Incorporation

A Theory of

Grammatical

Function Changing

Mark C. Baker

The University of Chicago Press
Chicago and London

2 Incorporation Theory

Having defined and motivated the domain of inquiry and having stated the leading idea, I now embed the study more explicitly in a theoretical framework where the thesis can be developed and applied concretely. The framework I adopt is the government-binding theory (GB) of Chomsky (1981, 1982, 1986a) and others. The discussion has two parts. First, I provide a general overview of this framework, describing (in part) its goals and methods as well as some of its specific concepts and principles. This allows "incorporation theory" to be properly located within the more general study of language. While it is impossible to be exhaustive in a few pages, I hope this overview will help those who have less familiarity with the system to understand the kind of project I am engaged in and why, as well as the technical terminology I use. At the same time, the discussion should aid those who are more familiar with the system by clarifying the particular form of the principles I assume in those (frequent) cases where well-known alternatives exist. The second part of the chapter begins to apply the concepts of the framework to incorporation processes specifically, thereby deriving some general results that are valid for each subtype of incorporation considered in chapters 3–6. Here the major technical/theoretical themes of this study are articulated.

2.1 The General Framework

2.1.1 Goals and Methods

The ultimate goal of linguistic inquiry in this particular tradition is to gain an understanding of exactly what someone knows by virtue of which we say that he knows a particular language—e.g., English or Japanese or Basque. As such, it is a subpart of human psychology and (perhaps) of biology, inasmuch as this knowledge of language is a subpart of the structure of the human mind, which in turn is (perhaps) dependent on the structure of the

human brain. Two particular subgoals can be distinguished; we seek theories of:

(I) the knowledge which a linguistically mature person has that (among other things) underlies his use of language;

(II) how that knowledge comes to be in the mature person;

where "knowledge" can presumably be interpreted as "cognitive structures." An important subpart of (II) is to have a theory of:

(II′) the knowledge by virtue of which the person can develop (I),

where "knowledge" again means something like "cognitive structures." (I) is called (particular) grammar (say, of language X); (II′) is called universal grammar (UG). The word "universal" is used here because UG is a common cognitive basis for all human languages and hence the source of nontrivial and nonaccidental similarities in their structure and properties. (II′) is the main focus of this work, although a good deal is said about (I) and (II) as well.

These basic goals of inquiry can be distinguished from other fundamental goals which motivate many who study or have studied language. Thus, the ultimate goal is not to understand the properties of any particular language (English, Japanese, Basque, . . .) in their entirety, where "language" this time is understood as a set of sentences or utterances apart from the human mind. It is not even to understand how they change over time, or what the properties of all such "languages" taken as a set are. Second, the ultimate goal is not to account for the full range of overt behaviors of humans which are somehow classed as "linguistic" behaviors. Finally, the goal is not to understand the particular communicative functions that characterize how and why humans make use of language and particular language structures. The orientation is patently mentalistic, and thus the focus is not on utterances, behaviors, or even purposes, even though any and all of these may give important evidence pointing toward the sometimes distant goal.

A natural question arises here: namely, WHY focus on (I) and (II), rather than on other goals, such as those briefly mentioned? It is not my purpose to justify this fully, and any answer is partly personal. Two types of reason can be mentioned, however. First, it is reasonable to think that (I) and (II) might to a large degree be logically prior to the alternatives mentioned. Thus, it is hard to see how English (or Japanese, or Basque) as a set of sentences can even be defined—much less understood—except as a way in which people with certain cognitive structures respond to in a particular way. If this is so, then the emphasis naturally shifts to the cognitive ded

tures. Similarly, overt linguistic behaviors are almost certainly just those behaviors that involve using linguistic knowledge in some way. Thus, the study of linguistic behavior again presupposes the cognitive structures (I) and (II′); moreover, it adds to them many independent and very difficult factors, such as the speaker's (or hearer's) desires and intentions, to say nothing of his or her muscular and perceptual fitness. Finally, in order to understand the functions and uses of something—in this case language, one frequently must have a good understanding of its structure and what it is. Functionally oriented syntactitians often emphasize the converse: that one often better understands the structure of something if one understands the function (e.g., that of a hammer, which is to drive nails). This is true in part, but if, as seems likely, the actual structure of human language is underdetermined by its communicative functions (as the shape and composition of a hammer is underdetermined by its function of driving nails!), then the functionally oriented goal will not give fruitful understanding apart from the formal one, at least in these domains. To the degree that the mentalistic orientation of (I) and (II) is logically prior to these other goals and orientations, progress will be made more quickly by focusing on it in particular.

The second type of reason for investigating (I) and (II) is simply that they are interesting and worthy of study, whatever other interesting research questions may be available. For one thing, this orientation allows the researcher to abstract away from what is accidental in behavior, culture, or language history to the extent that these are not represented in the average speaker's mind. Only in this way do some true patterns and true puzzles about human language come clear, such as those sketched in section 1.1. Moreover, it is from this perspective that linguistics becomes a vehicle for studying nontrivial properties of the human mind, a topic which I take to be of inherent interest to humans.

As often noted by Chomsky, when one pursues both (I) and (II) a creative tension arises, in particular with respect to (II′). On the one hand, it seems there must be great diversity of linguistic knowledge, since there is great superficial diversity of human languages. For example, some languages are highly "polysynthetic" with many causatives and noun incorporations, some have only a few, some none at all. Indeed, diversity seems so great that some have said (what cannot be true literally), "Languages differ from each other without limit and in unpredictable ways" (Joos (1957, 80)). The initial cognitive structures of the human (II′) must be consistent with all this diversity. On the other hand, this diverse knowledge must develop in each speaker with only limited amounts of evidence. For example, human implausible to assume that each speaker of a polysynthetic language

will have enough training and experience to allow her or him to discover inductively the complex restrictions on forming relative clauses based on double causative constructions, or how those restrictions differ from those on forming questions under similar circumstances (cf. 4.4.2 below). The relevant sentences are heard rarely or never, parents and elementary school teachers are quite unaware of the patterns, and the key issues never arise in simpler structures. Nevertheless, the mature speaker knows (implicitly) the precise restrictions. That information which is present in the final state of the person's mind but is not present in the person's experience is assumed to have its source in the initial state (II′). This points to initial cognitive structures that have rich, complex structure, placing very specific requirements on linguistic forms. Discovering the initial knowledge which is both broad enough to be consistent with the diversity of linguistic knowledge and specific enough to account for the complexity of linguistic knowledge is the fundamental challenge of the research program.

This situation suggests some things about what UG must be like. It must be highly developed and organized, with very restrictive principles which together determine most of linguistic structure, including almost all its subtleties. Moreover, these principles must allow for some specific parameters of variation. The values of these parameters—and hence the exact form of the principles—are determined by the language learner on the basis of (presumably) rather simple exposure to the language. However, when principles differ in two different languages because of this parameterization, the resulting differences in linguistic structures can be far from simple. This is because the allowed structures of the language are those which are consistent with all the principles of that language, and these principles may interact in complicated ways. The result is that a change in one principle may have complex and wide-ranging effects on the set of structures allowed. All the differences between languages thus need not be learned individually. With a universal grammar of this type, it is in principle possible to account for both the surface diversity among languages and the fact that each individual language can be learned in all its complexity. A cognitive system which is the result of setting each of the parameters provided by universal grammar in a particular way constitutes what Chomsky (1981) calls a CORE GRAMMAR, and this core grammar is the heart of a theory of (I), particular grammar. In addition, (I) will generally have some layers (also with internal structure) added to the core grammar as a result of historical residue, contact with other languages, dialect mixture, and the like. These additions constitute the PERIPHERY, and, unlike much of the core, they must be explicitly learned from positive evidence.

Considerations like these have led to a shift in focus in the extended

standard theory/government-binding theory away from specific and explicit syntactic rules. While such rules may in some cases be observationally adequate with respect to goal (I), they are generally deficient with respect to goal (II), since the more complex the phenomenon studied, the more complex the rule, and there is no account of how the language learner could acquire these complexities. Instead, the focus is more and more on the discovery of general principles and constraints, each of which in part determines the nature of a wide variety of different processes. To give a few examples, Ross (1967) observed that many transformational processes such as question movement, relativization, and topicalization seem to be subject to identical conditions (his "island" conditions) and proposed that these conditions be factored out of the statement of the transformational rules themselves and studied in their own right. Chomsky (1977) illustrates a further move: he claimed that those processes, as well as comparative formation, complex adjectival constructions, and others are in fact not independent transformational rules at all (in English), but rather specific instances of a general transformation "Move-*wh*" (to Comp). Residual differences between these constructions are the result of independent conditions which have different consequences in different circumstances, and not of inherent differences in the rule(s) that form the constructions. In another domain, Chomsky (1981) and Stowell (1981) show that explicit phrase structure rules of the familiar type seen in Chomsky (1965) are nearly or completely redundant and should be eliminated from the grammar, all their information being present already in the specifications of the subcategorization/selection properties of individual lexical items and in very general parameterized constraints of universal grammar (the X-bar Convention, case theory, Theta Role Assignment). Thus, while the generalizations about word order and phrasal groupings traditionally captured by phrase structure rules are true, the phrase structure rules themselves appear to be no more than side effects of deeper principles. In this last example, the shift in perspective reaches its natural limit, and the entire burden of explanation falls on the interplay of general conditions which are plausibly principles of UG, rather than on the existence of explicit rules.

