complexes from lexical items. This suggests that, at least as a good first approximation, Minimalist grammars define (partial) algebras by closing a finite set of (lexical) elements with respect to a small number of (partial) structure building operations (Keenan and Stabler, 2003).[6]

For each categorial feature N, V, A, P,..., let's suppose that we have corresponding selection features =N, =V, =A, =P,.... And in addition to the 'licensor' features +Wh, +Case, +Focus,..., we have corresponding 'licensee' features -Wh, -Case, -Focus,.... Call the set of categorial, selection, licensor, and licensee features F. Using standard spellings of words Σ to represent phonetic and semantic properties, we pair sequences $\sigma \in \Sigma^*$ of these elements with feature sequences $\alpha \in F^*$ using a binary type constructor :: for lexical items, obtaining pairs $\sigma :: \alpha$. A lexicon *Lex* is a finite set of string-feature sequence pairs $\sigma :: \alpha$. In derived, non-lexical expressions, sequences $\sigma \in \Sigma^*$ and features $\alpha \in F^*$ will be paired with a different constructor : to yield $\sigma : \alpha$.

We define structure building functions mapping trees to trees, so we regard *Lex* as providing a stock of 1-node labeled trees, where the labels are the structured arrays $\sigma :: \alpha$ of features. As mentioned above, in some of the prominent Minimalist theories, heads precede the first elements they merge with (their 'complements') and follow any later elements they merge with (their 'specifiers'). With this preliminary assumption, it is convenient to put the linear order into the syntactic trees. Nothing in the syntax will refer to this order, so we can regard it as coming from the phonetic interface, following standard Minimalist proposals. So instead of building pairs $\{X, \{X, Y\}\}$, we will build labeled ordered trees with a linear order signifying the temporal sequence (to which the syntactic operations will never refer), and marking the head of each complex not by linear position in a pair but by labeling internal nodes with symbols $>$ or $<$ that 'point' to the head. So for example, in the following tree, node 1 is the head, with complement 2 and specifier 3:

$$\overset{>}{3 \quad \overset{<}{1 \quad 2}}$$

A tree with one node heads itself, and in any tree with more than one node, we find the head of the tree by following the arrows from the root. The maximal projection of any head n is the largest subtree headed by n. At the leaves, we will have pairs of phonetic-syntactic feature sequences $\sigma :: \alpha$ or $\sigma : \alpha$. When no confusion will result, we sometimes write the 1-node tree with label $\epsilon : \epsilon$ simply as ϵ.

We can now define structure building operations *em* and *im* inspired by the Minimalist operations EM and IM. When the head of a tree is labeled $\sigma :: f\alpha$ or $\sigma : f\alpha$, so that its syntactic features begin with feature f, we sometimes refer to that tree as $t[f]$ and use t to represent the result of deleting the first feature f and possibly changing the type to :, so that the head of t is labeled $\sigma : \alpha$. Define the function *em* from pairs of trees to trees as follows,

$$\mathbf{em}(t_1[=f], t_2[f]) = \begin{cases} \overset{<}{t_1 \quad t_2} & \text{if } t_1 \text{ has exactly 1 node} \\ \overset{>}{t_2 \quad t_1} & \text{otherwise.} \end{cases}$$

Notice that *em* is triggered by a selection feature $=f$ and a corresponding category f, deleting both.

Now we define the unary function *im* which applies to a tree if and only if, first, its head has first syntactic feature $+f$, and second, it satisfies the following simple version of the 'shortest move constraint' (SMC): the tree contains at most one head with first syntactic feature $-f$. The value of the function is the result of replacing the maximal projection of the $t[-f]$ subtree with the empty subtree ϵ,

[6] Compare, for example, Chomsky (2000b, pp.100-101).

and put the corresponding subtree t with $\texttt{-}f$ deleted into specifier position. That is, letting $t\{t_1 \mapsto t_2\}$ be the result of replacing t_1 by t_2 in t, and letting $t_1^{>}$ be the maximal projection of the head of t_1,

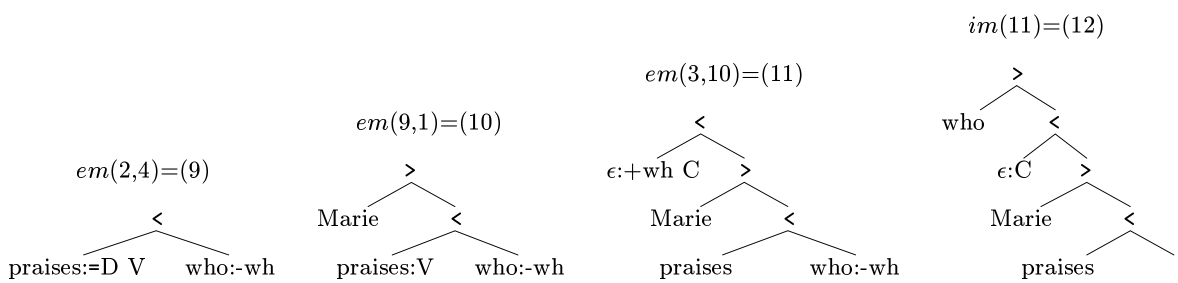
$$\mathbf{im}(t_1[\texttt{+}f]) = [_{>}\ t_2^{>}\ \ t_1\{t_2[\texttt{-}f]^{>} \mapsto \epsilon\}] \qquad \text{if (SMC) exactly one head in } t_1[\texttt{+}f] \text{ has } \texttt{-}f \text{ as its first feature.}$$

So *im* is triggered by a licensor feature $\texttt{+}f$ and a corresponding licensee feature $\texttt{-}f$, deleting both.

