7

The Minimalist Program

Alec Marantz
MIT

Contents

Prelude

This chapter, as its first section headline indicates, leads us from a discussion of the Government–Binding (GB) version of the Principles and Parameters approach to the Minimalist Program (MP). It simultaneously serves as an interpretation of Chomsky (1989, 1992) and as an introduction to chapter 8, which at the time of writing is the most recent original development of the Minimalist Program by Chomsky.

The chapter discusses how the MP conceptualizes grammatical representations and their wellformedness. The system reduces the set of four levels of representation D-structure, S-Structure, Logical Form (LF), and Phonological Form (PF) of standard GB theory to the two *interface levels* LF and PF. LF interfaces with semantic–conceptual systems of cognition and PF is connected to articulatory–perceptual modules. As a function of this reduction of the number of levels, the structure of the remaining levels and in particular their interface character (i.e. their role in connecting linguistic representations to interpretation elsewhere in cognition) takes on paramount importance. This is reflected by the following: every principle that constrains derivations either applies at LF, at PF, or at every step of the derivation where it is relevant. Consequently, all the benefits of principles and constraints that applied at D-structure or S-Structure but not throughout the derivation in the classical GB system now have to be captured in a different manner, either by the relevant constraints being reformulated as wellformedness conditions on derivations as a whole or by their becoming conditions on the wellformedness of one of the interfaces.

Another important departure of the MP from earlier versions of generative syntax concerns the notion that the grammaticality of one derivation may depend on the properties of another derivation. In particular, the MP invokes a number of *economy principles* such as "Shortest Move," "Procrastinate," and "Greed" that compare derivations involving the same lexical resources and discard all but the most economical derivations.

I would like to thank audiences and classes in Amsterdam and Seoul and at MIT for their help in clarifying what needed to be said in this chapter. In addition, I am indebted to Morris Halle and Gert Webelhuth for their comments on earlier drafts. I alone am responsible for any misinterpretations of the Minimalist Program contained in these pages; in no way does this chapter reflect any official or authorized view of the theory or of syntax in general. In addition, to the extent that I have correctly explained aspects of the Minimalist Program, my exposition should not be taken as endorsement of ideas or analyses.

0 From Government–Binding to Minimalism

The minimalist program (MP) continues the trend in syntactic theory begun in the late 1970s: the move from specific grammatical rules that describe particular syntactic constructions to general principles that interact to explain syntactic phenomena. This latest version of Chomsky's Principles and Parameters (P&P) approach to grammar grows out of a consideration of the interaction of mechanisms like Move-α and of principles like the Case Filter. In standard Government–Binding theories, operations such as Move-α apply freely. A single derivation resulting from the free application of such operations is evaluated in isolation; if it does not violate any principles, it is a grammatical derivation. In addition, only in a metaphoric sense do operations apply for a reason: an NP may move "to get Case" and thus pass the Case Filter, but this only means that the NP does move and does get Case. The notion that if the NP didn't move, it wouldn't get Case, plays no role in evaluating the grammatical derivation. Moreover, no comparison of the ways that an NP in a given position in a given structure might move "to get Case" is relevant to the grammaticality of any particular movement the NP may undergo.

A number of conceptual issues lead from standard GB theories to the MP. First, certain locality principles seem to have a "least effort" flavor to them. Both NP- and *wh*-movement seem to target the first potential position up from the source position of movement, from a slightly abstract point of view (see the discussion of "Shortest Move" in section 1 below for some relevant examples). NP-movement in passive and raising moves what appears to be the highest NP in a structure to the first A-position above this NP (the subject position); when this closest A-position is filled, movement to a still higher position is blocked. Here, only the shortest possible move, the one requiring the least effort, is allowed. Similarly, *wh*-movement moves a *wh*-constituent to the first available and appropriate A-bar position, usually the local spec of CP. Superiority effects (see e.g. (1c) below) suggest that the highest among a set of *wh*-phrases – the one closest to the spec of CP position – must move; *wh*-island effects suggest that when the closest spec of CP position is filled, movement of a *wh*-phrase beyond this position is blocked in some manner. In the Superiority case, the *wh*-constituent that would make the least overt effort to get to spec of CP must be the one that moves; in the *wh*-island case, the *wh*-constituent seems to be penalized for making the effort to skip the closest possible landing site – the local spec of CP.

Second, operations other than simple (and shortest) movement seem to be "last resort" options, applicable only when other options are prohibited. For example, the insertion of dummy "do" in *do*-insertion contexts in English happens only when other possibilities for realizing tense and agreement are blocked. Both "least effort" and "last resort" notions involve implicit or explicit comparison of derivations. "Least effort" implies the comparison of other amounts of effort a constituent might have made, given a particular configuration. If a filled subject position blocks NP movement to a higher A-position of a constituent c-commanded by the subject, the effort of moving to this filled subject position is being compared with the effort of moving even farther. "Last resort" implies the exhaustion of other possible resorts a derivation might have for avoiding the violation of some principle or filter.

Least effort and last resort suggest a striving for the cheapest or minimal way of satisfying principles. The MP relies explicitly on such "economy principles" in evaluating derivations. Another aspect to the MP is the move toward a minimum of principles and grammatical constructs themselves. Within P&P syntax, an increasing load has been placed on wellformedness conditions at the so-called "interface" levels of PF and LF. PF is assumed to be the structure that interfaces with the perceptual system in speech recognition and with the articulatory system in speech production. LF interfaces with a speaker's general knowledge and with extralinguistic cognitive systems (the systems involved in relating an LF to "meaning" in the intuitive sense). Despite the emphasis on interface conditions, pre-Minimalist systems relied as well on D- and S-Structure conditions and principles. For example, D-structure (DS) was often taken as the projection of lexical properties, including theta-role assignment and subcategorization. Principles of Binding Theory were taken to apply crucially at S-Structure (SS). The notion of PF and LF wellformedness can be linked to "visible" aspects of language – the possibility of phonetic and semantic interpretation; DS or SS wellformedness have no corresponding concreteness. On the LF side, the principle of Full Interpretation might rule out the presence of "excess" constituents in a structure, such as unbound variables or NPs without theta-roles. On the PF side, Full Interpretation might reject representations containing symbols with no phonetic realization. To the extent that ungrammatical derivations could be ruled out by general interface conditions at PF and LF, the grammar itself might be simplified, e.g. through the elimination of the independent levels of DS and SS and of principles that might have applied at these levels.

Conceptually, then, the major changes in the move to the MP are

these: constituents move for a reason, not freely; grammaticality depends on a comparison of derivations, not on the evaluation of a particular derivation in isolation; principles apply only at the interface levels of PF and LF or everywhere – DS and SS do not figure into the system. Below I will describe the implementation of these changes. However, this description will not provide anything like a complete technical introduction to Chomsky (1992). The original paper should be consulted for complete details.

1 Principles

In considering the principles that constrain derivations in the MP, it is important first to separate those that apply at the interface levels PF and LF from other principles. Other principles must apply everywhere that they are relevant in a derivation, since there are no other levels of grammar distinguished within this framework. In a conceptually clean version of minimalism, no operations occur explicitly in the course of a derivation to avoid a violation of a principle at one of the interfaces. Individual constituents might have particular needs that motivate them to move, however, and failure to meet these needs can lead to a derivation that is uninterpretable at one of the interfaces. For example, NPs have Case features that they must "check" in a derivation, and NPs will move to check these features. Failure to check a Case feature can lead to an uninterpretable structure at PF or at LF. (See sections 3 and 4 below for some comments on the technical aspects of feature checking.) A structure that fails to meet an interface condition at PF or LF is said to have "crashed." A derivation that crashes at an interface level has "failed to converge" at that level. Section 3 below will discuss the (morphological) features that require checking and the manner in which derivations crash or converge at the interface levels.

The evaluation of a structure at the interface levels for interpretability or crashes does not involve the comparison of the derivation of this structure with other possible derivations. Crashes, then, are local in the computational system. On the other hand, Economy Principles that operate across the grammar often involve the comparison of alternative derivations. The three major economy principles to be discussed here are Shortest Move, Greed, and Procrastinate. Although it might be possible to reduce some of the Economy principles to more general principles, Chomsky emphasizes that the particular forms these principles take in constraining the syntax cannot follow from general,

extralinguistic notions of economy. There are many ways a general desire to do the least possible, for example, could play out in the syntax. The particular grammatical Economy principles are just some of the infinite number of possible principles consistent with general economy considerations.