Here, we see how my proposals about the nature of GF changing and incorporation phenomena described in the last chapter fit into this general shift of perspective. The claim about GF changing processes is parallel to the Chomsky/Stowell claim about phrase structure: they are all simply side effects of the movement of words, as it is restricted by the principles of universal grammar. Parameters in these principles, moreover, account for the observed diversity across languages.

One's destination and the purpose of one's journey always affects one's path and even one's style of walking. In the same way, the research goals and values I have sketched have consequences for how data are used and how arguments are constructed. Since the focus is on mental structures rather than on particular "languages" or behavior types, we are not necessarily accountable for explaining every utterance, possible or actual, in every or any language—or at least we are not accountable for all utterances to an equal degree. Rather, we focus on those linguistic behaviors which for some reason are most likely to reveal the mental structures in their true light. The situation can be likened to the physicist who tries to determine the force of gravity. The force of gravity is a crucial factor in all instances of falling, and the desire to understand instances of falling is part of the physicist's motivation in studying this topic. Nevertheless, the careful investigation of a leaf falling to the ground on a breezy autumn day will not necessarily help the physicist in his or her inquiry; it may even be a serious setback to his or her understanding. Rather, he or she is better off starting with a steel ball bearing falling in a vacuum tube, even though such events are much rarer and have less impact on daily life than the autumn leaf events. Hopefully, this does not imply that the physicist is unrealistic or unresponsive to new or opposing evidence. The position of a researcher involved in this linguistic program is much the same, and unfortunately there is every indication that much of the linguistic behavior we have record of is like the autumn leaf—complicated by many other external factors. Thus, we not only may but must ignore some of what seems accidental about language as a result of historical factors, language contact factors, or chance idiosyncracies—the "periphery" of grammar—in order to make progress possible. This is a very different thing from being ignorant of, dishonest with, careless about, or insensitive to the data (although human fallibility may introduce these factors as well); it is a fact about the discipline even at its best. I do not claim to have the wisdom to reliably discern which linguistic behaviors are like autumn leaves and which are like steel ball bearings, but together we must attend to the guidelines of our conceptual arguments, follow our hunches, and get started. There is no other way.

Let me include a note on my own conceptual arguments and hunches. It has been observed (e.g. Chomsky (1981, 6)) that there are two valid ways to study universal grammar. One is to study one or two languages in great detail. Whatever of these details are too complex or subtle to have been learned from positive evidence can then be attributed in some form to UG. The other way is to study a great number of languages more superficially, and see what they all share. Whichever of these properties cannot be at-

tributed to general factors then can be incorporated into the theory of UG. These approaches are complementary, and their strengths and weaknesses balance each other. If one looks at only one language, an attractive line of investigation may lead one astray, when the structure of a second language would immediately have shown that line to be hopeless. On the other hand, if one looks at many languages, one can be bewildered by the variety and miss important details which hint at how processes are interrelated. In this study, I try to cross these perspectives by looking at an intermediate number of languages from different families in an intermediate amount of detail. In particular, I draw to varying degrees from the Bantu languages of Africa, the Iroquoian and the Eskimo languages of North America, and the Germanic and Romance languages of Europe. Where there is relevant literature on a particular topic and enough available information to evaluate it somewhat, other languages are discussed, including some from Oceania, the American Southwest, and India. This approach has the strengths of both the extreme approaches, although to a lesser degree, and likewise for the weaknesses. When a particular phenomenon has rather different properties in different languages of the sample for no clear reason, I generally put it aside, tentatively assuming it to be peripheral. No doubt much of interest has been lost here. However, when a particular phenomenon has strikingly similar properties in different languages, and moreover when those properties are subtle and complex, these are the phenomena which I fasten on with enthusiasm, figuring that a reason must be somewhere in the structure of core grammar. In these particular circumstances I sometimes assume something to be universal based on a far from exhaustive language sample; these hypotheses will with little doubt stand in need of refinement or modification given more evidence. Finally, when studying a phenomenon of this type, I sometimes use evidence from one language to establish conclusions which are then used in another, unrelated language. This type of reasoning, perhaps unnatural to some, is valid if one accepts the existence of UG in the above sense; then all humans share common cognitive structures, and these structures determine the shape of all their particular languages. Clearly, with the current limited understanding of what can reliably be attributed to UG, it is not wise to overuse this methodology, and care is needed. Nevertheless, in its place it is extremely productive; it enables a rather complete theory of a phenomenon to be developed, even when the evidence from any single language is only fragmentary, either because not enough is known about the language (by me) to build the whole case, or because no language happens to have the full range of "test constructions" that would be relevant.

As already described, we assume that the diversity and complexity of

languages come from a grammar in which the principles are simple, even elegant, but in which there are a number of them, so that their consequences interact in complex ways. If this is true, our theories, to the extent that they are accurate, will mirror this structure: postulates and assumptions will be rather simple, but many of them may be relevant to the analysis of any one structure. Thus we expect explanations of particular phenomena to consist of simple assumptions and relatively long and complex chains of reasoning, rather than of long and complex assumptions and simple chains of reasoning. When realized appropriately, this should be a sign of intellectual health and some depth of understanding. Hence, when the analysis of a single structure seems to last for pages, I hope the reader will interpret this charitably, rather than as needless tedium.

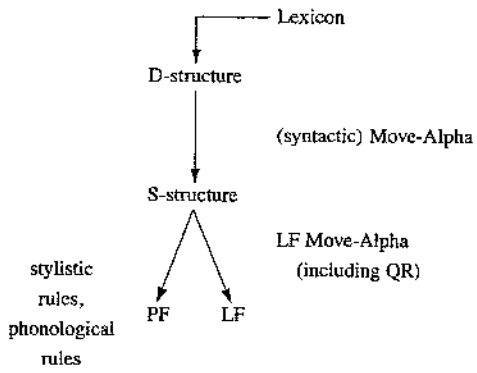
Finally, some importance is given to simplicity, elegance, symmetry, and nonredundancy in this research program. This is not an excuse to oversimplify the truths of the matter at hand, but rather an implicit belief about where those truths are to be found—a belief that needs always to be supported by other considerations before it can be accepted with finality. In part, this is simply the idea that when two similar effects are observed, it is worth the effort to look for a single common cause. There is no guarantee that there will in fact be one, just as there was no guarantee to physicists that the force of gravity would be expressible as a simple linear equation; the autumn leaf falling in the breeze might have been an accurate representation of gravity after all. Yet if there are single causes of multiple effects and simple basic principles, one will certainly never find them without looking for them; while if there are not, the search for such causes should prove fruitless in the final analysis. Thus, looking for these causes is the way to converge on the truth. It is in this spirit as well that I investigate the common properties of GF changing processes and processes of complex predicate formation.

2.1.2 The System of Levels and Rules

Given the conceptual background set as above, a language (or a GRAMMAR) is a mentally represented system for pairing sound and meaning in an unbounded way. We have seen that a large part of this system must be a universal grammar with a highly articulated structure and very specific principles, together with some parameters of variation. In this section and the following one, I describe aspects of a particular theory of this system which has developed to meet these specifications. This will provide a reference point for concepts to be used in our investigation of incorporation.

Government-binding theory typically includes the following levels of representation and processes relating them:

(1)



Formally, a representation at each of these levels (except the lexicon and perhaps phonological form) is a PHRASE MARKER that expresses relationships of constituency, normally represented as a tree or a labeled bracketing. LOGICAL FORM (LF) is the system's link with meaning; it is the level of interface between the language faculty and the conceptual faculties of the brain. Here predication relationships and the scope of quantifiers and operators of various kinds are explicitly represented, as well as the basic thematic relations among items. PHONOLOGICAL FORM (PF) is the system's link with acoustic form; it is the level of interface between the language faculty and the perceptual and motor faculties. Here the phonological shapes and groupings of items are directly represented. The rest of the system is a recursive, formal device which both PF and LF are linked to; in this way the nature of language as an unbounded system of pairing form and meaning is captured. The LEXICON provides the basic elements for the recursion; it (at least) lists the idiosyncratic properties of lexical items which constitute the atomic units of the syntax. In particular, the lexicon specifies what thematic relations these items may enter into with other phrases (i.e., what they subcategorize and assign theta roles to, what theta roles they may receive, etc.). Lexical items are combined together at D-STRUCTURE ("deep" or underlying structure), which is a formal syntactic level of representation at which the thematic relations among items and phrases are directly represented. Finally, D-structure is mapped into S-STRUCTURE. S-structure is a level which is not directly interpreted, but which must be properly related to each of the other three structures simultaneously, thereby tying together form, meaning, and the formal constraints inherent to language. The factoring of the formal syntax into two distinct parts, D-structure and S-structure, is not necessary a priori, and is controversial both outside and within the GB framework. One of the important theoretical implications of the current work is that it will strongly support this traditional distinction (see 8.1).

These levels of structure are related to one another in various ways. S-structure is derived from D-structure by successive applications of the generalized movement transformation MOVE-ALPHA, where "alpha" equals some category. The range of what "alpha" can be varies somewhat from language to language; this is the first locus of parameterization allowed by the theory of UG. For example, "alpha" includes [+*wh*] phrases (i.e., question phrases) in English, since these phrases move to the front of a clause which contains them, as in (2c):

(2) a. I think [Mary bought a novel].
b. I wonder [Mary bought what].
c. I wonder [what Mary bought —].