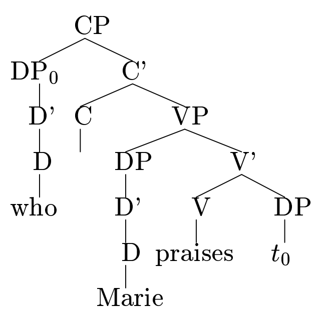
Let's call these simple grammars $G = \langle Lex, \{em, im\}\rangle$ *MGs* after the Minimalist grammars that inspire them. Since MG generating functions em, im are fixed, each MG is determined by its lexicon Lex. For any such grammar, let the structures $S(G)$ be the closure of Lex with respect to em, im. Let the *completed* structures be the trees in $S(G)$ with exactly 1 syntactic feature, namely, the 'start' category at the head. And let the set of sentences $L(G)$ be the phonetic yields of completed structures. For example, consider the following grammar G with 8 lexical items, numbered here for convenience:

(1)	Marie::D	Pierre::D	(5)
(2)	praises::=D =D V	knows::=C =D V	(6)
(3)	ϵ::=V +wh	ϵ::=V C	(7)
(4)	who::D -wh	and::=C =C C	(8)

Applying *em* and *im* to these items, we find some of the structures in $S(G)$:



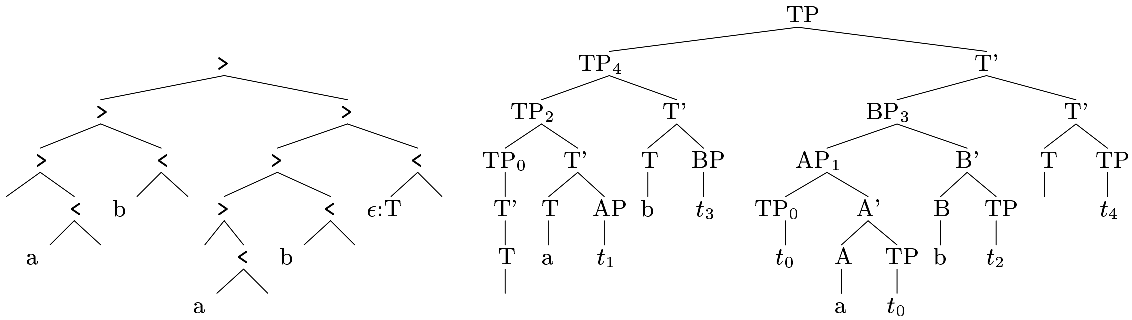
(To reduce clutter, when a node is labeled $\sigma : \epsilon$, we simply write σ, and when a node is labeled $\epsilon : \epsilon$ we do not write any label at all.) If C is the designated 'start' category, then these steps show that *who Marie praises* is in the set of sentences $L(G)$. We can also derive *Marie praises Pierre*, and *Pierre knows who Marie praises*, and infinitely many other sentences. In GB theory, the tree (12) would be something like this, co-indexing the moved element DP_0 with its trace t_0:



MGs can define non-context-free languages. For example, letting G_{xx} be the grammar defined by the following 7 element lexicon, with start category T, $L_{xx} = \{xx |\ x \in \{a, b\}^*\}$:

a::=A +l T -l	b::=B +l T -l b	
a::=T +r A -r	b::=T +r B -r	
ϵ::=T +r +l T	ϵ::T -r -l	ϵ::T

Among the derived structures in $S(G_{xx})$, we find the tree on the left below, pronounced *abab*, which in GB-like notation would be as on the right:



GB-style notation indicates the history of the derivation by coindexing each moved element X_i with its traces t_i, making it easy to see that there are several remnant movements in this MG derivation: there are two extractions from the moved phrase TP_4, one from the moved phrase TP_2, three from BP_3, and so on. It is important to observe that although this language is sometimes called a 'copy language', the grammar does not use any copying operation; no operation applies to an argument to yield a structure that contains two copies of that argument.

The MGs defined here are based on the slightly more complex grammars of Stabler (1997). They have been used to capture a range of Minimalist proposals, allowing careful study of their formal properties.

Theorem 1. (Vijay-Shanker, Weir and Joshi 1987; Michaelis 1998, 2001; Harkema, 2001)

$$\mathrm{L(CF)} \subset \mathrm{L(TAG)} = \mathrm{L(CCG)} \subset \mathrm{L(MG)} = \mathrm{L(MCTAG)} = \mathrm{L(MCFG)} = \mathrm{L(LCFRS)} \subset \mathrm{L(CS)},^{7}$$

where L(CF) is the set of languages defined by context free grammars; L(TAG) is the languages definable by tree adjoining grammars (Joshi, 1987); L(MCTAG) is the languages definable by set-local multiple-component tree adjoining grammars (Joshi, 1987; Weir, 1988); L(CCG) is the languages definable by combinatory categorial grammars as defined by Vijay-Shanker, Weir, and Joshi (1987); L(MG) is the languages definable by MGs; L(MCFG) is the languages definable by multiple context free grammars (Seki et al., 1991); L(LCFRS) is the languages definable by linear context free rewrite systems (Weir, 1988); and, L(CS) is the languages defined by context sensitive grammars.