Shortest Move is the most technically specific Economy principle and takes over much of the work performed by Relativized Minimality (Rizzi 1990), Subjacency, and the Head Movement Constraint in earlier versions of P&P theory. The basic idea is that a constituent must move to the first position of the right kind up from its source position. Ideally, a head-movement constraint violation like (1a), a super-raising violation like (1b) and a superiority violation like (1c) would all violate the Shortest Move economy principle.

(1a) *Have John will t left by the time we get there?
(1b) *John is likely for it to seem t to have left
(1c) *What did you persuade who to buy t?

The application of "Shortest Move" needs to be relativized to the type of constituent moving and to the relevant landing site. Heads, for example, should be prohibited by Shortest Move from skipping over any head position "between," in the relevant sense, the position they start in and the targeted landing site. So in (1a), the head "have" cannot skip the head position occupied by "will" to move to a higher head position, perhaps the position of C in the example. For constituents undergoing movement to an A-position, only intervening A-positions count. In (1b), "John" skips the subject A-position occupied by "it" to get to the subject position of "is likely." Similarly, A-bar movement of *wh*-constituents must not skip over A-bar specifier positions. We will return to the question of what counts as a possible intervening A-position for NP-movement in section 4.1 below.

Note too that an explicit account of Shortest Move will need to include more than just a notion of "intervening possible landing site" if it is to account for, e.g., Superiority examples like (1c). Here, we must move the higher of two *wh*-phrases in a derivation in which we derive an LF that yields a pair-seeking question, "For which person X and which thing or things Y did you persuade X to buy Y?" We want to be comparing two derivations with the same lexical items and the same interpretation. In one, the one yielding the grammatical, "Who did you persuade to buy what?" we move the higher *wh-* overtly and do whatever LF operations are necessary to create the multiple *wh-* interpretation with the lower *wh-*. In the other, the one that yields

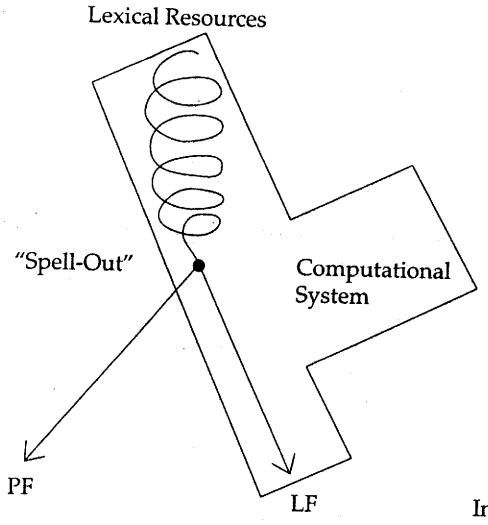
the word order in (1c), we have moved the lower *wh-* overtly. In neither case would the overt movement of the *wh-* cross a filled A-bar specifier position, nor is the overt movement by itself ungrammatical. We want the fact that moving the higher *wh-* involves a shorter move, measured in nodes crossed or some equivalent metric, to make this overt movement less costly than the movement of the lower *wh*-phrase. Since many analyses of multiple *wh-* constructions involve LF movement of the in situ *wh-* constituent(s), we need to make sure that the relative cost of the LF operations on the lower *wh-* added to the overt movement of the higher *wh-* is still less than the overt movement of the lower *wh-* and the LF operations on the higher *wh-*. Here the application of the Shortest Move principle involves a global comparison of different possible derivations, not a local application of a principle at one point in a derivation.

The application of Shortest Move, e.g. in the super-raising case in (1b), does not involve a comparison among "converging" derivations. Instead, the structure generated in the course of producing (1b) loses out to a more economical derivation that fails to converge. The structure generated in (1b) does in fact converge, since it would have all the relevant features checked off before the interface levels. However, *John* skips the position of the lower subject *it* when raising to the higher subject position. *John* could certainly make a shorter move by moving to the lower subject position. This more economical derivation, involving the shortest move that *John* could make from the most deeply embedded clause, would either be impossible – the landing site for the Shortest Move is already filled by *it* – or would fail to converge. Nevertheless, the more economical, but independently ill-formed, derivation blocks the (converging) derivation of (1b). Thus, in a sense, obeying Shortest Move takes precedence over convergence at the interface levels. Or, to put it differently, the evaluation of Shortest Move involves a comparison among all possible derivations, not just those that converge. If the Shortest Move dooms a derivation to crash at an interface, then there is no grammatical derivation using the lexical resources involved in the relevant comparison set of derivations.

Although the MP has no specific level of SS, there is a point in the computation of a grammatical representation where the derivation splits and heads toward the two interface levels, PF and LF. This point, called "Spell-Out," determines which movements will affect the pronunciation of a sentence – those that occur before Spell-Out – and which won't – those that occur after Spell-Out, on the way to LF. The assumption here is that the operations that occur between "Spell-Out" and PF are not of the same sort as those that operate within the computational

system on the road to LF. Or, to put it differently, the computational system (whose operations are described in section 2 below) runs on the line to LF, with the derivation from Spell-Out to PF subject to the rules and operations of a separate, phonological component.

(2)



"Procrastinate" is a principle that prefers derivations that hold off on movements until after Spell-Out, so that the results of such movements do not affect PF. Procrastinate is evaluated over convergent derivations; in effect, then, a derivation may violate Procrastinate in order to converge. In English, for example, main verbs do not raise to Tense before Spell-Out, and thus are pronounced within the VP, after certain VP adverbials and after negation (see section 4.2 below). This behavior of English contrasts with French, in which main verbs do raise to Tense before Spell-Out. English verbs are thus obeying Procrastinate, waiting until after Spell-Out to raise to Tense, which they must do by LF to check off their tense features. French verbs must violate Procrastinate to insure convergence. The assumption is that French tense features, unlike their English counterparts, are "strong" and visible at PF if not checked off. English tense features are "weak" and thus invisible at PF if not checked off. Because the English tense features are weak, English may wait to check off the tense features on main verbs until LF and Procrastinate says that if you can wait, you must wait.

Greed is perhaps the most problematic of the economy principles in the MP. The intuitive force of Greed is clear, but the examples in which it should apply seem to be explained by other principles. The principle of Greed states that a constituent may not move to satisfy the needs of some other constituent; movement is motivated for selfish reasons, to satisfy the needs of the moving constituent. For example, a constituent should not move to a position in order to check off features of a node in a checking relation with that position; the constituent should move only to check off its own features.

The problem for Greed is to make the theory of features and feature-checking clear enough to be able to decide where one is failing to be Greedy but nevertheless is satisfying other principles. Greed may have relevance for cases like those in (3) in which an NP has moved from a Case position.

(3a) *John seems [t is leaving]
(3b) *John seems to t [that Bill is leaving]
(3c) *John was said to t [that Bill is leaving]

The idea here is that NP-movement (raising, passive) from a Case position is to be ruled out by Greed. An NP in a Case position already has its Case (and phi-features – person, number and gender features) checked; further movement would only help check off the features of a higher Case position, i.e. further movement would be altruistic rather than greedy. To make Greed account for the ungrammaticality of (3), one must be explicit about what counts as feature-checking. If feature-checking in effect renders the relevant features invisible, then Greed would not be needed to rule out NP-movement from a Case position – since the NP in the Case position has all its relevant features checked in this position, whether the NP is greedy or altruistic, it has no features that can do any (further) checking. Then raising in, say, (3a) could not check features at the higher Case position since *John* has no features to check that haven't already been checked off in the lower Case position.

In languages like French with past passive participle agreement, it appears as if NPs may check at least their phi-features in more than one position. As the underlying object raises to the subject of the auxiliary verb in French passives, it appears to land in a specifier position to check gender and number phi-features with the passive participle. The agreement morphology that contributes the phi-features to the passive participle is visible on the verb in (4).

(4) Les filles sont [t [rencontrées au cinéma]].
the girls are met-fem-pl at the movie theater

The underlying object in (4) must check phi-features again in spec of AGRsP as the subject of the sentence (see the structure in (6) below; AGRsP is the functional phrase above T(ense)P within CP, while AGRoP is the functional phrase above VP). The fact of multiple agreement with the same argument in a single clause suggests that features do not disappear from NPs when they are checked, or at least that phi-features do not disappear. On the other hand, Case features are crucial to the examples in (3). We shall see in connection with the discussion of infinitivals in section 6 below that Greed, like Procrastinate, appears to be limited to deciding among competing derivations that converge at the interfaces. Since movements like the ungrammatical A-movements in (3) are motivated to insure convergence, if an NP could in fact carry multiple Case features or check Case more than once, Greed shouldn't prevent these movements in (3).