In (2b) the question word *what* appears in the position where it receives a thematic role as object of the lower verb; this is thus a valid D-structure, parallel to the grammatical (2a). (2b) is not an acceptable surface form, however; rather the question word moves, yielding (2c) at S-structure (and thereby at PF). In Chinese and Japanese, however, [+*wh*] phrases are not included in Move-Alpha in the syntax. Thus, the D-structure and S-structure representations of a sentence corresponding to (2c) are essentially identical to each other and equal to (2b):

(3) *Wo xiang-zhidao* [*Lisi mai-le sheme*].
I wonder Lisi bought what
'I wonder what Lisi bought.' (Chinese; Huang (1982))

A basic tenet of this work is that "alpha" includes categories of minimal bar level (i.e., words) as well as categories of maximal bar level (i.e., phrases) in some languages. Thus, "Move-N" is allowed in the syntax of Mohawk (see 1.2 (33) and (34)), although not in the syntax of English.

LF is related to S-structure primarily by QUANTIFIER RULE (QR) and similar processes which are essentially Move-Alpha in a different guise. LF Move-Alpha's effects are invisible because LF is not directly linked to PF, the representation of form. Its function is to place elements in positions which represent their semantic scope directly. Thus, to represent the fact that (3) is interpreted as an indirect question and not as a direct question or as a pure statement, LF Move-Alpha moves the interrogative element *sheme* to the beginning of its clause, yielding the following LF representation:

(4) *Wo xiang-zhidao* [*sheme* [*Lisi mai-le* —]].
I wonder what Lisi bought

This is identical to (2c). There are no major changes between S-structure and LF for the English sentence, and the similarity of the LF representations expresses the fact that the two sentences are semantically equivalent in these respects. Thus Move-Alpha includes [+*wh*] phrases at LF in Chinese, although not at S-structure. Later on, we will have some reason to think that Move-Alpha can include X^0s in LF as well as before S-structure in some languages.

PF is also derived from S-structure. This derivation accounts for processes like simplifying/adjusting hierarchical structure, deleting null elements, contractions, and perhaps stylistic movements. Little about these processes is known in detail, but see Marantz (1984) and Sproat (1985b) for suggestions. For discussion of how morphological "spell-out" rules and phonological rules fit into the grammar, see 2.2.5 and 8.2.

Finally, the syntactic levels of description of a given sentence are only properly related to one another if they jointly satisfy a fundamental principle of GB theory: the Projection Principle. Intuitively, this states that representations at each syntactic level (LF, D- and S-structure) are "projected" from the lexicon in that they all represent the lexical selection properties of items categorially (cf. Chomsky (1981, 29); see 2.2.2 for a more formal statement). To illustrate, the verb *bought* selects for a direct object NP (its theme) as a lexical property; hence the VP it heads must have a direct object NP at D-structure, S-structure, and LF. This principle determines much of how D-structure is constructed from lexical items. Moreover, it follows that *bought* must have an object in the S-structure (2c) as well, even though nothing is there overtly. Hence (2c) is more fully represented as:

(5) I wonder [what$_i$ [Mary bought [$_{NP_i}$ *e*]]].

The Projection Principle thus has the important consequence that categories moved by Move-Alpha often must leave behind phonetically null copies, called TRACES. Similar considerations imply the existence of traces at LF in structures like (4). Taken together, a moved category and the traces it has left behind constitute a more abstract unit called a CHAIN. Elements of a chain are related to one another by a particular type of coindexing, CHAIN COINDEXING (Chomsky (1986b)), which is generally represented with letter subscripts.

This study is primarily concerned with D- and S-structures and the mapping between them. The Projection Principle plays a central role.

2.1.3 The System of Constraints

The basic structure of representations and rule types is only the beginning of the structure of UG on this theory. For the conceptual reasons discussed in 2.1.1, systems of general principles and constraints are at least as crucial to GB. These principles are generally broken down by Chomsky (e.g., (1982)) and others into subsystems. I introduce each in turn.

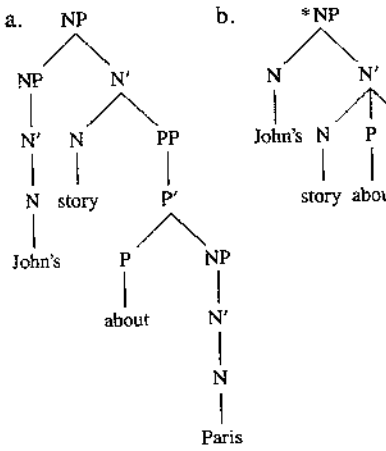
X-BAR THEORY

This subtheory constrains the set of phrase markers allowed; its requirements hold fundamentally at D-structure, where they determine how lexical items are pieced together into phrases. Although the details will not be particularly essential, I assume the X-bar theory of Chomsky (1986b) for concreteness. Basic lexical categories include Noun, Verb, Adjective, and Preposition (more generally Adposition or Particle). Higher level, phrasal categories are projections of these lexical categories, according to the following schemata:

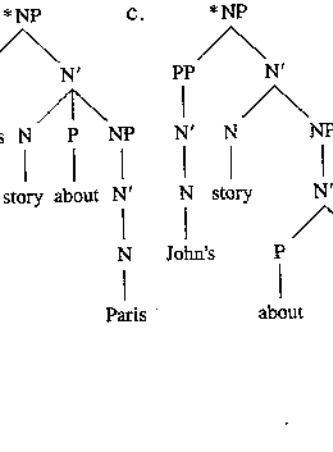
(6) a. $X' = X\ XP^*$
b. $XP = X'\ XP^*$

where "X" ranges over the category types.[1] The asterisk means that any number of the phrases so marked can appear, and order is subject to cross-linguistic variation—a second place where parameterization enters UG. XPs on the right-hand side of (6a) are called COMPLEMENTS; XPs on the right-hand side of (6b) are called SPECIFIERS. X in (6a) is the HEAD (sometimes X′ in (6b) also). (7a) shows a representation valid with respect to X-bar theory; (7b) and (7c) are not valid:

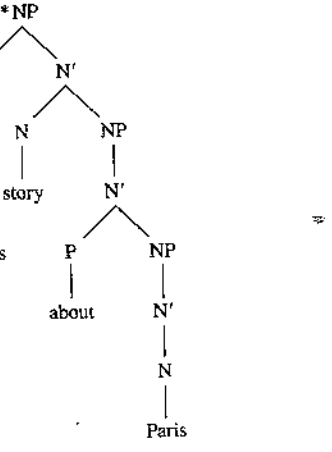
(7) a.



b.



c.

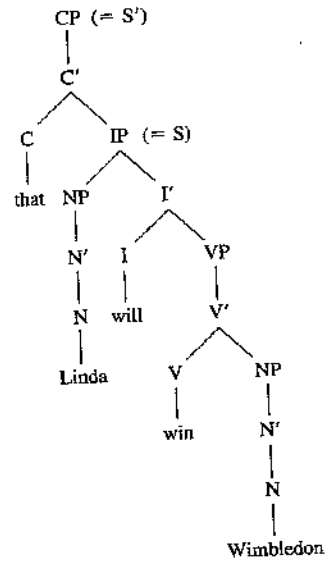


(7b) has nonmaximal phrases serving as complements and specifiers; (7c) has phrases headed by categories of the wrong type. In this way, X-bar theory constrains linguistic structure.

With regard to the structure of clauses, I follow Chomsky (1986b) in

assuming that the nonlexical categories of complementizer (C) and Infl (I) are also heads that form projections in accordance with (6). Thus, S is the maximal projection of I (hence IP); S′ is the maximal projection of C (hence CP). The full structure of a typical clause is thus:

(8)



The subject is thus the specifier of I′. The specifier of C′ is the landing site of *wh*-movement (e.g., the place of *what* in (2b)), which was said to arrive in "Comp" in earlier versions of the theory. I will still use the term "Comp" in this sense when no confusion will arise, since the term is more familiar. Indeed, I will frequently suppress much of the X′ structure for clarity and convenience when possible; (8) will be the last tree drawn in complete detail.

X-bar theory defines the notion MAXIMAL PROJECTION (XP), which is then used to define a fundamental structural relationship of linguistic theory, C-COMMAND (cf. Aoun and Sportiche (1983)):[2]

(9) A C-COMMANDS B iff A does not dominate B and for every maximal projection C, if C dominates A then C dominates B.

This is a formalization of an intuitive relation "higher in the tree than." Thus in (7a) the NP *John's* c-commands the NP *Paris* but not vice versa; the N *story* c-commands them both. This notion is used by other subsystems of grammar.

THETA THEORY

This subtheory is concerned with how semantic/thematic dependencies are represented in grammar. Ultimately, it is this theory that divides the possible semantic dependencies into linguistically significant classes—the theta roles—and characterizes how each theta role is normally represented in linguistic structure. This is not a developed aspect of the theory, however. Theta roles may be assigned by a lexical head to a complement of that head as defined by X-bar theory (e.g. *win* to *Wimbledon* in (8)), or they may be assigned compositionally by the head and its complements to the nearby subject position (specifier of I′ or specifier of N′; e.g. *win* with *Wimbledon* to *Linda* in (8)). The former type are called INTERNAL THETA ROLES, the latter EXTERNAL THETA ROLES (cf. Williams (1981b)).