These proofs of equations in Theorem 1 are constructive, showing how, for example, given an arbitrary MG grammar, we can construct a multiple context free grammar (MCFG) which defines exactly the same language. The needed constructions are quite straightforward, suggesting a similarity in their recursive mechanisms. The translation from MGs to the well-studied MCFG, for example, allows n step MG derivations to correspond to isomorphic n-step MCFG derivations.[8] In fact, some of these recipes for translating between grammars have been automated as a kind of compilation step. The MG languages are 'mildly context sensitive' in the sense of Joshi (1985), and they can be recognized in polynomial time (Harkema, 2000).

The MGs defined above use structure-building rules that are rather similar to standard Minimalist proposals, but better insight into theoretical proposals might be obtained by enriching and adjusting the simple MG with more of the mechanisms reviewed in §1. Many such studies have been carried out, revealing a surprising expressive equivalence of many ideas:

[7] For linguists who believe that there is be a fixed, universal set of features that trigger movement in human languages, it may be of interest to note that the references cited establish an infinite subhierarchy between L(CF) and L(MG). Let a k-MG be an MG in which there are k different features f such that $+f$ appears in the lexicon. Then L(0-MG)=L(CF), and for any $k \geq 0$, and $\mathrm{L}(k\text{-MG}) \subset \mathrm{L}((k+1)\text{-MG})$.

[8] Given the controversies around the use of empty categories in Chomskian grammar, it is interesting to note that while the straightforward translation of MGs to MCFGs preserves the empty categories, Seki et al. (1991, Lemma 2.2) show how, given any MCFG, it is possible to construct an equivalent grammar with no empty categories (except possibly the complete, empty sentence) and no rules that delete string arguments, but this construction can increase grammar size exponentially.

Theorem 2. (Michaelis 2001, 2002; Kobele 2002; Stabler 1997, 2001, 2003, 2010)

$$\mathrm{L(MG)=L(MGH)=L(MT)=L(DMG)=L(CMG)=L(PMG)},$$

where MGH extends MG with head movement and with covert phrasal movement (QR); MT is a version of Brody's 'mirror theory'; DMG modifies MGs so that selection of heads and specifiers can be on the left or on the right; CMG extends MGs by first conflating licensors with selectors, and licensees with categories, and also allowing certain features to persist when checked; and finally PMG extends MGs by designating certain categorial features as 'phases' which block extraction from their complements.

The proofs of these results are again constructive, providing recipes for conversions between formalisms with slightly different mechanisms.

The convergence of formalisms revealed by Theorems 1 and 2 might be taken as confirmation of the hypothesis that human languages are MG definable, confirming Joshi's (1985) hypothesis that human languages are 'mildly context sensitive'. MG variants not conforming to this hypothesis are easily defined (Gärtner and Michaelis, 2007; Kobele, 2005; Kobele and Michaelis, 2005), but mainly, these involve theoretical proposals that are not well motivated.

The idea that the convergence represented by Theorems 1 and 2 supports the hypothesis that human languages are MG definable has been seriously attacked, though, as being too weak and as being too strong. Notice that these attacks concern not just a particular Minimalist proposal or even Minimalism in general, but a wide range of proposals in various traditions of syntax.

The idea that the convergence is too weak comes mainly from the fact that, while the recognition problem for MG definable languages is polynomial, the 'universal recognition problem' for MGs – a problem includes grammar size along with input size parameters – is intractable, EXP-POLY time complete (Kaji et al., 1994, 1992; Satta, 1992). So for example, if one thinks of the language learner as exploring the class of MGs using a universal parsing strategy (possibly involving some kind of MG compilation to obtain a feasible recognizer), then both grammar size and input size matter, and this intractability result suggests that some reformulation of the problem may be required.[9] Perhaps we should try to see human languages as falling in a more restricted class. Let a k-MG be an MG in which there are k different features f such that $+f$ appears in the lexicon (see footnote 7). Then we could consider the stronger hypothesis that human languages are in L(k-MG) for some particular k, or in L(TAG).

The idea that the convergence represented by the previous Theorems is already too strong – i.e. in spite of the convergence of independent formalisms and traditions, it is simply false that human languages are MG definable – comes from a number of fronts. Various linguists have pointed out apparent regularities in human languages which are not definable by MG-equivalent formalisms. These include Old Georgian case marking (Michaelis and Kracht, 1997), German scrambling (Rambow, 1994), and Chinese number names (Radzinski, 1991). More central to the Minimalist program is evidence of copying in syntax, mentioned above. Intuitively, the structure building rules of MGs and similar formalisms merge and rearrange substrings, but copying rules double their arguments. Chomsky's IM and similar proposals are explicitly and intentionally presented as copying proposals: a multidominance structure is created by IM in which a single element (of unbounded size) may have more than one parent, allowing it to be pronounced (or partially pronounced) in more than one position. If the pronunciation of structures built by IM can involve pronouncing a complete phrase in multiple places, as suggested in §1.2 above, this can affect the expressive power of the grammar significantly. Kobele (2006) defines such an extension of MGs, and shows that the definable languages fall in the class of parallel multiple context free (PMCFG) languages, where L(MCFG)$\subset$L(PMCFG). The MG variants of Theorem 2 can all be modified in this way to allow copying in certain instances of im, producing a similarly equivalent but now more expressive range of grammatical options, defining languages in L(PMCFG) with harder but still polynomial recognition problems (Seki et al., 1991). It appears that most Minimalist syntacticians are persuaded that a step of this kind is empirically well-supported, but the question is certainly not settled. We seem to have arguments for more restrictive grammars and for less restrictive grammars. Much current mathematical interest is focused on this problem.

[9] Cf., Barton, Berwick, and Ristad (1987, §8) on GPSG. While GPSGs define only context free languages, recognizable in polynomial time, the universal recognition problem for GPSG is EXP-POLY time hard.