2 Derivations and the Computational System

In older versions of P&P Grammar, there was a clear notion of the "starting point" of a derivation. For example, in some theories, the operation of the phrase structure rules (or principles of the Base) would be followed by the insertion of lexical items into a tree generated by these PS rules to produce a DS. A similar view identifies the starting point as a set of items drawn from the lexicon. Combining this set into a single constituent structure tree satisfying X-bar theory and the subcategorization features of the lexical items would yield a DS. The MP gives up any notion that the starting point of the derivation is a single-constituent structure tree; instead the MP claims that syntactic structures are built through generalized transformations that may insert already formed trees into trees. However, although derivations have no DS starting point, still the principle that derivations must compete (since a grammatical derivation is the most economical one from a set of competing derivations) requires some sort of "base" – something shared by competing derivations. It is tempting to try comparing derivations whose LFs have the same interpretation, but the notion of "same interpretation" or even "same LF" is difficult to make precise.

Chomsky relies instead on the notion that competing derivations are those that make use of the same lexical resources. There is a sense, then, in which MP derivations start from a set of lexical resources (see (2) above). Computation, as we will see immediately, involves putting lexical items together and competition among derivations involves comparison of computations on the same set of lexical items. I depart here from Chomsky's (1992) exposition of tree-building to provide an account of the mechanisms that I find more intuitive.

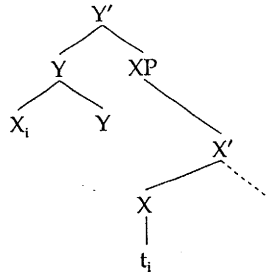
Consider the construction of a sentence to take place in a working area onto which one has already spilled some of the contents of the lexicon. The lexical items in the working area are all those that the computational system will have access to in the course of the derivation; moreover, the economy principles will compare derivations using this same set of lexical items. We will understand the lexicon to contain unlimited tokens of each lexical entry (e.g. as many tokens of the lexical item *John* as one wants); thus, the set of items in the working area may include more than one token of a lexical entry. The lexical items in the working area are completely formed words, fully inflected for case, agreement, tense, etc., a point we will return to below. A move in this working area consists of taking some piece already in the area and adding on to it in some way, perhaps by adding another piece in the area. The expanded constituent now becomes available, like all the lexical items already in the area, for further expansion or combination. At some point, the point of "Spell-Out" in (2), one may decide to take a constituent from the working area and submit it to the LF and PF components for interpretation. In the case of a grammatical representation of a sentence, the constituent removed from the working area will be a CP (or IP) containing all the lexical items originally spilled onto the working area and will (eventually) meet the interface conditions at LF and PF.

The crucial notion of competition or comparison of derivations in the MP is relativized to the derivations one might perform with a single set of lexical items in the working area. For example, suppose that there is a general economy principle of "laziness" that prefers doing nothing computationally over doing something. Clearly, if this were the most highly valued principle, we'd never produce constructions that are well-formed at the interface levels of PF and LF since the laziest thing to do with a set of lexical items is to leave them alone, in an unstructured pile, in the working area. Suppose, then, the need to create a structure well-formed at the interface levels (the desire for convergence at the interfaces) is stronger or more highly ranked than laziness. Then the computational system should do the minimal amount of work with the lexical items in the working area to insure convergence. In principle, then, one would compare all the different derivational paths one could take with the lexical items to find the laziest path that yields convergence. The derivation on this path would win the competition among derivations for this particular set of lexical items and be the only grammatical derivation for this set.

In the working area (i.e. before Spell-Out), any operation on a constituent must expand the constituent structurally, to create a larger constituent. (This expansion or "extension" requirement will need an exception, to be discussed below, to allow head-movement.) The basic operation is one of Projection, in the sense of X-bar theory, where an X projects an X′ or an X′ projects an XP (in the further development of the MP described in chapter 8, Chomsky abandons X-bar theory and proposes a conceptually simpler system of syntactic composition). The computational system targets some constituent in the working area and projects from this constituent. Within this framework, the question of whether adjunction structures may be created in the working area becomes one of whether X can project to X, X′ to X′, and XP to XP. When projecting, one has the option of including an empty category as sister to the constituent being projected, i.e. as a sister to the targeted constituent. This empty category must immediately be filled by either another constituent from the working area or by a constituent from within the targeted constituent itself. If we fill the empty category with another constituent from the working area, we have performed a generalized transformation, one that puts together what have otherwise been separate derivations. If we fill the empty category with a constituent from within the category that we are expanding, we have movement, a singulary transformation. However, movement in the MP is equivalent only to copying, not to copying and deletion. Chains formed by movement consist, then, in a sequence of copies of the "source" constituent. Where the chain is pronounced and where it is interpreted is a matter for independent principles to decide. One possible empirical advance of the MP is in its treatment of "reconstruction effects" (see section 5 below). The analysis of chains just described allows for reconstruction without any special operation; unless something else is said, a "moved" constituent exists at its source location as well as at the head of the chain of movement.

At least one sort of movement that we would wish to occur before Spell-Out cannot be thought of as expanding a targeted constituent. Suppose, as in (5), we do head-movement of the head X of an XP to adjoin to the head Y of YP, where XP is the sister of Y in Y′. This movement might expand Y through adjunction, but it could not expand Y′, which would be the minimal constituent one could target to move X to Y.

(5)



It is possible to say, as Chomsky suggests, that adjunction before Spell-Out need not expand a targeted constituent. This would allow head-movement before Spell-Out. However, if adjunction need not expand a targeted constituent, then the principle of expansion cannot insure a strictly cyclic derivation prior to Spell-Out. Without the necessity to expand the target, head-movement could reach down and move a head deeply embedded in a targeted constituent and adjoin it to another head also deeply embedded. In fact, nothing essential to the basic computational mechanisms here would prevent downward movement and adjunction of a head.

At any point in the computation, one may take the contents of the working area and submit it to the PF and LF components. As noted above, this splitting of the derivation on the separate paths to the interfaces is known as "Spell-Out." After Spell-Out, the generalized transformation of the computational system is no longer active. Thus, if the contents of the working area have not been combined into a single structure at Spell-Out, nothing in PF or LF will be able to put the pieces together, and the unjoined set of items will presumably not be well-formed (as a constituent) at the interfaces. If a derivation is to meet interface conditions, then, all the contents of the working area – all the lexical items – must be combined prior to Spell-Out.

If operations in the working area are limited to the projection and expansion of constituents, we insure that derivations will appear to obey a strict cycle condition. One cannot go back inside a constituent and do further expansion, nor is downward movement a possibility. Questions arise about operations – limited to singulary transformations – in the PF and LF components. Clearly movement at LF must occur within constituents already built in the working area. For example, in English main verbs will raise to Tense and to AGR at LF in the MP (see below); this raising occurs within already constructed AGRPs and TP. Relative to operations already performed before Spell-Out, then, movement at LF will violate a strict cycle condition. If, however, we adopt the same project and expand formalism for movement at LF as for movement before Spell-Out, then still at LF downward movement will be impossible since movement will always attach some constituent from within a target phrase to a constituent containing the target phrase (head-movement will still require special consideration). However, one must still ask whether any independent principles are required to insure a bottom-up derivation at LF for upward movement operations. Will problems arise, for example, if we allow LF-movement in a higher clause followed by LF-movement in an embedded clause?

On the PF side of the grammar after Spell-Out, operations are assumed not to conform to the basic mechanisms of the computational system. Importantly, then, movement to feature-checking positions does not occur on the way to PF after Spell-Out. If features are unchecked at Spell-Out, they remain so at the PF interface and will cause a derivation to crash, if they are visible at PF ("strong").

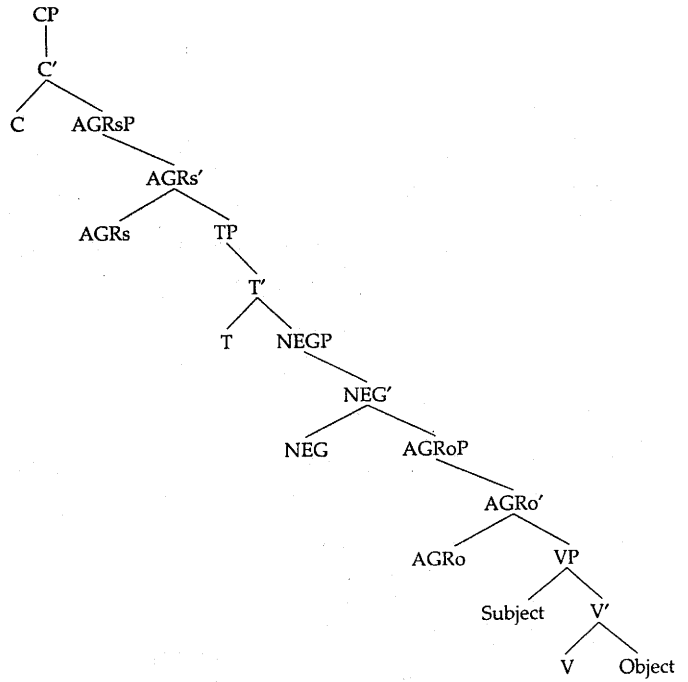
3 Features and Convergence

At the heart of the MP are what Chomsky calls "morphological features." These include features associated with tense, case, and agreement. Items from lexical categories such as V, N, and Adj, are fully inflected in the lexicon. The addition of, e.g., tense and agreement morphology (such as phonologically overt affixes) to a Verb in the lexicon involves the simultaneous addition of tense and agreement features, features that play a role in the computational system of language but which play no role at the PF or LF interfaces. These features, then, if potentially visible to these interfaces, must be eliminated prior to PF and LF. Failure to eliminate morphological features prior to an interface at which they are visible causes a derivation to crash (fail to converge) at this interface.