When a lexical head takes more than one internal argument, I assume that the linear ordering among those arguments is in principle free; where it is not free, the restrictions follow from independent considerations, notably case theory requirements (Stowell (1981), Chomsky (1981)). I assume that the class of theta roles includes at least Agent, Patient/Theme, Goal, Instrument, Benefactive, Location, Direction, and Possessor in something like their usual senses (cf. Fillmore (1968), Gruber (1965), Jackendoff (1972)). Furthermore, I assume that (at least at D-structure) in all languages agent theta roles are external, and patient/theme theta roles are internal when a verb has both, although this is controversial (cf. Marantz (1984); see 8.1). Following Stowell (1981), I represent the theta assignment relation between two items by a second type of coindexing, theta indexing (Chomsky (1986b): "head-marking"), represented with arabic numeral subscripts.

The fundamental principle of theta theory is the THETA CRITERION, a biuniqueness condition on theta role assignment. This can be stated as (cf. Chomsky (1982)):

(10) Every term of LF that requires a theta role (each ARGUMENT) is associated with one and only one position to which theta roles are assigned, and each theta role determined by the lexical properties of a head is uniquely associated with one and only one argument.

The following paradigm illustrates the sense of this constraint:

(11) a. I arrived.
b. *I arrived a dog.
c. I hit a dog.
d. ?*I hit. (meaning 'I hit something')
e. *I hit a dog a cat. (meaning 'I hit a dog and a cat')
f. *I hit. (meaning 'I hit myself')

Arrived has one theta role as a lexical property; *hit* has two. Only in (11a) and (11c) do these verbs have the right number of arguments. In (11b) an argument (*dog*) is not associated with a theta position; in (11f) an argument (*I*) is associated with two theta positions. In (11d) a theta role (the patient) is not associated with any argument; in (11e) it is associated with two. The proper grammaticality distinctions are thus made by the Theta Criterion.

In the statement of (10), theta roles are assumed to be assigned from a specified position to a specified position (the THETA POSITION); both arguments and theta assigners "are associated" with these key positions either by actually occupying them, or—given the existence of Move-Alpha—by being the antecedent of a trace that occupies them. In other words, the Theta Criterion holds of whole chains, and not just of individual items. A term that does not need or receive a theta role in the sense of (10) is called an ADJUNCT.

PREDICATION THEORY

This subtheory is partly related to theta theory. Its fundamental principle is that "predicates" must be associated with a maximal projection (usually called the predicate's "subject," a somewhat different sense of the term from that used so far). A predicate in the sense of this condition can be any maximal projection which does not itself receive a theta role (cf. Williams (1980), Rothstein (1983)). To count as associated, the predicate and its subject must c-command each other. In addition, there may be (parameterized) constraints on the relative linear order between subject and predicate, supplementing the constraints that come from X-bar theory.

VP never gets a theta role in the normal sense of the term; hence it is always a predicate. The predication condition thus implies as a special case that clauses must always have subjects (the "extended" part of the "Extended Projection Principle" of Chomsky (1981)). The following paradigm illustrates this point:

(12) a. John (generally) believes Mary/that justice will prevail.
b. *[$_S$[$_I$is (generally) [$_{VP}$believed Mary/that justice will prevail]]].
c. [$_S$Mary$_i$ [$_I$is (generally) [$_{VP}$believed t_i]]].
d. [$_S$It [$_I$is (generally) [$_{VP}$believed that justice will prevail]]].

The thematic subject present in an active clause (12a) is systematically absent in a passive. Nevertheless, the VP must be predicated of something given the Predication Condition; hence (12b) is ungrammatical as it stands. The requirement of a subject can be met either by moving an NP into the needed position (12c), or by inserting a "dummy," pleonastic NP (12d). This last example shows that the predication condition is not purely semantic, but rather a grammaticalization of an intuitive semantic relationship.

Most languages have a word order principle which requires that a subject precede its syntactic predicate (Rothstein (1983), Travis (1984)). This principle will play a role in the account of crosslinguistic causativization possibilities (cf. 4.3.3).

GOVERNMENT THEORY

This subtheory defines GOVERNMENT, a locality relation holding between two items:

(13) A GOVERNS B iff A c-commands B and there is no category C such that C is a barrier between A and B (cf. Chomsky 1986b).

Intuitively A must be "higher" in the tree than B, and so close that no category of the wrong type—no barrier—contains B but not A. The precise notion of BARRIER is a very technical matter, and I therefore postpone completing the definition until section 2.2.3; for the time being, one can simply assume that all and only maximal projections except S (if S = IP) are "barriers" (cf. Aoun and Sportiche (1983)). Thus, referring to the phrase structure in (7a), *story* governs *John's* but not *Paris* because the PP is a barrier; in (8), *win* governs *Wimbledon* but not *Linda* since it does not c-command it. I assume (without argument, but see 8.1) that this last case is typical: that at D-structure all languages contain a VP node which is a maximal projection, so that the V will fail to c-command and govern the subject (specifier of I') of its clause—although various things can happen in the course of the derivation to change this state of affairs. The government relation is central to GB theory, since it is a key condition in many of its principles.

One important example of such a principle can be introduced immediately: the EMPTY CATEGORY PRINCIPLE (ECP), a condition on the traces left by Move-Alpha that must be satisfied at the level of LF:

(14) a. Traces must be PROPERLY GOVERNED.
b. A PROPERLY GOVERNS B iff A governs B, and A and B are coindexed.

where "coindexed" in (14b) seems to be able to mean either theta coindexing or the chain coindexing induced by Move-Alpha (cf. Chomsky (1981), Stowell (1981), Kayne (1983)). A classic paradigm illustrating the ECP is in (15) and (16):

(15) a. Who$_i$ [$_S$ t_i [$_{VP}$fixed the car]]?
b. What did [$_S$Angelo [$_{VP}$fix$_1$ t_1]]?
c. How$_i$ did [$_S$Angelo [$_{VP}$fix the car] t_i]?

(16) a. *Who do [you wonder [$_{S'}$whether [t [$_{VP}$fixed the car]]]]
b. ?What do [you wonder [$_{S'}$whether [Angelo [$_{VP}$fixed$_1$ t_1]]]]
c. *How do [you wonder [$_{S'}$whether [Angelo [$_{VP}$fixed the car] t]]]

"Short" question movement (i.e., movement to the nearest Comp) is equally possible for all kinds of phrases ((15)), but "long" movement (movement directly to a higher Comp) is far more acceptable for object phrases than for either subject or adjunct phrases ((16b) vs. (16a,c)). These distinctions are made by the ECP. The object NP is governed by a verb which is theta-coindexed with it; hence the ECP is satisfied regardless of where the antecedent of the trace is, accounting for the relative goodness of (15b), (16b). The subject NP and the adjunct NP are not governed by a theta-coindexed element, however; the subject because the verb fails to c-command it, the adjunct because it is not theta-coindexed at all, having no theta role. Hence, these categories can only move such that their antecedent will govern the trace left behind, satisfying the ECP by chain coindexing instead. This condition is met in (15a,c), assuming that S is not a barrier to government here; it is not satisfied if the *wh*-phrase moves any farther, however, as in (16a,c)—S′ at least will be a barrier in these cases. The ECP will play a crucial role in what follows, explaining a similar difference between object movement and subject/adjunct movement in incorporation structures.

CASE THEORY

This subtheory involves the assignment of (abstract) Case to categories and the distribution of NPs which this induces. Certain lexical items—notably transitive verbs, prepositions, and tensed Infls—are lexically specified as being Case assigners. They assign their Case to a category (usually an NP) provided that they govern that category. For example, in English Infl governs the subject NP and is a nominative Case assigner; V governs the object and is an accusative Case assigner. It follows that subjects are nominative and objects are accusative and not vice versa:

(17) a. That he would strike her (surprises me greatly).
b. *That him would strike she (surprises me greatly).

The difference between nominative and accusative shows up morphologically only in pronouns in English but is assumed to be true of all noun phrases abstractly. Case comes in various subvarieties: structural (also called grammatical), inherent, and semantic. Which categories assign what types of Case under what conditions (e.g. linear adjacency, left- or rightward) is a very important source of parameteric variation, as we shall see (cf. Kayne (1983), Stowell (1981), Chomsky (1986a)).

It is usually necessary for an NP to receive Case from a Case assigner

(the CASE FILTER of Rouveret and Vergnaud (1980), Chomsky (1980)), because of the following VISIBILITY CONDITION on LF (Chomsky (1986a), following Aoun):

(18) An NP position which is the head of a chain (i.e. the last position of a moved category) can only bear a theta-index if it receives Case.

Since an NP will normally need to be theta-indexed given the Theta Criterion, it must also get Case. (18) explains contrasts like the following:

(19) a. *Him (he, his) to strike her (would surprise me).
b. For him to strike her (would surprise me).

The infinitival Infl *to*, unlike the tensed I in (17), is not a Case assigner in English; hence the subject NP cannot receive Case in (19a), violating (18). (19a) can be salvaged by a marked process of English in which the prepositional complemetizer *for* is included. This governs the subject (cf. (15a)) and assigns accusative Case, yielding the grammatical (19b).

It has been suggested that the Visibility Condition be extended in various ways. For example, it seems that subjects of predicates must receive Case, even when they are expletive and need no theta-index (*It to rain would annoy me). The notions of this subtheory will also be crucial for the analyses that follow. In chapter 3, I propose that (18) needs to be generalized somewhat, by extending the notion of what counts as "visibility" to include agreement systems and incorporation as well as Case assignment in the narrow sense.