3 Logical and categorial analyses

3.1 MSO and tree automata

As noted by Higginbotham (1997, §7.3), logical studies of GB theory (Rogers 1994, 1999; Kracht 1995) concluded that it is a context free grammar notation, up to co-indexing of constituents. But those formalizations blocked 'remnant movement' with the PBC, as discussed in §1.4 above. Lacking the prohibition of remnant movement, MGs define a class of languages that strictly includes the context free languages. The logical connection between automata and logical definability established by Büchi (1960) and Elgot (1961), which set the stage for the extension to tree automata and Rogers' study of GB theory, has inspired analyses of non-context-free language classes like this (Mönnich, 2007; Kobele, Retoré, and Salvati, 2007; Kolb et al., 2003; Michaelis, Mönnich, and Morawietz, 2000; Michaelis, Mönnich, and Morawietz, 2001). In particular, Bloem and Engelfriet (2000) have shown how monadic second order logic can be used to specify an output tree as a certain kind of modification of an input tree, and such a transducer is 'direction preserving' if the directed edges of the input tree correspond to similarly directed edges of the output tree. Using this notion,

Theorem 3. (Mönnich, 2007) For every MG, there is a strongly equivalent direction preserving tree transducer definable in monadic second-order logic.

Although trees are complicated objects and so the formal theory of tree automata remains a specialized topic, this perspective on Chomskian proposals is remarkably simple and illuminating.

3.2 Type logical grammar

A different kind of logical perspective on Minimalist proposals is provided by multimodal higher-order type logic (Moortgat, 1996, §4). That system has Turing power (Carpenter, 1999; Moot, 2002) so of course it can encode any MG derivation, but a particularly transparent representation is suggested by Vermaat (2004), with a recipe for representing any MG analysis as a type-logical proof.[10] To represent trees, we use the binary modalities $\bullet_>, \bullet_<$, and for each MG grammar licensee $+f$, we use unary modalities $\Box_f$, $\Diamond_f$ to control the association of string positions related by im, as allowed by the following structural postulates (for $i \in \{<, >\}$):

$$
\begin{array}{rl}
\Diamond_f(A \bullet_> B) \rightarrow \Diamond_f A \bullet_> B & [P1] \\
\Diamond_f A \bullet_> (B \bullet_i C) \rightarrow B \bullet_i (C \bullet_< \Diamond_f A) & [P2] \\
\Diamond_f A \bullet_> (B \bullet_i C) \rightarrow B \bullet_i (\Diamond_f A \bullet_> C). & [P2]
\end{array}
$$

Then the 4 step MG derivation of *who Marie praises* shown in the previous section corresponds to the following proof.

$$
\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{\mathrm{C} \vdash \mathrm{C}/_<\mathrm{V} \qquad \dfrac{\mathsf{Marie} \vdash \mathrm{D} \qquad \dfrac{\mathsf{praises} \vdash (\mathrm{D}\backslash_<\mathrm{V})/_<\mathrm{D} \qquad \dfrac{\mathsf{who} \vdash \Box_{wh} D}{\langle\mathsf{who}\rangle^{wh} \vdash D}[\Box_{wh}E]}{\mathsf{praises} \circ_< \langle\mathsf{who}\rangle^{wh} \vdash \mathrm{D}\backslash_>\mathrm{V}}[/_<E]}{\mathsf{Marie} \circ_> (\mathsf{praises} \circ_< \langle\mathsf{who}\rangle^{wh}) \vdash \mathrm{V}}[\backslash_>E]}{\mathrm{C} \circ_< (\mathsf{Marie} \circ_> (\mathsf{praises} \circ_< \langle\mathsf{who}\rangle^{wh})) \vdash \mathrm{C}}[/_<E]}{\mathrm{C} \circ_< ((\langle\mathsf{who}\rangle^{wh} \circ_> (\mathsf{Marie} \circ_> \mathsf{praises})) \vdash \mathrm{C}}[P2]}{\langle\mathsf{who}\rangle^{wh} \circ_> (\mathrm{C} \circ_< (\mathsf{Marie} \circ_> \mathsf{praises})) \vdash \mathrm{C}}[P3]}{\langle\mathsf{who} \circ_> (\mathrm{C} \circ_< (\mathsf{Marie} \circ_> \mathsf{praises}))\rangle^{wh} \vdash \mathrm{C}}[P1]}{\mathsf{who} \circ_> (\mathrm{C} \circ_< (\mathsf{Marie} \circ_> \mathsf{praises})) \vdash \Box_{wh}\mathrm{C}}[\Box_{wh}I]
$$

Notice that the lexical premises are here placed at the top of the proof tree, with the conclusion at the root on the bottom. It is easy to recognize in this proof the first merging of the lexical elements, followed by the movement of the wh-phrase to initial position where the wh modality of *who* can 'unlock' the whole phrase.

[10] Although multimodal type logics have Turing power, a "landscape" of grammars that includes weaker systems is presented by Moortgat (1996). And Moot (2002, §9.3) surveys complexity results for multi-modal type logics and observes (Thm.9.16) that with only 'non-expanding' structural postulates, these logics define only context sensitive languages.

Considering this representation of Minimalist derivations, Vermaat points out that they are unnatural in the type logical perspective in a number of fundamental respects. *First*, most work in categorial approaches avoids the use of empty categories as a way to license type conversions, since the logic itself determines in a non-ad-hoc way which type conversions are valid (Moortgat, 1996, Prop.2.18, for example). *Second*, as in the earlier GB theory, in most Minimalist proposals, all the elimination steps are first order. That is, the type that is eliminated is always a simple category, and never a higher order type.[11] *Finally*, and more fundamentally, movement is here treated by the structural postulates which, under a Curry-Howard correspondence, will have no semantic consequences. In type logical grammar, it is much more natural to treat movement as the elimination of an empty assumption, which corresponds by Curry-Howard to λ-abstraction.