The functional categories of AGR and T in this system are the locus of tense and agreement features that may check off or eliminate the corresponding features on a Verb that moves up and adjoins to these categories. AGR and T also contain Case and phi-features that they may check off against features of NPs (DPs) that raise to their specs. In the MP, these functional nodes never contain items from the lexicon; they are not the positions in which inflectional affixes are inserted. Rather, inflectional affixes are attached to items of the lexical categories in the lexicon. The functional nodes of AGR and T serve only to carry the morphological (inflectional) features necessary to check off the features on Ns and Vs.

The hierarchical structure of a CP will look something like that shown in (6) (NEGP will not be discussed in this chapter). Chomsky leaves open what principles determine the hierarchical ordering of the various XPs in (6) as well as the question of whether this hierarchical ordering might vary from language to language. In addition, there may be additional functional projections between VP and CP or above CP.

(6)



The AGRP above T(ense)P is labeled AGRs for "subject agreement" and the AGRP above VP AGRo for "object agreement." However, Chomsky claims that there is no difference between the sorts of features found at these two AGR positions and hence no difference in the category labels of the two positions. That is, hence is no intrinsic difference between the features of AGRs and AGRo, only a positional difference between the two AGR nodes. The contrast between the NOM Case usually licensed by AGRs and the ACC Case associated with AGRo results from the raising of T to AGRs and V to AGRo; the Cases are associated with features of T and V respectively. This is not to say

that the AGR features in AGRs and AGRo will be identical in every sentence. The AGR features in AGRs will include the phi-features (person, number, gender) of the subject DP that is raised to spec of AGRs while the AGR features in AGRo will include the phi-features of the object DP that is raised to spec of AGRo. However, the phi-features themselves – i.e. the AGR features – will come from the same pool of features and will not be identifiable as subject or object AGR features except by the location of the AGR node in which they reside.

To check its AGR and Tense features, the V will raise to AGR, to T, and to AGR, either before Spell-Out or at LF. The corresponding tense and agreement features on T and on the AGRs are known as the "V" features of T and AGR, i.e. the features that check off the features of the V. These functional nodes, as well as the Verb, also have "N" features, or features that check off features on NP (DP) arguments. The "Case Filter" of the earlier P&P theory divides into two parts in the MP. First, Ns in the lexicon (and/or perhaps Ds) must be assigned morphological case features before they are inserted into structures via generalized transformations. This requirement insures that DPs will have a need to move to a position to check off these morphological features prior to an interface level at which the features might be visible and cause a derivation to fail to converge. The second part of the Case Filter, that the Case features on DPs be checked off in the syntax, is now part of a much wider and more central system in the MP. Head-movement of Verbs to inflectional nodes as well as the movement of DP arguments to Case positions are now motivated by this generalized form of the Case Filter – i.e. the requirement that morphological features get checked off before the interface levels.

On the MP program, there is only one, generalized, relation that allows one element to license another, by checking off the latter's features. This basic relation derives from what Chomsky takes to be the basic relation of Agreement – the relation between a head and its specifier. Unlike in previous versions of the theory, no licensing takes place between a head and its complement; the relation of "government" plays no direct role in this theory. The licensing or "checking" domain of a head H includes the specifier of H, a head position adjoined to H, a position adjoined to the maximal projection of H, and a position adjoined to the specifier of H. Chomsky's definitions of structural relations puts all these positions in the "checking domain" of H. For the basic cases we will discuss here, the important positions in the checking domain of H are the specifier of H, where DPs will check Case and phi-features against H, and the position adjoined to H, since the Verb will adjoin to functional heads to have its (V) features checked by these heads.

4 Parameters and Basic Word Order

4.1 *Deriving Basic Clause Structure*

Within the MP, differences between languages are attributed to differences between the features of lexical items in the languages and specifically between the features of lexical items belonging to the functional categories AGR and Tense. Recall that Vs and Ns are taken from the lexicon fully inflected with inflectional affixes. The functional nodes in the syntax are not associated with affixes (nor with any phonological content whatsoever) but simply with certain features – Tense, Case, and Agreement features among others. Nevertheless, specific bundles of these features of the category AGR and T are lexical items and differences between the sets of bundles available in the lexicon account for cross-linguistic syntactic differences between languages in the MP.

Consider, then, how a basic sentential structure is derived for English within this framework. We might be deriving a simple transitive sentence, such as, "Hortense touched the porcupine." Subject and object DPs would be initially inserted via generalized transformations within the VP (see (6) above). That is, assume that two DPs, "Hortense" and "the porcupine," have been created with lexical items from the working area and redeposited in the working area. We target a (fully inflected) V, "touched," from the working area, project a V' with an empty complement position to the V and immediately replace the empty complement with one of the previously constructed DPs, "the porcupine," which will be the object. Next we target the V', project VP with an empty spec of VP position and replace this empty spec of VP with the other DP previously constructed, "Hortense," which is now the subject. The V will contain ACC Case features to check against the object DP at LF, as well as past Tense features to check against T. It also must have AGR features of both subject and object DPs to check against the AGRs that it will adjoin to.

Case features and phi-features (person, number, gender) would be associated with each DP argument. Chomsky leaves somewhat open where these features of the DP come from, i.e. which lexical items contribute these features to the DP as a whole. One possibility is that the features of the DP are features of the head D of the DP and that other constituents within the DP that bear affixes with, say, Case, gender or number features from the lexicon must check these features against the features in D at some point in the derivation. Another possibility is that the Case and phi-features of the DP are inherited in some manner from the features of the head N of the NP that serves as complement to D.

At some point in the derivation of our basic transitive sentence, an AGR from the working area would be targeted and would project an AGR' and the VP would be inserted as the sister to the AGR. In our example, for the derivation to converge, the AGR would need to contain the phi-features of "the porcupine." This AGR is identified as AGRo by its position governing VP, but not by any of its intrinsic features. As explained above, features of the functional categories are said to be either "strong" or "weak" with respect to their visibility at the PF interface. "Strong" AGR features are visible at PF if they are not checked off before the interface. "Weak" features are not visible. Visibility at PF does not correspond directly to any notion of "rich phonological content" since the functional nodes do not in any case contain any phonological material – no affixes. Whether an AGR or a T node has strong or weak features, a phonologically null constituent will, in general, appear at that node. In English, the "N-features of AGR" – the person, number, and gender features that AGR will check off against the corresponding features on a DP in spec of AGRP – are weak and therefore may appear (invisibly) at PF without causing a crash (without disturbing convergence). The principle of Procrastination demands that movement wait until after Spell-Out as long as waiting is compatible with convergence. Since an object DP need not move to spec of AGRP to check off the weak features there until after Spell-Out, Procrastinate demands that it not move before Spell-Out. Thus objects in English do not visibly move to the spec of AGRP position. Since this position is not filled before Spell-Out, it is not projected. Recall that if the spec of AGRP were projected, it would be filled by an empty category that would immediately need to be replaced with a constituent from the working area (in a generalized transformation) or with a constituent from within AGR' (a singulary transformation). There is no entity in the MP consisting of a position that is projected but not filled (although a position might of course be filled by an empty category, such as PRO or *pro*).

After we have created AGRP, a Tense is taken from the working area and projects a T' into which the AGRP containing the VP is inserted as a complement to T. The T in our example must contain the "V-feature" past tense to check against the past tense feature on the verb, "touched." The N-features of T in English – Nominative Case features in the case of finite tense – are strong. Thus, unless the (subject) DP raises to check the Nom features before Spell-Out, the derivation will crash at PF. However, the assumption is that the N-features of T, like the Case (N-)features of V, are checked in conjunction with AGR. Therefore, T will raise to a higher AGR and check its N-features against a DP in spec of AGRP. For reasons discussed in Chomsky (this volume), the subject raising to spec of this higher AGRP need not and will not pass through the spec of TP position, which will therefore not be projected.