BOUNDING THEORY

This subtheory contains locality conditions; in particular, the SUBJACENCY CONDITION that limits how far Move-Alpha can take a category in one step (Chomsky (1973)). In essence, Subjacency states that a phrase cannot be moved out of more than one category of a certain type (a bounding category). Since the details will not be particularly crucial, I will assume a somewhat dated version of Subjacency, in which the bounding nodes are a stipulated subset of the set {NP, S, S′} (Chomsky (1977), Rizzi (1982)). Subjacency accounts for contrasts like the following:

(20) a. Who do [$_S$you believe [$_{S'}$$t$ [$_S$I said [$_{S'}$$t$ [$_S$I saw t]]]]]?
b. *Who do [$_S$you believe [$_{NP}$my statement [$_{S'}$$t$ that [$_S$I saw t]]]]?
(21) a. What do [$_S$you know [$_{S'}$$t$ that [$_S$he gave t to his father]]]?
b. *What$_i$ do [$_S$you know [$_{S'}$to whom$_j$ [$_S$he gave t_i t_j]]]?

Assume that NP and S are the bounding nodes in English. *Wh*-movement can move a phrase into any specifier of C (i.e. Comp) as long as the position is not filled. Then in (20a) and (21a), the question word can move to

the front step by step, never crossing more than one S node. The result is acceptable, although complex. In (20b), however, the second stage of the movement necessarily goes out of an NP as well as an S. In (21b), the lower specifier of C position is filled by *to whom*, so movement must be in one step and hence out of two S categories. Thus, in both these two examples subjacency is violated and the acceptability of the constructions is degraded.

Exactly what counts as a bounding category is yet another locus of parameterization. Thus, Rizzi (1982) shows that in Italian, sentences like (21b) are grammatical, although those parallel to (20b) are not (see 4.4.2 for examples). One explanation of this is that in Italian the bounding nodes are NP and S′ instead of NP and S, as assumed for English. It is easy to check that with this new definition only one bounding category at a time is crossed in (21b) (and (20a), (21a)), although two must still be crossed in (20b).

Bounding theory will come into focus only in chapter 4, where it is used to provide evidence on the true nature of incorporation structures.

BINDING THEORY

This subtheory is concerned with the relations of anaphors and pronominals to their antecedents. Here the basic notions are the BINDING CONDITIONS, which make the following specifications:

(22) A. Anaphors (e.g. reflexives, reciprocals) must be bound in their governing category.
B. Pronouns must not be bound in their governing category.
C. "Denoting expressions" must not be bound.

where A BINDS B if and only if A c-commands B and A and B are coindexed, coindexing this time being a representation of referential dependency. In these conditions, the GOVERNING CATEGORY is a local domain; it is roughly a category which contains both a subject (in the X′ sense) and an item which governs the element in question. (22) accounts for the following contrasts:

(23) a. Mark thinks that [Sara$_i$ likes herself$_i$].
*Sara$_i$ thinks that [Mark likes herself$_i$].
b. *Mark thinks that [Sara$_i$ likes her$_i$].
Sara$_i$ thinks that [Mark likes her$_i$].
c. *Mark thinks that [she$_i$ likes Sara$_i$].
*She$_i$ thinks that [Mark likes Sara$_i$].

Here we see that anaphors do indeed need an antecedent "nearby" ((23a)), pronouns may have an antecedent as long as it is not "nearby" ((23b)), and

names (one type of "denoting expression") may not have an antecedent at all ((23c)). All this is as (22) states.

An insight of GB is that chain coindexing—i.e. the relationships established by Move-Alpha—is equivalent to referential coindexing with respect to (22). In other words, traces act like overt NPs for binding theory. In particular, the trace of movement to subject position is an anaphor; hence (24) is parallel to (23a):

(24) a. It seems that [the vase$_i$ { broke / was broken } t_i].
b. *The vase$_i$ seems [Pete broke t_i].

The trace of movement to specifier of C position, on the other hand, is a "denoting expression"; (25) is parallel to (23c):

(25) a. *Who$_i$ does Mark think [she$_i$ likes t_i]?
b. *Who$_i$ does she$_i$ think [Mark likes t_i]?

Filling out the paradigm is the empty pronoun PRO, which is found in many languages (e.g. Spanish (*pro*) *vimos a Juan*, '(We) saw Juan'). This element has essentially the same referential properties as English *she* and is subject to (22B).

One more phonetically unrealized element should be mentioned at this point: PRO, the null subject in equi/control structures. Chomsky (1981) assumes that PRO is both a pronoun and an anaphor. (22A) and (22B) then produce contradictory requirements, UNLESS PRO is ungoverned, thereby having no governing category. Since in general the only ungoverned position is the subject of an infinitival clause, PRO will appear only there. Thus:

(26) a. I$_i$ want [PRO$_i$ to meet the ambassador].
b. *I$_i$ want [the ambassador to meet PRO$_i$].
c. *I$_i$ hope [PRO$_i$ will meet the ambassador].

In (26a), PRO is ungoverned; in (26b) it is governed by the verb *meet* and in (26c) by the I *will*, resulting in ungrammaticality. Note that since PRO is always ungoverned it will never receive Case; for some reason it is exempted from the Case Filter (18).

Binding theory also has some parametric variation. The only thing that will be of crucial concern to us is that anaphors in some languages, unlike *himself* in English, require that their antecedent be a subject. Binding theory will not be a central concern here, but it, like bounding theory, will be used at various points to give evidence about the nature of incorporation structures. Some innovations of Chomsky (1986a) will be assumed in 3.2.2.

CONTROL THEORY

In the system outlined, binding theory determines the distribution of PRO but not its possible antecedents. Thus, it is possible that a distinct subtheory, called control theory, is needed for this task (but see Manzini (1983a)). Little is established in this area, and the matter will come up only in passing in chapters 4 and 6.

This completes the review of the major concepts of government-binding theory.

2.1.4 Grammatical Functions in Government and Binding

The reader may have noticed that I have laid out the structure of GB with no direct mention of grammatical functions, even though they should be central to my topic. This is no accident, because GFs have a derivative rather than a fundamental role in this theory. Chomsky generally defines the grammatical functions in terms of phrase structure configurations and the primitives of X-bar theory (Chomsky (1965; 1986b)). Thus, the SUBJECT of a clause is defined as the X' theory specifier of Infl or N (written [NP, S] or [NP, NP]); the (DIRECT) OBJECT of a clause is defined as the (NP) X' theory complement of an X^0 category (also written [NP, VP], [NP, N'], etc.); and so on.

Nevertheless, in relating to GB the literature on GF properties and GF changing written in other linguistic traditions, an important point should be made. For concreteness, focus on the GF "direct object." However linguists may use this term, all agree that (for example) the NP *Linda* in (27) is an object:

(27) Rover bit **Linda.**

Now, GB is divided up into modular subtheories; hence, NPs in other structures may form a natural class with this NP with respect to some of the subtheories but not others, depending on the particular concepts important to each. Consider the following range of structures:

(28) a. Rover [$_{VP}$swam **the river**] (after biting Linda).
b. Linda$_i$ [$_{VP}$seems [$_S$**t**$_i$ to have been scarred by the bite]].
c. Linda [$_{VP}$considers [$_S$**Rover** to be dangerous]].
d. Linda and Rover would [$_{VP}$prefer [$_{S'}$(for) **each other** to die]].
e. Linda [$_{VP}$hopes [$_{S'}$that **Rover** will never return]].

Which of the boldfaced NPs is an object of the matrix verb? Intuitively we call an NP an object if it is like the object of (27) in relevant ways. However, whether the ways in which these NPs behave like the object of (27) are relevant or not depends crucially on which subtheory one has in mind. The NP in (28a) is identical to that of (27) with respect to X' theory (and most of the others), but perhaps not with respect to theta theory—it is not if it is linguistically significant that it has a PATH thematic role rather than a patient role (cf. Jackendoff (1983)). The NP in (28b) is not similar to that of (27) with respect to X-bar theory (or theta theory) but is similar with respect to government theory, in that both are governed by the matrix verb. The same holds for the NP in (28c), which is also similar to that of (27) with respect to case theory: both receive (structural) Case from the main verb. The NP in (28d) is not an X' theory sister of the matrix verb, nor a thematic dependent of the verb, nor is it governed or case-marked by the verb. Even so, it still forms a natural class with the NP of (27) with respect to binding theory: both have the entire matrix clause as their governing category, allowing reflexives. Finally, the highlighted NP in (28e) is not grouped with that of (27) by any subtheory; it is a canonical subject. Any of the other classes defined in (28) can be called a class of "objects" in a meaningful sense.

The NP in (28e) does, however, form a natural class with each of the highlighted NPs in (28b–d) with respect to at least one subtheory. For example, they all act like subjects for theta theory, and partially so for binding theory, in that they all create an "opaque" domain (Chomsky (1980)) in which anaphors must and pronouns must not have antecedents. Thus, the notion "subject" is also relative to the particular concerns of the moment.

From these examples, we see that in GB it is very natural to make the traditional GF names into relational terms, which have meaning only given a particular subtheory. Hence, when someone gives evidence that a certain nominal is an object, we must ask which subtheories this evidence is relevant to. Moreover, the framework predicts that NPs will often show hybrid properties, acting as an object with respect to some subtheories and as a subject with respect to others. This will prove to be an important explanatory virtue of this system.[3] In what follows, I use terms like "subject," "object," etc. without specification when the relevant subtheories are clear from the context. Two senses of the GF terms are particularly important: the X-bar sense, since the structure is fundamental to the other theories; and the Government/Case sense, since these determine the surface morphological features which are most obvious to the observer. To distinguish these senses, I sometimes use terms like "structural object" for the former and "NP with (surface) object properties" for the latter.