From the Minimalist perspective, on the other hand, the type logical analyses seem unnatural in a number of respects too. *First*, the use of empty categories, as for example in the case of optionally pronounced complementizers and relative pronouns, is linguistically natural (and see footnote 8). *Second*, while MG derivations are quite naturally emulated in type logical grammar, a number of the easy MG extensions look like they will be more challenging for the type logician. In particular, while it is indeed unnatural to treat movement with structural postulates lacking semantic consequences, the type logical alternative of treating movement as a kind of assumption discharge looks problematic too, particularly because of the possibility that the moved element may be pronounced in more than one place, as discussed in §1.2 above. The addition of copying to MGs is notationally and conceptually straightforward, and can be done in such a way that the resulting grammars are all PMCFGs with known, polynomial recognition methods (Kobele, 2006). It is not clear how type logical grammar should handle this.[12] More generally, while type logical grammar naturally captures the basic predicate argument relations of MGs, it is less clear how it will extend to agreement, case marking, and other distinctions of human languages. *Finally*, while the type logical framework can emulate MG derivations, no type logical characterization of the MG definable languages is known.

Given these differences, it is no surprise that various mixtures of the two traditions are being explored (Kanazawa and Salvati, 2007; Retoré and Salvati, 2007; Amblard, 2007; Vermaat, 2006; Lecomte and Retoré, 1999), not so much to faithfully realize earlier Chomskian proposals but rather to attain an elegant and empirically well supported approach.

4 The future

Many developments in the Minimalist program could not be mentioned in this short review. Only a few of these have been carefully studied, but very many are now within easy reach. The general tendency towards simpler mechanisms in Minimalist syntax has allowed substantial and rigorous comparisons of proposals, both within the Chomskian tradition and across traditions. Particularly significant is the convergence of Minimalist mechanisms and proposals in the tradition of tree adjoining grammar: minimalist grammars (without copying) are "mildly context sensitive" in Joshi's sense, and are naturally formalized by grammars that are very similar to other well understood mildly context sensitive formalisms. It is conceivable that, with further results of this sort, Minimalism will bring the rich empirical and theoretical currents of Chomskian syntax into a more accessible form, beginning an era of more sophisticated language studies that transcend traditional and disciplinary boundaries.

[11] This assumption of GB and Minimalist theorizing is occasionally made explicit. For example, Koopman and Sportiche (1991, p.215) say, "No category takes as a complement a syntactic category corresponding to a non-saturated predicate." Various Minimalist proposals, like the 'parallel merge' (Citko, 2005) mentioned in footnote 2, can be seen as ways of allowing non-first-order steps.

[12] Note again that it is a simple matter to apply Vermaat's recipe to the example derivation for $abab \in L_{xx}$ in §2 above. But in the first place, that grammar does not have copying in the sense of an operation that doubles the pronounced length of its arguments. And in the second place, Vermaat does not establish that the lexical type assignments provided by her recipe define exactly the language L_{xx}. Similarly Shan (2005, Figure 2.30) gives a type logical grammar for L_{xx}, but does not establish that his grammar does, in fact, define exactly L_{xx}. In contrast, it is relatively easy to establish that simple MGs or MCFGs for this language are correct with an induction on derivation length (Cornell, 1996, for example).

References

Abels, Klaus. 2007. Towards a restrictive theory of (remnant) movement: Improper movement, remnant movement, and a linear asymmetry. *Linguistic Variation Yearbook 2007*, 7:53–120.

Abels, Klaus and Ad Neeleman. 2006. Universal 20 without the LCA. Ms., University College, London.

Amblard, Maxime. 2007. *Calculs de Représentations Sémantique et Syntaxe Générative: Les Grammaires Minimalistes Catégorielles.* Ph.D. thesis, Université Bordeaux 1.

Baker, Mark C. 2003. *Lexical Categories: Verbs, Nouns, and Adjectives.* Cambridge University Press, NY.

Barton, G. Edward, Robert C. Berwick, and Eric Sven Ristad. 1987. *Computational Complexity and Natural Language.* MIT Press, Cambridge, Massachusetts.

Bloem, Roderich and Joost Engelfriet. 2000. A comparison of tree transductions defined by monadic second order logic and by attribute grammars. *Journal of Computer and System Sciences*, 61(1):1–50.

Boeckx, Cedric. 2004. Long-distance agreement in Hindi: Some theoretical implications. *Studia Linguistica*, 58:23–36.

Boeckx, Cedric. 2008a. *Aspects of the Syntax of Agreement.* Routledge, NY.

Boeckx, Cedric. 2008b. *Bare Syntax.* Oxford, NY.

Boeckx, Cedric. 2008c. *Understanding Minimalist Syntax: Lessons from Locality in Long-distance Dependencies.* Blackwell, Oxford.

Boeckx, Cedric and Kleanthes Grohmann. 2007. Putting phases in perspective. *Syntax*, 10(2):204–222, August.

Boeckx, Cedric and Sandra Stjepanovic. 2001. Head-ing toward PF. *Linguistic Inquiry*, 32:345–355.

Bošković, Željko. 2007. On the locality and motivation of move and agree: An even more minimal theory. *Linguistic Inquiry*, 38(4):589–644.