Note again here that head-movement prior to Spell-Out causes a small technical problem for a strict interpretation of the principle that all operations prior to Spell-Out expand the constituent that they target. If we wish to move and adjoin a T from the TP serving as complement to the higher AGR in (6), we must target AGR′, for only the AGR′ contains both the AGR and T heads involved in the movement. But, adjoining the head T of the TP to the head AGR of the AGR′, as we must in English, does not technically expand the targeted AGR′. The AGR head itself may only be targeted when AGR′ is projected. It may be that adjunction in general must be allowed before Spell-Out and allowed to violate the "always expand the targeted constituent" principle. Alternatively, the definitions of "targeting and expanding" a constituent would have to be redefined to allow for head-movement and adjunction as a type of expansion.

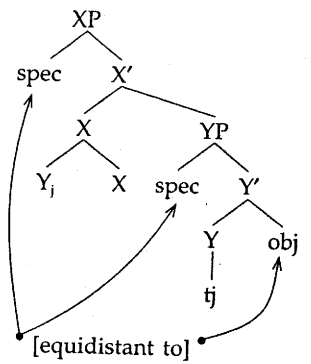
Making the N-features (Case features) of T strong in English is a way to implement the Extended Projection Principle (EPP) of earlier approaches, i.e. the requirement that sentences have (overt) subjects. Within Chomsky's particular set of assumptions here, it is not possible to capture the EPP for English by making the N-features of AGR strong. Recall that AGRs and AGRo are not distinguished by their intrinsic features. If the N-features of AGR are strong, then the N-features of the AGR above VP will be strong, in addition to the N-features of the AGR above TP. Thus, if the N-features of AGR were strong in English, objects as well as subjects would have to raise to spec of AGRP before Spell-Out and we would expect to pronounce the object (higher than and) before the verb.

Following along our sample derivation of "Hortense touched the porcupine," to produce a structure like (6) above, we target the higher AGR′, project AGR(s)P, and move the subject from within the VP to the spec of AGRP position. Assume that we have already adjoined T to AGR and that the combination of AGR and T may check the Case and phi-features of the subject DP "Hortense." The movement of the subject from spec of VP to spec of AGRP crosses two potential spec positions, spec of the lower AGRP and spec of TP. However, neither of these positions has been projected, and neither will be projected prior to Spell-Out. Therefore, this movement does not violate the Shortest Move economy principle. To be more accurate about the determination of grammaticality here: we evaluate a derivation in which we move the subject from spec of VP to spec of AGRP against other alternative derivations from the same lexical resources in the working area. Since the distance of movement is computed relative to projected positions of the same sort as the end-point of movement, there is no alternative derivation that gets the subject to spec of AGRP in which the subject takes a shorter movement.

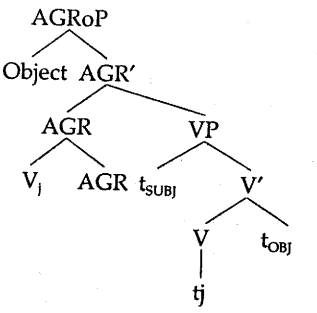
After Spell-Out, "at LF" (i.e. in the computation of LF), the object, "the porcupine," will move to spec of the lower AGR(o)P to check its Case and phi-features. Here it is assumed that the ACC Case feature is associated with the V but checked by the V in combination with AGR, just as NOM is associated with finite Tense but checked by Tense in combination with AGR. The movement of the object to spec of AGRoP crosses a projected spec position, the spec of VP which is occupied by the tail of the A-chain of the subject "Hortense" (i.e. by the trace of the subject). Unless some other consideration applies here, this movement over a projected spec position should be a violation of Shortest Move, here clearly connected to Relativized Minimality. A derivation in which the object moves to spec of VP position on its way to spec of AGRoP (and covers or adjoins to the trace of the subject) would eventually violate other principles or at least be uninterpretable at LF. However, when considering whether a movement would violate Shortest Move, all that matters is that there is a shorter movement available, not that taking this shorter move in some derivation from the same lexical resources could lead to convergence. Shortest Move, unlike Greed and Procrastinate, is ranked higher than convergence in comparing derivations from the same lexical resources (see the discussion of (1b) above).

So how is a violation of Shortest Move avoided when the object moves to spec of AGRoP? At LF, the verb in English will raise to AGR, with the AGR+V combination raising to the T+AGR combination created before Spell-Out. The V has the N-features of (accusative) Case that it will check off, in combination with AGR, against a DP in spec of AGR(o)P. In addition, it has Tense features and AGR features of the subject, which it will check off when it gets to the highest AGR node. Chomsky defines a notion of equidistance that allows the spec of XP and the spec of a YP that is a complement to X to be equidistant from an element inside Y′ just in case Y adjoins to X, as shown in (7a).

(7a)



(7b)



In the situation of interest, (7b), the spec of AGRoP and the spec of VP are equidistant to the complement of V after V raises and adjoins to AGR. Since these specs are equidistant from the object, a derivation in which the object raises over spec of VP to spec of AGRP does not lose out for economy reasons to a derivation in which the object lands in the spec of VP.

The sample derivation discussed above began with the assumption that we had all the proper lexical items in our working area to allow for convergence. In particular, the subject N (or D) "Hortense" was assumed to have NOM Case from the lexicon, the Case that would be checked by the N-features of T in combination with the AGR above TP. The object N (or D), "porcupine" or "the," was assumed to have ACC Case from the lexicon, i.e. the Case that would be checked by N-features of V in combination with the AGR above VP. Suppose, however, that the N in the object DP came from the lexicon with nominative Case and the N in the subject DP came with accusative Case. Could there be a convergent derivation in which the complement to V ends up in spec of the higher AGRP and the specifier of VP ends up in the spec of the lower AGRP, a derivation in which we derive "The porcupine touched Hortense," with the meaning of "Hortense touched the porcupine"?

For the core case of English sentences containing main verbs, such a derivation would not be grammatical. Recall that for the object to raise over the spec of VP at all and not violate Shortest Move, the verb must raise to AGR and the object must land in the spec of the lower AGRP. The strong N-features of T in English require that the DP that ends up in spec of the higher AGRP move to this position before Spell-Out. However, main verbs (i.e. verbs that are not auxiliary verbs) do not raise before Spell-Out in English. Since main verbs do not move before Spell-Out, the object may not raise over a (filled) spec of VP position before Spell-Out and thus may not fill the spec of the higher AGRP before Spell-Out, as required if the underlying object were to have NOM Case checked and end up as the surface subject.

Of course we must ask why it is that main verbs do not raise before Spell-Out in English. Could we violate whatever principle is being obeyed by their failure to raise in order to insure convergence, should we happen to put a DP with NOM Case in the complement position of a transitive VP? We assume that the V-features of T and AGR are weak in English. Thus the verb need not raise to T and AGR to check its features before Spell-Out and, by Procrastinate, it therefore may not raise. However, we have seen that Procrastinate can be violated to insure convergence, as when the subject raises to spec of AGRP before Spell-Out in English. Therefore, perhaps the verb could raise before

Spell-Out in English to allow a NOM DP to escape the VP and hea toward spec of the higher AGRP to have its Case checked.

Other considerations rule out such derivations of our "inverse" sen tence, "The porcupine touched Hortense," although it is not clear whic principles should be considered crucial to derive this result. Recall tha movement of a head Y to the head X where YP is complement to makes the spec of YP and the spec of XP equidistant from material i the complement of Y, as shown in (7a). If we further move X (with adjoined) to the head Z of ZP where XP is the complement to Z, w make spec of ZP and spec of XP equidistant from material in th complement to X. However, although the spec of ZP and the spec of X are equidistant from material below X' and the spec of XP and the spe of YP are also equidistant from material below Y', the spec of ZP an the spec of YP are not equidistant from the complement of Y; equi distance is not transitive in this way. Thus an object may move ove the spec of VP if the V raises to AGRoP, but it may only move as far a the spec of AGRoP from the complement of V position in one step i spec of VP is filled; no matter how high the AGR to which V is adjoine may move, the spec of VP will be equidistant at most with the spe of AGRoP to the object in the complement of V position.