2.2 TOWARD A FORMAL THEORY OF INCORPORATION

With the basics of the GB theory laid out, it is time to focus on the aspects of the framework that need clarification and refinement so that they can be

applied to Incorporation (defined in 1.2) clearly and contentfully. We do this in the context of exploring in detail the consequences that the grammar has for X^0 movement. Some concepts are applied to basic examples immediately, but the major goals are to develop a theoretical core and to derive tools for use in what follows. Thus, the discussion here is for the most part rather abstract and technical, compared to some of what follows. With this in mind, the reader is invited to skim or to read especially carefully, according to her/his interests.

2.2.1 D-Structure and the Uniformity of Theta Assignment

The first concept to be clarified is that of D-structure. Chomsky (1981, 43f.) characterizes D-structure as "a pure representation of thematically relevant Grammatical Functions (=GF-theta)." This means essentially that at D-structure all phrases appear in the position that the theta-role they receive is assigned to (cf. also the "logico-semantic" structure of Marantz (1984)). As an example, *whose luggage* and *Jerry's luggage* must be in the position marked *x* in the D-structures of (29a) and (29b), because they bear the same theta role as the phrase *Jerry's luggage* in (29c):

(29) a. Whose luggage did the airline [lose *x*]?
b. Jerry's luggage was [lost *x*] by the airline.
c. The airline [lost Jerry's luggage].

There have been several attempts to eliminate D-structure from the grammar as a level with independent status, relying instead on algorithms that form chains defined directly on S-structure (e.g. Chomsky (1981, chapter 6), Rizzi (1983), Sportiche (1983)); nevertheless, there is growing evidence that D-structure in fact exists (see Burzio (1986), Chomsky (1986a), Baker (1985a)). If this latter conclusion is correct, D-structure's character as a linguistic representation of thematic structure must be taken seriously. In this light, I propose to strengthen the notion of D-structure, so that it is a representation of thematic structure more generally. Toward this end, let something like the following be a guiding principle of grammar which characterizes the level of D-structure:

(30) The UNIFORMITY OF THETA ASSIGNMENT HYPOTHESIS (UTAH):
Identical thematic relationships between items are represented by identical structural relationships between those items at the level of D-structure.

This hypothesis clearly includes the idea that D-structure directly represents "GF-theta" as a special case but is somewhat more general. In order to make this fully formal one would need, among other things, a more

exact theory of theta-roles than we now have;[4] I will leave it at a rather intuitive level.

The UTAH can constrain linguistic analyses in meaningful ways. For example, it supports the so-called Unaccusative Hypothesis (Perlmutter (1978), Burzio (1981; 1986)), according to which the sole argument of certain nonagentive intransitive verbs is a structural object at D-structure. This NP then becomes the subject by S-structure via Move-Alpha. Given this analysis, sentences like those in (31) have the D-structures given in (32):

(31) a. Julia melted the ice cream into mush.
b. The ice cream melted into mush.
(32) a. [$_S$Julia [$_{VP}$melted [the ice cream] into mush]].
b. [$_S$*e* [$_{VP}$melted [the ice cream] into mush]].

The D-structures in (32) are those that the UTAH implies: the same thematic relationship holds between *the ice cream* and *melted* in both sentences in (31), and this is represented by their having the same structural relationship at D-structure in (32). In fact, this analysis has been shown to be correct for alternations such as this by much evidence in Italian and other languages (e.g. see references cited above).

On the other hand, the UTAH is not consistent with the analysis of the dative shift construction put forth by Kayne (1983, Chapter 7). On his analysis, the thematic paraphrases in (33) have the nonparallel D-structures in (34):

(33) a. Brian gave a nickel to Sophia.
b. Brian gave Sophia a nickel.
(34) a. Brian [$_{VP}$[$_{V'}$gave a nickel] to Sophia].
b. Brian [$_{VP}$gave [$_S$Sophia a nickel]].

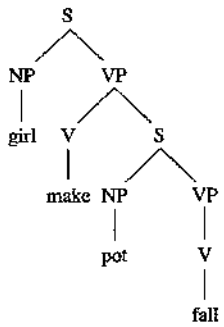
Sophia bears the goal role of the verb in both sentences, yet this is not represented in the same way in (34a) and (34b). Thus, the UTAH can be used to guide the construction of analyses—both by the linguist and by the child—in a nontrivial way.[5]

The UTAH has consequences for GF changing processes as well. Consider again the thematic paraphrases involving causatives in Chichewa (Bantu):

(35) a. *Mtsikana a-na-chit-its-a kuti mtsuko u-gw-e.*
girl do-CAUSE that waterpot **fall**
'The girl made the waterpot fall.'
b. *Mtsikana a-na-gw-ets-a mtsuko.*
girl **fall-CAUSE** waterpot
'The girl made the waterpot fall.'

In both these sentences, *mtsuko* 'waterpot' bears the same thematic relationship (theme) to the verbal root *-gw-* 'fall'; thus the UTAH suggests that these items should stand in the same structural relationship in the D-structures of both. This in turn implies that the verb root is an independent constituent in an embedded clause in the D-structure of (35b), just as it is in the D-structure of (35a):

(36)



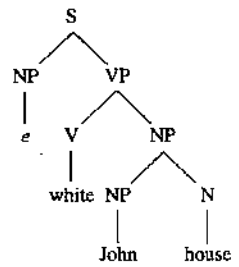
A similar conclusion follows for the Noun Incorporation thematic paraphrases such as the example from Mohawk (Postal (1962)):

(37) a. *Ka-**rakv** ne sawatis hrao-**nuhs**-a?*.
3N-**be.white** John 3M-**house**-SUF
'John's house is white.'

b. *Hrao-**nuhs**-**rakv** ne sawatis*.
3M-**house**-**be.white** John
'John's house is white.'

The nominal *-nuhs-* 'house' bears the same thematic relation to the stative verb *-rakv* 'be white' in both sentences; therefore it must occur in the same D-structure configuration in both. Assuming that *-rakv* is unaccusative, this configuration must be:

(38)



Generally, whenever a part of a word shows syntactic signs of assigning or receiving a thematic role in the same way that morphologically independent constituents do, the UTAH will imply that that part of the word appears in an independent structural position at D-structure, so that the thematic relationship can be represented in the canonical way.[6] Thus, the Uniformity of Theta Assignment Hypothesis points away from a lexical analysis of causative, applicative, and noun incorporation structures and gives theoretical motivation for analyses in terms of syntactic X^0 movement.

2.2.2 S-Structure and the Projection Principle

Given that the UTAH determines certain properties of the D-structure representations of "GF-changed" sentences, the PROJECTION PRINCIPLE (PrPr) should determine properties of their S-structure (and LF) representations. Chomsky (1981, 38) states the PrPr in the following way:

(39) (i) If B and A are immediate constituents of C at level L_i, and C = A′, then A theta marks B in C.

(ii) If A selects B in C as a lexical property, then A selects B in C at level L_i.

(iii) If A selects B in C at level L_i, then A selects B in C at level L_j.

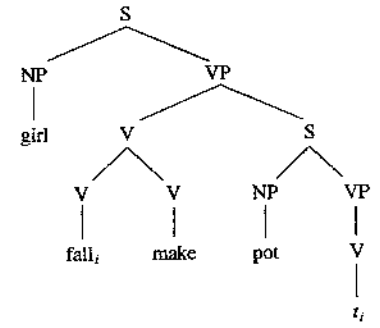
where L ranges over the "syntactic levels" D-structure, S-structure, and LF. Part of the content of this principle (explicit in (iii)) is that transformational processes neither create nor destroy categorial structure that is relevant to the lexical properties of items, including the thematic relationships that they determine.

Unfortunately, there is potential ambiguity as to what type of item the variable "A" refers to in this principle. Let us take a particular example. In (35b) above, the items whose properties must be represented categorially at every level could be taken to be both the root *-gw-* 'fall' and the affix *-ets-* 'make'; otherwise it could be taken as being only the combination of the two *-gw-ets-* whose properties must be so represented. This ambiguity arises as long as all three are assumed to be in the lexicon. If the second interpretation, in which the two morphemes are combined, is taken, (35b) will presumably have the structure of an ordinary transitive sentence at every syntactic level.

The UTAH, however, resolves the ambiguity, implying that this second interpretation is unavailable in some cases, and that the two morphemes must be independent at D-structure. Given this, the PrPr takes over and determines that the lexically determined theta-marking properties of both items must be categorially represented at the other levels as well. Thus, in

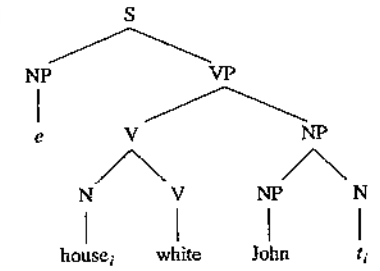
our example, the causative morpheme *-ets-* must take a clausal complement at S-structure (and LF) because it takes one at D-structure. Similarly, *-gw-* must assign an external theta role to a subject position, because it does so as a lexical property and at D-structure. In short, the PrPr implies that X^0 movement must preserve structure by leaving traces, just as XP movement does. Thus, the S-structure of (35b) must not be that of a simple transitive verb, but rather (40), where there is a trace left in the D-structure position of the embedded V:

(40)



By the same token, the S-structure of (37b) must be (41), again with a trace in the base position of the incorporated noun:

(41)



Similar consequences follow for any instance of incorporation where the UTAH requires that two items be separate at D-structure.