Bošković, Željko and Jairo Nunes. 2007. The copy theory of movement. In Norbert Corver and Jairo Nunes, editors, *The Copy Theory of Movement.* John Benjamins, Philadelphia.

Brody, Michael. 2000. Mirror theory: syntactic representation in perfect syntax. *Linguistic Inquiry*, 31:29–56.

Büchi, J. Richard. 1960. Weak second-order arithmetic and finite automata. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 6:66–92.

Carpenter, Bob. 1999. The Turing-completeness of multimodal categorial grammars. In *JFAK: Essays dedicated to Johan van Benthem on the occasion of his 50th birthday*, Amsterdam. Institute for Logic, Language, and Computation, University of Amsterdam. http://www.illc.uva.nl/j50/.

Chomsky, Noam. 1970. Remarks on nominalization. In R.A. Jacobs and P.S. Rosenbaum, editors, *Readings in English Transformational Grammar.* Ginn, Waltham, Massachusetts. Reprinted in Chomsky's (1972) *Studies on Semantics in Generative Grammar.* The Hague: Mouton.

Chomsky, Noam. 1981. *Lectures on Government and Binding.* Foris, Dordrecht.

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding.* MIT Press, Cambridge, Massachusetts.

Chomsky, Noam. 1995a. Bare phrase structure. In Gert Webelhuth, editor, *Government and Binding Theory and the Minimalist Program.* MIT Press, Cambridge, Massachusetts, pages 383–439.

Chomsky, Noam. 1995b. *The Minimalist Program.* MIT Press, Cambridge, Massachusetts.

Chomsky, Noam. 2000a. Minimalist inquiries: The framework. In Roger Martin, David Michaels, and Juan Uriagereka, editors, *Step by Step: Essays on Minimalism in Honor of Howard Lasnik.* MIT Press, Cambridge, Massachusetts, pages 89–155.

Chomsky, Noam. 2000b. *New Horizons in the Study of Language and Mind.* Cambridge University Press, NY.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz, editor, *Ken Hale: A Life in Language.* MIT Press, Cambridge, Massachusetts.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti, editor, *Structures and Beyond: The Cartography of Syntactic Structures Volume 3.* Oxford University Press, Oxford.

Chomsky, Noam. 2007. Approaching UG from below. In Uli Sauerland and Hans-Martin Gärtner, editors, *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics.* Mouton de Gruyter, NY, pages 1–30.

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero, and Maria Luisa Zubizarreta, editors, *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud.* MIT Press, Cambridge, Massachusetts.

Cinque, Guglielmo. 2005. Deriving Greenberg's Universal 20 and its exceptions. *Linguistic Inquiry*, 36(3):315–332.

Citko, Barbara. 2005. On the nature of merge: external merge, internal merge, and parallel merge. *Linguistic Inquiry*, 36:475–496.

Collins, Chris. 1997. *Local Economy.* MIT Press, Cambridge, Massachusetts.

Collins, Chris. 2002. Eliminating labels. In Samuel D. Epstein and Daniel Seeley, editors, *Derivation and Explanation.* Blackwell, Oxford.

Cornell, Thomas L. 1996. A minimalist grammar for the copy language. Technical report, University of Tübingen, SFB 340 Technical Report #79.

Dalrymple, Mary, Stuart Shieber, and Fernando Pereira. 1991. Ellipsis and higher order unification. *Linguistics and Philosophy*, 14:399–452.

den Besten, Hans and Jerold A. Edmondson. 1983. The verbal complex in continental West Germanic. In Werner Abraham, editor, *On the Formal Syntax of West Germanic.* John Benjamins, Philadelphia.

den Besten, Hans and Gert Webelhuth. 1990. Stranding. In G. Grewendorf and W. Sternefeld, editors, *Scrambling and Barriers.* Academic Press, NY.

Elgot, Calvin C. 1961. Decision problems of finite automata design and related arithmetics. *Transactions of the American Mathematical Society*, 98:21–52.

Fiengo, Robert. 1977. On trace theory. *Linguistic Inquiry*, 8:35–61.

Fiengo, Robert and Robert May. 1995. *Indices and Identity.* MIT Press, Cambridge, Massachusetts.

Gärtner, Hans-Martin and Jens Michaelis. 2007. Some remarks on locality conditions and minimalist grammars. In Uli Sauerland and Hans-Martin Gärtner, editors, *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax-Semantics.* Mouton de Gruyter, NY, pages 161–196.

Greenberg, Joseph. 1963. Some universals of grammar with particular reference to the order of meaningful elements. In Joseph Greenberg, editor, *Universals of Language: report of a conference held at Dobbs Ferry, New York, April 13-15, 1961.* MIT Press, Cambridge, Massachusetts.

Harkema, Henk. 2000. A recognizer for minimalist grammars. In *Sixth International Workshop on Parsing Technologies, IWPT'00.*

Harkema, Henk. 2001. A characterization of minimalist languages. In Philippe de Groote, Glyn Morrill, and Christian Retoré, editors, *Logical Aspects of Computational Linguistics*, Lecture Notes in Artificial Intelligence, No. 2099, pages 193–211, NY. Springer.

Hawkins, John A. 1983. *Word Order Universals.* Academic Press, NY.

Higginbotham, James. 1997. GB theory: An introduction. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language.* Elsevier, NY, pages 311–360.

Hiraiwa, Ken. 2005. *Dimensions of Symmetry in Syntax: Agreement and Clausal Architecture.* Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Hornstein, Norbert, Howard Lasnik, and Juan Uriagereka. 2007. The dynamics of islands: Speculations on the locality of movement. *Linguistic Analysis*, 33:149–175.