As a result of the definition of equidistance, if we tried to move NOM marked object up to spec of the higher AGR(s)P, it would b necessity move through the spec of the lower AGR(o)P. In the simpl situation under consideration, where we derive "The porcupine touche Hortense," this lower spec of AGRP is a Case-checking position, position for checking the ACC Case assumed to be on the (underlying subject DP, "Hortense." Thus Greed might prevent the object fron leaving this spec of AGRP position to move onto the higher spec o AGRP position and check NOM Case, assuming that Greed in genera prevents movement from a Case-checking position to another Case checking position. Alternatively, there might be reasons why the sub ject DP "Hortense" should be prevented from moving into spec of th lower AGRP at LF, covering the trace of the object. In short, the "re verse" derivation in which a subject moves to spec of AGRoP and th object moves into spec of AGRsP is prohibited if principles prevent th object from moving through the spec of AGRoP on its way to AGRsP since Shortest Move requires this movement.

4.2 *Parametric Differences*

Recall that Chomsky assumes that the significant parametric differ ences between languages are limited to lexical differences, specifically

differences in the features of the lexical elements that occupy the functional category nodes. For basic parametric differences in constituent (word) order, Chomsky looks to the AGR and T nodes and their N-features and V-features. The N-features are those that are checked off against a DP in spec of AGRP (or, potentially, spec of TP) and the V-features are those that are checked off against a V that adjoins to a functional head. These features may be either weak – invisible at PF even if unchecked – or strong – visible at PF if unchecked – in a language. For AGR and T independently, there are four different combinations possible of weak and strong N- and V-features (i.e. strong N with strong V, strong N with weak V . . .). Since the strength of N- and V-features on AGR is at least conceptually independent of their strength on T, the four combinations of features on AGR can combine freely with the four combinations on T to yield 16 possible language types defined by the strength of their morphological features in AGR and T. We saw that English has strong N-features on T but weak V-features. The strong N-features require a DP (subject) to move to spec of AGRsP before Spell-Out; the weak V-features permit the V to stay in VP before Spell-Out. Both the N- and V-feature of AGR must be weak. If the V-features were strong, the verb would have to raise out of the VP before Spell-Out. If the N-features of AGR were strong, the object would have to raise to spec of AGR(o)P before Spell-Out.

In French, main verbs do raise to Tense before Spell-Out in finite clauses. Evidence for this difference between English and French comes from, e.g., the relative positions of tensed main verbs and VP-adjoined adverbs, as in (8). English leaves the main tensed verb inside the VP before Spell-Out, and thus to the right of the adverb. French raises the verb to T before Spell-Out, and it is pronounced to the left of the adverb.

(8a) Elmer lave souvent son chat
Elmer washes often his cat
(8b) Elmer often washes his cat

Thus the V-features of either AGR or T or both must be strong in French, in contrast to English. Since objects do not raise out of the VP before Spell-Out in French, the N-features of AGR must be weak, as in English. In adopting this account of the differences between French and English, Chomsky is building on the pioneering work of Emonds and the recent analysis of Emonds's (1978) insights in Pollock (1989).

Chomsky also suggests that the N-features of AGR and T might be weak in verb-initial languages such as Irish. With weak N-features on all functional heads, the subject and object of verbs in these languages could remain in the VP until after Spell-Out. If the V-features of T and/or AGR were strong, the verb would be forced to raise from the VP prior to Spell-Out in these languages, yielding a VSO order of major constituents. The relative positioning of VP adverbs, negative morphemes and the inflected verb in various languages has been widely studied within the MP recently.

5 Copy Theory of Movement

In addition to the various changes to standard P&P theories we have described above, Chomsky's MP adds an explicit endorsement of the copy theory of movement. Thus, when a singulary transformation replaces a projected empty category in position X with a constituent from position Y, both position X and position Y – both members of the chain created by "movement" – contain a copy of the "moved" constituent. After Spell-Out, on the way to both PF and LF, the grammar must decide which members of the chain are pronounced and which are interpreted. In standard cases of A- or A-bar movement, e.g. passive or *wh*-movement, the constituent whose copies form the chain is pronounced at the head of the chain, in either the Case(-checking) position for A-movement or the A-bar operator position for A-bar movement. Similarly, any thematic role associated with the constituent is usually assigned to the copy at the tail of the chain, so in a sense the constituent is interpreted in this position.

The copy theory of movement and the adoption of generalized transformations provides an approach to various types of "reconstruction effects." For example, in standard cases of A-bar reconstruction effects, the proper application of standard Binding Theory requires putting back some or all of an A-bar moved constituent. Treating (9a) with standard Binding Theory at LF necessitates finding the reflexive back in the trace position, as in (10a). Under a copy theory, although the whole moved constituent is pronounced at the head of the chain (with perhaps PF deletion of the material in the copy of the moved constituent at the tail of the chain), at LF we may delete all but the *which* at the head of the chain and nothing but the *which* at the tail of the chain, deriving (in 10a) a possible input to interpretation and producing a suitable candidate for Binding Theory. In the case in (9b), where complete reconstruction would produce a Condition C violation (cf. (9c)), we may choose a different sort of deletion pattern at LF, as shown

in (10b). Here we delete all the material at the tail of the chain and retain the material in the operator at the head of the chain.

(9a) Which pictures of himself did Mary say John saw t
(9b) Which pictures of $John_i$ did Mary say he_i saw t
(9c) *Mary said he_i saw those pictures of $John_i$
(10a) $Which_x$ Mary said John saw [x pictures of himself]
(10b) Which x, x a picture of John, Mary said he saw x

Chomsky also suggests that adopting generalized transformations will allow him to incorporate insights about the interaction of adjunction and Binding Theory from Lebeaux (1988). Consider the basic contrast in (11); in (11a) coreference between *John* and *he* is allowed, in contrast to (11b) where a similar coreference relation seems ill-formed.

(11a) Which claim that $John_i$ made did he_i regret?
(11b) ?*Which claim that $John_i$ runs did he_i deny?

Assume that the *that* clause in (11b) is the complement to the N *claim*. Assume also that relative clauses, like *that John made* in (11a), are adjoined to DP (or NP). Finally, assume that adjunction falls outside the requirement that an operation in the computational system before Spell-Out must always expand the targeted constituent. Thus relative clauses, as adjuncts, may be added to a DP after this DP has already been incorporated into a sentence via a generalized transformation, and even after this DP has undergone movement. Putting these assumptions together with the copy theory of movement, one may explain the contrast in (11). In (11a), the relative clause is adjoined to *which claim* after this constituent has raised to spec of CP. Thus the copy of the *which claim* in the position c-commanded by *he* in (11a) does not contain *John*, and no Condition C violation is expected. On the other hand, a complement to N must be incorporated into the structure when the N′ is projected. Thus in (11b), the complement *that John runs* must be part of the copy of the constituent *which claim that John runs* in the trace position c-commanded by *he*. We expect then a Condition C violation in (11b) but not in (11a), since there is a derivation of (11a) in which *he* does not c-command (a copy of) *John* at LF.

There is some tension between the explanation of reconstruction effects in (9) and the anti-reconstruction effects in (11). Following the analysis of (9b) above, we might be able to delete most of the trace copy of *which claim that John runs*, leaving *which x, x a claim that John runs* in the operator position and removing *John* from a position c-commanded by *he* at LF. Chomsky discusses approaches to the contrast between reconstruction and anti-reconstruction effects in Chomsky (1992).

6 Some Residual Issues

Although recovering within the MP all the data previously accounted for in earlier versions of P&P syntax is beyond both Chomsky (1992) and this summary, I will briefly discuss here a few remaining issues and constructions. First, although main verbs do not raise from the VP to AGR or T prior to Spell-Out in English, auxiliary verbs do. To account for the behavior of English auxiliary verbs, one might give them strong N-features for the phi-features of subjects or strong Tense features. In either case, the strong features on auxiliaries would force their movement before Spell-Out, in contrast to main verbs. However, Chomsky proposes instead that the behavior of English auxiliaries might be principled and follow from their lack of semantic content. "Have" and "be" in particular, used as auxiliaries, bear tense affixes and appear in connection with participles to express tense and aspect, but carry no particular semantic value in themselves. If "have" and "be," being semantically vacuous, were truly invisible after Spell-Out in the LF part of the grammar, they could not be raised to Tense and AGR to check off features there. Movement prior to Spell-Out of these verbs would then be forced, in violation of Procrastination, since Procrastination only compares derivations that converge at LF and PF. In addition, Greed will be violated by the pre-Spell-Out movement of auxiliaries. Since the auxiliaries are invisible at LF, any morphological features they carry would presumably be invisible as well. Therefore, when the auxiliary verbs move prior to Spell-Out to insure convergence, they are acting altruistically to check off the features of AGR and T, not with Greed to check off their own features.