In his discussion of (39), Chomsky makes it clear that "B," the theta role receiver, refers to a position rather than a category; due to Move-Alpha, that position can contain either the selected category or its trace. Now we see that a similar remark must be made about "A," the theta role assigner; it too must refer to a position, and the position may be occupied by either the selector or its trace.

Notice that the "surface" structures assigned to sentences like (35b) and

(37b) are different from those assigned by other theories, even those which derive the sentences syntactically (e.g. "Old" Transformational Grammar, Marantz (1984)) due to the presence of the null structure. The obedience to a strong Projection Principle is a distinctive characteristic of my approach.[7]

In closing, I point out the creative tension between the Projection Principle and the Uniformity of Theta Assignment Hypothesis; together they constrain the theory and make it interesting. The Projection Principle says that certain conceivable transformational processes (e.g. Raising to Object; Chomsky (1981)) are ruled out in principle; transformations cannot modify syntactic structure beyond a well-defined point. However, it is possible to escape much of the empirical bite of the Projection Principle by claiming that structures such as causatives and applicatives are simply base-generated, with identical structures throughout the syntax. In the limit, this tactic would move all such grammatical relationships into the lexicon. Explanation of their properties is still necessary at that level, and little is gained. In effect, the Projection Principle is emptied of explanatory content. The UTAH, on the other hand, leads away from base generation in many cases. Yet unless the transformational component is limited by principles like the Projection Principle, it makes little difference what D-structure is assigned to a given form, because anything could happen en route to the interpreted levels of PF and LF. In this case, the UTAH would have little explanatory content. However, in a theory which contains both, each provides a check against the undisciplined avoidance of the other. This is the kind of situation which can give rise to deep and true explanations. Thus, linguistic theory is stronger with both in balance.

2.2.3 Head Movement and the Empty Category Principle

So far, we have developed the notion of D-structure and clarified the Projection Principle such that a single, morphologically complex unit on the surface may be derived by combining constituents which are independent at D-structure for principled reasons. This sets the stage for analyzing linguistic phenomena in terms of Incorporation. The next step is to explore more closely the idea that Incorporation is no more than the syntactic movement of an X^0-level category.

Within the GB framework, saying that something is syntactic movement is not a vague or meaningless claim. Rather MOVEMENT is a technical term; its use implies that incorporation is a subcase of the generalized transformation Move-Alpha—in particular, the subcase where the bar-level feature of alpha is taken to be zero. If this is correct, significant generalizations should be captured by saying that Incorporation is fundamentally the same as more familiar and well-studied instances of Move-Alpha like NP move-

ment in raising, or *wh*-movement in question formation. Based on his study of these latter phenomena, Chomsky (1981, 55 ff.) discovers the following properties of the Move-Alpha relation as it holds between a trace and its c-commanding antecedent:

(42) (i) The trace is (properly) governed.
[i.e. it is subject to the ECP]
(ii) The antecedent of the trace is not in a theta position.
(iii) The antecedent-trace relation satisfies the subjacency condition.

These properties are not necessarily true of other, superficially similar linguistic relationships, such as the construal relation that holds between PRO and its antecedent, as Chomsky shows. Thus, they can be taken as a valid characterization—perhaps in part a definition—of the movement relation. If Incorporation is in fact movement in the technical sense, it should obey these three conditions. (42) will then express the significant generalizations that group together incorporation and other "movement" processes.

Consider first property (42ii). For XP movement, this implies that NPs can never move into an object position, and they can only move into the subject position when the VP assigns no theta role to that position, as in unaccusative verbs and raising verbs. In fact, this property does not need to be stipulated independently; it follows from the Theta Criterion (10), which requires a biunique relationship between theta roles assigned by items and phrases that need theta roles. If an NP moved from a position where a theta role is assigned to another such position, it would be associated with two theta roles, violating of this condition. Following Koopman's (1984) discussion of verb movement, I observe that the movement of theta role assigners must obey the same constraint as the movement of theta role receivers in this regard: if a theta role assigner moved from a position where it assigns a theta role to one argument to a position where it assigns that theta role to another argument, the biuniqueness between theta roles and arguments is again broken. Thus, the notion "theta position" in (42ii) is to be understood (somewhat more broadly than Chomsky intended) as "position from which a theta role is assigned" as well as "position to which a theta role is assigned." In other words, a theta position is any position which is relevant to the establishment of thematic relationships.

A glance at the putative Incorporation structures in (40) and (41) shows that they satisfy this property of movement; the antecedent of the trace is in a position which is (Chomsky) adjoined to a lexical item—surely not in general a position of either theta role assignment or reception. In fact, given that X-bar theory holds at D-structure, adjoined positions will not in general exist at this level, where the set of thematically relevant positions is defined (cf. Jackendoff (1977), Stowell (1981)).

More interesting is the question of whether Incorporation-type X^0 movement must satisfy condition (42i): i.e., whether the trace left by such a movement is subject to the ECP. This will prove to be the heart of the matter, and the discussion will necessarily be rather long and technical.

Intuitively, the ECP is a requirement that the position (and perhaps the content) of a phonetically null trace must have something nearby which locally identifies it, a requirement that can be met by either an item that theta-marks it or the antecedent itself. In fact, there is a strict locality condition on Incorporation that comes to mind in this connection. Travis (1984, 131) gives this condition shape in terms of the following constraint:

(43) HEAD MOVEMENT CONSTRAINT (HMC)
An X^0 may only move into the Y^0 which properly governs it.

Travis bases this generalization on observations about Germanic Verb and Infl movement, together with the ideas on noun incorporation in Baker (1984) (cf. also principle (VII) of Sadock (1985, 413), which is equivalent to (43)). Notice that each of the putative instances of Incorporation introduced so far obeys this condition: a verb, noun, or preposition moves into the verb that governs it. I postpone the task of establishing that this property is true in general; for the time being let us just assume that the HMC is descriptively correct. Note, however, that this is unlikely to be an independent principle of grammar. In particular, it uses the notion "proper government," which is the hallmark of the ECP. I will endeavor to show that the HMC can be derived from the ECP; in fact it is simply the empirical evidence that traces of X^0 movement are subject to this principle, as are all other traces of movement. In order to show this, some particular assumptions are necessary.

Assume that the trace of an X^0, known to exist by the Projection Principle, must be properly governed. This means that it must be governed by an element which is either theta-coindexed with it (i.e. a head) or by an element which is chain-coindexed with it (i.e. an antecedent). Now suppose that X^0-level categories are never theta-marked by an argument taker; only the XP-level categories which they head are. This makes sense from a number of perspectives. Formally, it is more or less implied by the combination of X-bar theory and theta theory: by X-bar theory, only XP-level categories can be sisters of (complements of) a lexical head, and by theta theory, (direct) theta marking takes place only under sisterhood. Thus, XPs are theta-marked and not X^0s. From a semantic viewpoint, this also makes sense. To take a particular example, the linguistic relation of theta marking

between a verb and a nominal is supposed to correspond to a given semantic relationship that holds between the referent of the nominal expression and the action or state type named by the verb.[8] Now it is the category NP which refers, and not the category N. Thus, it is reasonable to say that the V theta marks the NP but not the N. This is illustrated with the following example:

(44) I finally found [[someone] who really cares about me].

Here the point is not that the speaker located anyone in general—the potential referent of the head N on its own—but rather a very particular person, the referent of the NP as a whole, including the restrictive relative. Thus, XPs can be theta-marked but Xs cannot. Formally, this means that theta indexes are assigned to the XP node under sisterhood as above, and that these theta indexes do NOT percolate to the head X^0 of that XP, even though other types of features do percolate.[9]

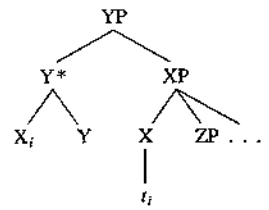
This discussion implies that although a lexical head may govern the trace of an X^0 (cf. Belletti and Rizzi (1981)), it can never properly govern the trace, since the X^0 never bears a theta index. It then follows from the ECP that it must be governed by its antecedent. This consequence can be stated in the following form:

(45) An X^0 must govern its trace. (<=ECP)

Given that X^0 movement must leave a trace, (45) will be virtually equivalent to (43) if we can show the following: an X^0 governs its former position if and only if it appears united with a Y^0 which governs the XP that X headed at D-structure.