Johnson, Kyle. 2008. *Topics in Ellipsis.* Cambridge University Press, NY.

Joshi, Aravind. 1985. How much context-sensitivity is necessary for characterizing structural descriptions. In D. Dowty, L. Karttunen, and A. Zwicky, editors, *Natural Language Processing: Theoretical, Computational and Psychological Perspectives.* Cambridge University Press, NY, pages 206–250.

Joshi, Aravind. 1987. An introduction to tree adjoining grammars. In A. Manaster-Ramer, editor, *Mathematics of Language.* John Benjamins, Amsterdam.

Kaji, Yuichi, Ryuchi Nakanishi, Hiroyuki Seki, and Tadao Kasami. 1992. The universal recognition problems for multiple context-free grammars and for linear context-free rewriting systems. *IEICE Transactions on Information and Systems*, E75-D(1):78–88.

Kaji, Yuichi, Ryuchi Nakanishi, Hiroyuki Seki, and Tadao Kasami. 1994. The computational complexity of the universal recognition problem for parallel multiple context-free grammars. *Computational Intelligence*, 10:440–452.

Kanazawa, Makoto and Sylvain Salvati. 2007. Generating control languages with abstract categorial grammars. In L. Kallmeyer, P. Monachesi, G. Penn, and G. Satta, editors, *Proceedings of the 12th conference on Formal Grammar (FG'07)*, Stanford, California. CLSI Publications.

Kayne, Richard S. 1989. Null subjects and clitic climbing. In O. Jaeggli and K. Safir, editors, *The Null Subject Parameter.* Kluwer, Dordrecht.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. MIT Press, Cambridge, Massachusetts.

Keenan, Edward L. and Edward P. Stabler. 2003. *Bare Grammar*. CSLI Publications, Stanford, California.

Kehler, Andrew. 2002. *Coherence, Reference, and the Theory of Grammar*. CSLI Publications, Stanford, California.

Kobele, Gregory M. 2002. Formalizing mirror theory. *Grammars*, 5:177–221.

Kobele, Gregory M. 2005. Features moving madly: A note on the complexity of an extension to MGs. *Research on Language and Computation*, 3(4):391–410.

Kobele, Gregory M. 2006. *Generating Copies: An Investigation into Structural Identity in Language and Grammar*. Ph.D. thesis, UCLA.

Kobele, Gregory M. and Jens Michaelis. 2005. Two type 0 variants of minimalist grammars. In *Proceedings of the 10th conference on Formal Grammar and the 9th Meeting on Mathematics of Language, FGMOL05*.

Kobele, Gregory M., Christian Retoré, and Sylvain Salvati. 2007. An automata-theoretic approach to minimalism. In James Rogers and Stephan Kepser, editors, *Model Theoretic Syntax at 10. ESSLLI'07 Workshop Proceedings*.

Kolb, Hans-Peter, Jens Michaelis, Uwe Mönnich, and Frank Morawietz. 2003. An operational and denotational approach to non-context-freeness. *Theoretical Computer Science*, 293(2):261–289.

Koopman, Hilda. 1983. *The Syntax of Verbs: From Verb Movement Rules in the Kru Languages to Universal Grammar*. Foris, Dordrecht.

Koopman, Hilda and Dominique Sportiche. 1991. The position of subjects. *Lingua*, 85:211–258. Reprinted in Dominique Sportiche, *Partitions and Atoms of Clause Structure: Subjects, agreement, case and clitics*. NY: Routledge.

Koopman, Hilda and Anna Szabolcsi. 2000. *Verbal Complexes*. MIT Press, Cambridge, Massachusetts.

Kracht, Marcus. 1995. Syntactic codes and grammar refinement. *Journal of Logic, Language and Information*, 4:41–60.

Kracht, Marcus. 2008. On the logic of LGB-type structures, Part I: Multidominance structures. In Fritz Hamm and Stephan Kepser, editors, *Logics for Linguistic Structures*. Mouton de Gruyter, NY.

Kroch, Anthony S. and Beatrice Santorini. 1991. The derived constituent structure of the West Germanic verb-raising construction. In Robert Freidin, editor, *Principles and Parameters in Comparative Grammar*. MIT Press, Cambridge, Massachusetts.

Lasnik, Howard and Mamoru Saito. 1994. *Move Alpha: Conditions on Its Application and Output*. MIT Press, Cambridge, Massachusetts.

Lecomte, Alain and Christian Retoré. 1999. Towards a minimal logic for minimalist grammars. In *Proceedings, Formal Grammar'99*, Utrecht.

Martins, Ana Maria. 2007. Double realization of verbal copies in European Portuguese emphatic affirmation. In Norbert Corver and Jairo Nunes, editors, *The Copy Theory of Movement*. John Benjamins, Philadelphia.

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry*, 37(1):157–181.

McDaniel, Dana. 2004. *Conditions on Wh-Chains*. Ph.D. thesis, The City University of New York.

Merchant, Jason. 2003. *The Syntax of Silence*. Oxford, NY.

Michaelis, Jens. 1998. Derivational minimalism is mildly context-sensitive. In *Proceedings, 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, and Christian Retoré, editors, *Logical Aspects of Computational Linguistics*, Lecture Notes in Artificial Intelligence, No. 2099, pages 228–244, NY. Springer.

Michaelis, Jens. 2002. Notes on the complexity of complex heads in a minimalist grammar. In *Proceedings of the 6th International Workshop on Tree Adjoining Grammars and Related Frameworks, TAG+6*.