This solution to the problem of distinguishing main and auxiliary verbs, together with other assumptions of the MP, leaves Chomsky without an account of *do*-insertion (to replace that from Chomsky's earlier "Economy" framework (1989)). We know that auxiliary verbs raise before Spell-Out over *not* to T and AGR in negative sentences like (12a) and over the subject to C in inversion contexts like (12b).

(12a) John is not leaving here soon enough for me
(12b) Is John leaving here at all?

Assume that a tensed main verb agreeing with a subject would need to move to T and AGR to check features by LF. If something prevented the main verbs from moving to T and AGR after Spell-Out, Procrastinate could be violated to assure convergence, and main verbs like auxiliaries should raise before Spell-Out to the AGRs and T in negated sentences – (13a) – and on to C in inversion contexts – (13b). Compare the Procrastination-violating movement of auxiliary verbs prior to Spell-Out to insure convergence.

(13a) *John leaves not until he's told
(13b) *Leaves John before the food arrives?

A number of possible MP approaches to *do*-insertion suggest themselves (and see Watanabe (1993)). Given the assumption that we compare only derivations using the same lexical resources and the assumption that all inflected verbs and all functional nodes are lexical items, we cannot compare derivations of constructions like (13) with constructions containing "do," as in (14). The lexical resources in (13) and (14) are different and thus the derivations of these structures should not compete.

(14a) John does not leave until he's told
(14b) Does John leave before the food arrives?

Therefore, the key to explaining the *do*-insertion facts in the MP as presented has two parts. First, we must make sure that the most economical derivations involving the lexical resources in (13) fail to converge. For example, if Shortest Move were to be violated by the movements of the main verbs in (13) (but not by the parallel movement of auxiliary verbs in (12)), then we're in business, since Shortest Move generally takes precedence over convergence. It is not immediately clear, however, how the derivation of the structures in (13) might violate Shortest Move.

To complete the account of *do*-insertion, there is no problem allowing the derivations of the structures in (14), which should be economical given the choice of lexical resources they use. Rather, the second problem in explaining *do*-insertion is to rule out the use of "do" outside of *do*-insertion contexts, e.g. in (15).

(15) *John does leave before he's told. (non-emphatic: unstressed *does*)

Here again we cannot compare the derivation in (15) to the derivation of "John leaves before he's told," since these employ different lexical resources. A new account of *do*-insertion within the MP needs to find an alternative way to account for the apparent "last resort" character of the use of "do."

Another difference between standard P&P theories and the MP involves the treatment of Exceptional Case Marking (ECM) constructions. On the view of Case checking outlined above, structural Case is never checked under government but only in a spec-head relation with AGR and T or V. To have ACC Case checked, then, objects in English must raise from the VP at LF and land in spec of AGRoP. Similarly, in ECM constructions such as (16), the lower subject must raise at LF out of the lower clause and to the position in the higher spec of AGRoP marked by the ACC in (16).

(16) I [ACC [believe [$John_i$ [to have been chosen t_i for the job]]]]

Although this "raising to object" takes place at LF, not before Spell-Out, raising to object should have recognizable consequences since such principles as those of binding theory apply at LF. Lasnik (1993a) explores some of these, reviving arguments from Postal (1974).

In ECM constructions like (16), interesting issues arise around the overt presence of *John* in spec of AGRsP of the lower clause, raised from the trace position in lower VP (the lower object position). Since *John* doesn't check its Case or phi-features until it moves up to spec of the higher AGRoP at LF (the "ACC" position in (16)), the movement from its base position within the lower VP to spec of AGRsP in its own clause seems unmotivated, a violation of Procrastinate. Apparently Procrastinate is not being violated to insure convergence since it does not appear that strong features are being checked in the lower spec of AGRsP.

To explain the behavior of ECM and raising constructions, a revised theory of infinitivals has been articulated (see Chomsky and Lasnik (1991)). The infinitival Tense of control constructions, those with subject PROs, are assumed to check Case on PRO in spec of AGRsP, perhaps NOM Case, but to be restricted to checking the features only of PRO. So a PRO that reaches the spec of AGRsP of such a control infinitival must stay there, with all its features checked, and a non-PRO that reaches this spec will fail to have its features checked and will cause a crash. Subject position of an infinitival in control constructions (and in any construction that permits an "arbitrary" PRO subject) is thus just like the subject position of a tensed clause, with the exception that only PRO is allowed to occupy this position.

On the other hand, the Tense of an infinitival clause like that in (16), an infinitival Tense of an ECM or raising construction, must check some features against any DP in spec of AGRsP, be it PRO or overt.

in (10b). Here we delete all the material at the tail of the chain and retain the material in the operator at the head of the chain.

(9a) Which pictures of himself did Mary say John saw t
(9b) Which pictures of $John_i$ did Mary say he_i saw t
(9c) *Mary said he_i saw those pictures of $John_i$
(10a) $Which_x$ Mary said John saw [x pictures of himself]
(10b) Which x, x a picture of John, Mary said he saw x

Chomsky also suggests that adopting generalized transformations will allow him to incorporate insights about the interaction of adjunction and Binding Theory from Lebeaux (1988). Consider the basic contrast in (11); in (11a) coreference between *John* and *he* is allowed, in contrast to (11b) where a similar coreference relation seems ill-formed.

(11a) Which claim that $John_i$ made did he_i regret?
(11b) ?*Which claim that $John_i$ runs did he_i deny?

Assume that the *that* clause in (11b) is the complement to the N *claim*. Assume also that relative clauses, like *that John made* in (11a), are adjoined to DP (or NP). Finally, assume that adjunction falls outside the requirement that an operation in the computational system before Spell-Out must always expand the targeted constituent. Thus relative clauses, as adjuncts, may be added to a DP after this DP has already been incorporated into a sentence via a generalized transformation, and even after this DP has undergone movement. Putting these assumptions together with the copy theory of movement, one may explain the contrast in (11). In (11a), the relative clause is adjoined to *which claim* after this constituent has raised to spec of CP. Thus the copy of the *which claim* in the position c-commanded by *he* in (11a) does not contain *John*, and no Condition C violation is expected. On the other hand, a complement to N must be incorporated into the structure when the N' is projected. Thus in (11b), the complement *that John runs* must be part of the copy of the constituent *which claim that John runs* in the trace position c-commanded by *he*. We expect then a Condition C violation in (11b) but not in (11a), since there is a derivation of (11a) in which *he* does not c-command (a copy of) *John* at LF.

There is some tension between the explanation of reconstruction effects in (9) and the anti-reconstruction effects in (11). Following the analysis of (9b) above, we might be able to delete most of the trace copy of *which claim that John runs*, leaving *which x, x a claim that John runs* in the operator position and removing *John* from a position c-commanded by *he* at LF. Chomsky discusses approaches to the contrast between reconstruction and anti-reconstruction effects in Chomsky (1992).

6 Some Residual Issues

Although recovering within the MP all the data previously accounted for in earlier versions of P&P syntax is beyond both Chomsky (1992) and this summary, I will briefly discuss here a few remaining issues and constructions. First, although main verbs do not raise from the VP to AGR or T prior to Spell-Out in English, auxiliary verbs do. To account for the behavior of English auxiliary verbs, one might give them strong N-features for the phi-features of subjects or strong Tense features. In either case, the strong features on auxiliaries would force their movement before Spell-Out, in contrast to main verbs. However, Chomsky proposes instead that the behavior of English auxiliaries might be principled and follow from their lack of semantic content. "Have" and "be" in particular, used as auxiliaries, bear tense affixes and appear in connection with participles to express tense and aspect, but carry no particular semantic value in themselves. If "have" and "be," being semantically vacuous, were truly invisible after Spell-Out in the LF part of the grammar, they could not be raised to Tense and AGR to check off features there. Movement prior to Spell-Out of these verbs would then be forced, in violation of Procrastination, since Procrastination only compares derivations that converge at LF and PF. In addition, Greed will be violated by the pre-Spell-Out movement of auxiliaries. Since the auxiliaries are invisible at LF, any morphological features they carry would presumably be invisible as well. Therefore, when the auxiliary verbs move prior to Spell-Out to insure convergence, they are acting altruistically to check off the features of AGR and T, not with Greed to check off their own features.

This solution to the problem of distinguishing main and auxiliary verbs, together with other assumptions of the MP, leaves Chomsky without an account of *do*-insertion (to replace that from Chomsky's earlier "Economy" framework (1989)). We know that auxiliary verbs raise before Spell-Out over *not* to T and AGR in negative sentences like (12a) and over the subject to C in inversion contexts like (12b).

(12a) John is not leaving here soon enough for me
(12b) Is John leaving here at all?