Michaelis, Jens and Marcus Kracht. 1997. Semilinearity as a syntactic invariant. In Christian Retoré, editor, *Logical Aspects of Computational Linguistics*, pages 37–40, NY. Springer-Verlag (Lecture Notes in Computer Science 1328).

Michaelis, Jens, Uwe Mönnich, and Frank Morawietz. 2000. Algebraic description of derivational minimalism. In *International Conference on Algebraic Methods in Language Proceesing, AMiLP'2000/TWLT16*.

Michaelis, Jens, Uwe Mönnich, and Frank Morawietz. 2001. On minimalist attribute grammars and macro tree transducers. In Christian Rohrer, Antje Rossdeutscher, and Hans Kamp, editors, *Linguistic Form and its Computation*. CSLI Publications, Stanford, California, pages 287–326.

Mönnich, Uwe. 2007. Minimalist syntax, multiple regular tree grammars, and direction preserving tree transductions. In James Rogers and Stephan Kepser, editors, *Model Theoretic Syntax at 10. ESSLLI'07 Workshop Proceedings*.

Moortgat, Michael. 1996. Categorial type logics. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*. Elsevier, Amsterdam.

Moot, Richard. 2002. *Proof Nets for Linguistic Analysis*. Ph.D. thesis, Utrecht University.

Müller, Gereon. 1998. *Incomplete Category Fronting*. Kluwer, Boston.

Muysken, Pieter. 1982. Parameterizing the notion 'head'. *Journal of Linguistic Research*, 2:57–75.

Nkemnji, Michael A. 1995. *Heavy Pied-Piping in Nweh*. Ph.D. thesis, University of California, Los Angeles.

Nunes, Jairo. 2001. Sideward movement. *Linguistic Inquiry*, 32:303–344.

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

Rambow, Owen. 1994. *Formal and Computational Aspects of Natural Language Syntax*. Ph.D. thesis, University of Pennsylvania. Computer and Information Science Technical report MS-CIS-94-52 (LINC LAB 278).

Retoré, Christian and Sylvain Salvati. 2007. Non-associative categorial grammars and abstract categorial grammars. In R. Muskens, editor, *Proceedings of the Workshop on New Directions in Type-theoretic Grammars (NDTTG'07)*, Amsterdam. Foundation of Logic, Language and Information (FoLLI).

Rizzi, Luigi. 1997. The fine structure of the left periphery. In Liliane Haegeman, editor, *Elements of Grammar*. Kluwer, Boston, pages 281–337.

Rizzi, Luigi. 2004. Locality and left periphery. In Adriana Belletti, editor, *Structures and Beyond: Cartography of Syntactic Structures, Volume 3*. Oxford, NY, pages 104–131.

Roberts, Ian. 2006. Clitics, head movement, and incorporation. Manuscript, Downing College, University of Cambridge.

Rogers, James. 1994. Studies in the logic of trees with applications to grammar formalisms. Technical Report TR-95-04, Department of Computer and Information Sciences, University of Delaware, Newark, Delaware.

Rogers, James. 1999. *A Descriptive Approach to Language-Theoretic Complexity*. Cambridge University Press, NY.

Satta, Giorgio. 1992. Recognition of linear context-free rewriting systems. In *Proceedings of the 30th annual meeting on Association for Computational Linguistics*, pages 89–95, Morristown, New Jersey. Association for Computational Linguistics.

Seki, Hiroyuki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. 1991. On multiple context-free grammars. *Theoretical Computer Science*, 88:191–229.

Shan, Chung-chieh. 2005. *Linguistic Side Effects*. Ph.D. thesis, Harvard University.

Sportiche, Dominique. 1998. Movement, agreement and case. In Dominique Sportiche, editor, *Partitions and Atoms of Clause Structure: Subjects, agreement, case and clitics*. Routledge, New York.

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

Stabler, Edward P. 2001. Recognizing head movement. In Philippe de Groote, Glyn Morrill, and Christian Retoré, editors, *Logical Aspects of Computational Linguistics*, Lecture Notes in Artificial Intelligence, No. 2099. Springer, NY, pages 254–260.

Stabler, Edward P. 2003. Comparing 3 perspectives on head movement. In A. Mahajan, editor, *From Head Movement and Syntactic Theory, UCLA/Potsdam Working Papers in Linguistics*. UCLA.

Stabler, Edward P. 2010. Computational perspectives on minimalism. In Cedric Boeckx, editor, *Oxford Handbook of Minimalism*. Oxford University Press, Oxford.

Travis, Lisa. 1984. *Parameters and effects of word order variation*. Ph.D. thesis, Massachussets Institute of Technology, Cambridge, Massachusetts.

Uriagereka, Juan. 1999. Multiple spell-out. In S. D. Epstein and N. Hornstein, editors, *Working Minimalism*. MIT Press, Cambridge, Massachusetts, page 251282.

Vermaat, Willemijn. 2004. The minimalist move operation in a deductive perspective. *Research on Language and Computation*, 2(1):69–85.

Vermaat, Willemijn. 2006. *The Logic of Variation: A cross-linguistic account of wh-question formation*. Ph.D. thesis, Utrecht University.

Vijay-Shanker, K., David Weir, and Aravind Joshi. 1987. Characterizing structural descriptions produced by various grammatical formalisms. In *Proceedings of the 25th Annual Meeting of the Association for Computational Linguistics*, pages 104–111.

Weir, David. 1988. *Characterizing Mildly Context-Sensitive Grammar Formalisms.* Ph.D. thesis, University of Pennsylvania, Philadelphia.