These features must be strong to insure movement of the subject to this position before Spell-Out, in violation of both Procrastination and Greed. However, whatever these strong features are, they must not check off the Case and phi-features of a subject, which must move on from the subject position of the embedded infinitival to spec of a higher AGRsP or AGRoP. In raising to subject constructions, the further movement to spec of AGRsP takes place before Spell-Out in English; in the ECM constructions, as we have seen, the further movement to spec of AGRoP takes place at LF. These strong features of non-finite Tense in raising and ECM constructions might be called "Extended Projection Principle" features, because, like the earlier EPP, they insure that sentences will have an "overt" subject (phonologically overt, trace, *pro*, or PRO).

Note that the movement of the subject out of the lower VP in the ECM construction in (16) had to be forced by (strong) features that needed checking in the landing site of the movement. In general, movement is not optional in the MP and overt movement is always forced by strong features. This implies that overt *wh*-movement in English must be forced by strong *wh*-features in C, for example.

If all movement is forced and no movement is optional, the MP faces a challenge in apparent optional movement of constituents, e.g. in Scrambling in Japanese and other languages with apparent freedom in word order. Some differences in constituent order might be accounted for through the imaginative use of some of the functional heads in (6). For example, suppose that all the N features of T and AGR are potentially weak in Japanese. Then no DP need raise from the VP to check Case or phi-features prior to Spell-Out. Now if some of these N-features are optionally strong, movement of some of the constituents from VP but not others could be forced. For example, the N-features of V might optionally be strong, forcing movement of V to AGRo and movement of the object to spec of AGRoP prior to Spell-Out. If all other relevant features are weak, the object will appear to have scrambled over the subject. The optionality in Scrambling on this account would not be in the movement itself but in the choice of lexical items, a choice between a lexical item with strong or weak features. Other approaches to Scrambling within the MP are of course conceivable and are being explored in the literature.

7 Reactions and Extensions

In previous versions of P&P syntax, functional heads such as Tense (or Infl) were often the sites of lexical insertion of actual affixes – lexical items with phonological as well as morphological features. Certain types of word formation, then, were accomplished via syntactic operations on syntactic heads (see Marantz (1984) and Baker (1988a), for example). Attempts were made to explain aspects of the internal structure of words through their construction via syntactic operations.

Under the MP, all inflected words are formed in the lexicon. The question arises, then, whether all explanations of the distribution of morphemes within words should be left to whatever principles govern affixation in the lexicon. In addition, Chomsky explicitly limits the computational system to the path from lexical resources to LF. If any operations on tree structures occur between Spell-Out and PF, they would seem to fall outside the mechanisms, if not the principles, characteristic of the computations in the syntax proper.

In proposing and detailing the theory of "Distributed Morphology," Halle and Marantz (1992) suggest that functional heads like Tense and AGR should in fact serve as the locus of lexical ("Vocabulary") insertion and that all word formation should occur in the syntax, as a result of the syntactic combination of heads. In addition, we argue that the operations taking place in a derivation between Spell-Out and PF are of the same sort and obey the same principles as the operations in the rest of the syntax. The last section of Halle and Marantz (1992) lays out what is at stake in the differences between the assumptions of the MP and those of Distributed Morphology.

Between the introduction of the MP and the writing of this chapter, a large research effort has applied the core of the program to a wide range of languages, with various modifications and extensions of MP ideas and principles. Some energetic work within the MP has centered on the analysis of ergative languages and ergative constructions. For example, Murasugi (1992) modifies principles and assumptions of the MP to allow the subject to land in the spec of AGRoP and the object in the spec of AGRsP in ergative languages. Bobaljik (1993) provides an alternative analysis of ergativity more consistent with the MP as described here.

Much recent MP research has concentrated on the overt appearance of various arguments in the specifier positions above VP. For example, Carnie and Bobaljik (Carnie 1993) propose that the subject in Irish raises to spec of TP before Spell-Out, rather than remaining within the VP. VSO order results when the verb raises to AGRs. Comparative research on the Germanic languages has explored "object shift" and other apparent movements that place arguments to the left of adverbs and other constituents thought to mark the left edge of the VP. Recall that the verb must raise to AGRo to allow the object to move past the spec of VP position and escape the VP; this raising makes the spec of VP and

spec of AGRoP equidistant to the object in complement to V position. If the object is to leave the VP prior to Spell-Out, the verb must also raise prior to Spell-Out, i.e. overtly. One apparent consequence of the MP, then, is to predict a correlation between overt movement of the V from the VP and "object shift" – overt movement of the object out of the VP (see Bures (1992) and Jonas and Bobaljik (1993)).

Longer works that propose significant modifications and/or extensions of the MP include Branigan (1992), Watanabe (1993), Lasnik (1993a), Kitahara (1994b), and Zwart (1993). For a collection of recent papers that include applications of the MP as well as extensive bibliographic references, see Bobaljik and Phillips (1993) and Phillips (1993).

8 The End of Syntax

In closing, I would like to discuss a certain radical flavor to the MP in Chomsky (1992) and in his Bare Phrase Structure theory of chapter 8. In contrast to the wide-ranging discussion of somewhat intricate data from a number of languages found in Chomsky (1981), for example, Chomsky's latest papers (1992, this volume) treat very little data, and the discussion of data itself is somewhat programmatic. We should not interpret this move to minimalist syntax as a rejection of the enormous volume of extraordinary work within the P&P approach since the early 1980s. On the contrary, this detailed and highly successful work on a wide range of languages has inspired Chomsky to envisage the end of syntax per se. From one point of view, explanations in current syntactic work are emerging at the interfaces with phonology, and, perhaps more extensively, with semantic interpretation (as this is commonly understood). The syntactic engine itself – the autonomous principles of composition and manipulation Chomsky now labels "the computational system" – has begun to fade into the background. Syntax reduces to a simple description of how constituents drawn from the lexicon can be combined and how movement is possible (i.e. how something other than the simple combination of independent constituents is possible). The computational system, this simple system of composition, is constrained by a small set of economy principles, which Chomsky claims enforce the general requirement, "do the most economical things to create structures that pass the interface conditions (converge at the interfaces)."

The end of syntax has no immediate consequences for the majority of syntacticians, since most of us have been investigating the interfaces whether we acknowledge this or not. After all, word order is phonology and we have always investigated "sentences" (strings or structures of terminal nodes) under particular interpretations, i.e. with particular assumptions about their LF interface. Chomsky's vision of the end of syntax should have the positive consequence of forcing syntacticians to renew their interface credentials by paying serious attention to the relevant work in phonology and semantics. We should not interpret the diminished role of the computational system within the MP grammar as somehow an abandonment of a previously "autonomous" syntax. The question of the autonomy of syntax has had different content at different times, but whatever the meaning of "autonomous," syntax in the MP is as autonomous, or non-autonomous, as it ever was. As always, syntax – here the computational system – stands between the interfaces and is neither a phonological nor a semantic component. And, as always, syntax trades in representations that are themselves neither phonological nor semantic. A vision of the end of syntax – the end of the sub-field of linguistics that takes the computational system, between the interfaces, as its primary object of study – this vision encompasses the completion rather than the disappearance of syntax.

Related Material in Other Chapters

Several other chapters contain information relevant and related to the material discussed in the present chapter. Chapter 1 deals with the assumptions about phrase structure and Case in standard GB theory. It also discusses the Projection Principle which is intimately tied to the concept of a derivation involving mappings between levels of representation. The chapters on LF, the Binding Theory, and the ECP contain much valuable data analyzed within pre-minimalist versions of the Principles and Parameters approach that at the time of writing still awaits treatment within the Minimalist Program (MP). Many binding-theoretic facts are particularly relevant in this respect because parts of the Binding Theory (e.g. the licensing of parasitic gaps) have traditionally been thought to involve S-Structure conditions, i.e. conditions that cannot be incorporated into the MP in their original form. The material contained in the chapter on the ECP is important as well in that the MP discards the notion of head-government that was central to the definition of proper government in Chomsky (1981). It remains to be shown

that all the classical ECP cases analyzed in terms of head-government can be reduced in an equally satisfactory manner to the principles of the MP, in particular the "Shortest Move" constraint.

The concerns of the present chapter also overlap with several topics covered in chapter 6 (Morphosyntax), including the treatment of agreement, clitics, and phenomena such as passive, applicatives, etc. That chapter also contains a detailed discussion of the relative merits of various versions of the lexicalist hypothesis which determines the degree to which syntactic word formation is possible.

Chapter 8 by Noam Chomsky develops the Minimalist Program further.

8

Bare Phrase Structure

Noam Chomsky
MIT

Contents