For an X^0 (or any category) to govern its trace, two conditions must be met, in accordance with the definition of government sketched in (13). The first is that it must c-command its trace. Here some clarification is useful. Consider an abstract incorporation structure:

(46)



The leading idea of the c-command relation is that the first node of a particular type that dominates the c-commander must also dominate the node to be c-commanded. The crucial question then is exactly what type of category counts for c-command, and here definitions differ. If the zero level node Y* formed by adjunction counts, X will not c-command its trace; if it does not, X will c-command its trace. Clearly, we must assume that Y* does not count in order for Incorporation structures to be possible at all. If we simply continue with the Aoun and Sportiche (1983) definition of c-command given in (9), this result follows immediately, since with this definition only maximal projections are considered. Y* is not a maximal projection; hence it is not a maximal projection that contains X and not its trace. Indeed, there is no such maximal projection, and therefore X does c-command its trace as required. Note that this result is essentially identical to that assumed in analyses of clitics like that of Borer (1983, 35f.), in which a clitic governs a complement position of the head that the clitic is attached to.[10]

The second requirement that must be met in order for an X^0 to govern its trace is the locality requirement proper: there must be no BARRIER category that intervenes between the two. The exact definition of a barrier must be handled with some care. The term "barrier" is introduced by Chomsky (1986b); I will assume a somewhat different technical implementation of the leading ideas of that work. Chomsky develops the insight that what constitutes a barrier to government between two nodes must be made relative to those nodes themselves. Thus, consider the following structures:

(47) a. John decided $[_{S'} e$ $[_S$PRO to $[_{VP}$see the movie]]].
b. John prefers $[_{S'}$ for $[_S$Mary to $[_{VP}$see the movie]]].
c. How did John want $[_{S'} t'$ $[_S$PRO to $[_{VP}$fix the car t]]]?

In (47a), *decide* does not govern the embedded subject position, since PRO can appear in this position. Therefore, either S′ or S (or both) must be a barrier to government here. Nevertheless, S cannot be a barrier to government in (47b), because the complementizer *for* assigns Case to the subject and must therefore govern into the S. Furthermore, S′ cannot be a barrier to government in (47c), because (following Lasnik and Saito (1984)) the *wh*-word *how* must properly govern its trace in Comp across this boundary to satisfy the ECP. Therefore, neither S′ nor S can be an absolute barrier to government; one of them must be a barrier in (47a) RELATIVE TO THE PARTICULAR POSITIONS OF THE ELEMENTS INVOLVED.

In this context, Chomsky considers two distinct notions of what creates a barrier for government, both with roots in the literature. One is that maximal projections of certain kinds block government (cf. Aoun and Sportiche (1983)); Chomsky proposes that in fact it is maximal projections which are not theta-marked arguments that create barriers. The second idea is a "minimality" notion, in which government between two nodes A

and B is blocked if there is another lexical head C which is closer to B than A is (Rouveret and Vergnaud (1980), Reuland (1983)). On this approach, a category which contains such a C as well as B, but does not contain A, is a barrier between A and B. Chomsky claims that in fact both notions are needed. If the Head Movement Constraint is correct and is a consequence of the ECP, we have good evidence that confirms this conclusion. Thus, suppose that both (48b) and (48c) are impossible incorporations, where X, Y, and Z stand for lexical categories, and the links represent the theta marking relationships (see 3.1, 4.1, 5.1):

(48) a. $[_{YP}X_i + Y\ [_{XP}t_i\ ZP]]$

b. $*[_{YP}X_i + Y\ [_{XP}t_i\ ZP]]$

c. $*[_{YP}Z_i + Y\ [_{XP}X\ [_{ZP}t_i]]]$

Here (48b) is ruled out by the first notion of barrierhood, since "XP" is a non–theta-marked category containing the trace but not X. The second notion does not rule it out, however. On the other hand, (48c) is ruled out given the second notion of barrierhood, but not the first: both XP and ZP are theta-marked and hence not barriers in the first sense; but XP does contain the trace and a lexical head but not the antecedent, so it is a barrier in the second sense. Thus both notions seem to be required.

We can now give a definition of "barrier" consistent with these requirements. For Chomsky, barriers are relative only to the potentially governed element; I propose that they be doubly relativized with respect to both the potential governor and the potential governee in the following way:

(49) Let D be the smallest maximal projection containing A. Then C is a BARRIER between A and B if and only if C is a maximal projection that contains B and excludes A, and either:
(i) C is not selected, or
(ii) the head of C is distinct from the head of D and selects some WP equal to or containing B.

Definition (49) has the same basic structure as Chomsky's definitions (1986b, 14, 42), although there are several differences of detail. The core case of government is the relation between a word and its theta-marked complement. (49i) expresses the fact that an adjunct breaks a government path between A and B; if B is contained in an adjunct, we say that it is not THETA-CONNECTED to A. (49ii) is the Minimality Condition, expressing the fact that an intervening theta assigner also breaks a government path; if B is contained in a category with such an item, we might say that it is not DIRECTLY theta-connected to A. Then A governs B if and only if it is "directly theta-connected" to B. This terminology emphasizes the sense in which the government relation is a grammaticalized formal extension of the predicate-argument relation. The phrase "distinct from the head of D" in (49ii) is consistent with the idea of the minimality condition (surely A is not counted as a closer governor than itself), but is redundant so far; this clause will become important in the next subsection.

(49) is a simplification of Chomsky's approach to barriers in that it eliminates the notion of inheritance of barrierhood and several stipulations involving the nonlexical categories complementizer and Infl; the cost of this simplification is the abandoning (at least for purposes of this work) of Chomsky's primary goal—a definition of barrier which will also be appropriate as a definition of "bounding node" for Subjacency. The simplifications come from introducing into the definition the notion of SELECTION. Selection is a natural generalization of the theta role assignment relation to include the nonlexical categories:

(50) A selects B if and only if:
(i) A assigns a theta role to B, or
(ii) A is of category C and B is its IP, or
(iii) A is of category I and B is its VP.

Intuitively, a given item A selects another item B if it needs to occur with (some such) B in order to satisfy its inherent lexical properties. Lexical categories need arguments given the Theta Criterion, and C and I need IP and VP sisters, more or less by definition. Notice that C and I do NOT select their specifiers, Comp and the subject respectively. This selection asymmetry in the case of the nonlexical categories is the reason why S and S′ behave somewhat differently from other categories with respect to government, a fact which has often been noticed.

Before continuing, I illustrate these definitions by showing how they apply to the structures in (47) (repeated here) to give the desired results:

(51) a. John decided $[_{S'}e\ [_{S}$PRO to $[_{VP}$see the movie]]].

b. John prefers $[_{S'}$for $[_{S}$Mary to $[_{VP}$see the movie]]].

c. How did John $[t''$ [want $[_{S'}t'\ [_{S}$PRO to $[_{VP}$fix the car $t]]]]]]$?

Does *decided* govern PRO in (51a)? The answer is no: S′ is a barrier between them (clause (ii)), since it contains PRO and excludes *decided* and its null complementizer head selects the IP which contains PRO. Notice, however, that S′ would not be a barrier if it did not have a null complementizer head. Thus, I assume that the special property of so-called "Exceptional Case Marking" (ECM) verbs like *believe* is that they allow the deletion of the head of their complement. (For an alternative, see note 4 to

chapter 8.) This causes the verb to govern the embedded subject, allowing the Case marking of a lexical NP and barring PRO:

(52) John believes [$_{S'}$—[$_{S}$Mary/*PRO to have seen that movie]].

This account of ECM verbs is attractive in that it leaves the specifier of C position completely intact; this position is in fact needed to house the traces of moved adjunct question phrases; see Lasnik and Saito (1984, 274). Next, does *for* govern *Mary* in (51b)? This time the answer is yes. The only possible barrier is S, but S is selected by *for*, and its head *to* does not select the subject PRO, as discussed above. Hence S is not a barrier by either (49i) or (49ii). Finally, is t' in Comp governed in (51c)? Here I follow Chomsky (1986b) in assuming that successive cyclic movement proceeds by adjoining the *wh*-phrase to VP between one Comp and the next; hence t'' exists as a possible antecedent governor. The only category that includes t' and excludes t'' is S′ this time. S′ is selected by *want* and hence not a barrier by clause (i). The head of S′, another null complementizer, does not select t' in its specifier; hence S′ is not a barrier by clause (ii) either. Therefore, government holds between t'' and t'.[11] Note that if S′ were not selected by *want*, then it would be a barrier between the traces. This accounts for the deviance of (53) compared to (51c):

(53) How did John [t'' [leave [$_{S'}$ t' [$_{S}$PRO to [$_{VP}$fix the car t]]]]]?

where the infinitive is a purposive rather than a complement. t' here fails to be governed by its antecedent, violating the ECP. I conclude that the definitions in (49) and (50) adequately cover the government properties of the clausal system, including some of the more complex examples.

Finally, we are ready to return to the topic of X^0 movement, to see that these definitions give the correct results with respect to the abstract test cases in (48), repeated below. Recall that an X^0 movement structure will only be acceptable if the X^0 governs its trace (see (45)):

(54) a. [$_{YP}X_i$ + Y [$_{XP}t_i$ ZP]]
b. *[$_{YP}X_i$ + Y [$_{XP}t_i$ ZP]]
c. *[$_{YP}Z_i$ + Y [$_{XP}$X [$_{ZP}t_i$]]]

In (54a), the only maximal projection which contains t but not X is XP. However, XP is selected by Y, and its head (t itself) surely does not select any category containing t. Thus XP is not an actual barrier between the two, and X governs t. In (54b), the structural configuration is similar, but this time XP is not selected by Y (or anything else); thus it is an actual barrier (by (49i)). Hence there is no government between X and the trace, and the structure is ungrammatical by the ECP. Finally, in (54c) XP is a potential barrier between Z and t. It is selected, but its head X also selects ZP, which contains the trace. Hence it too is an actual barrier (by (49ii)). The result is that X governs the trace but is not coindexed with it, while Z is coindexed with the trace but does not govern it. Neither suffices as a proper governor for the trace, and (54c) is also ungrammatical by the ECP, as desired. Thus, the definition of government in (13) together with the definition of barrier in (49) has the correct range of consequences in these Incorporation cases as well. These examples make clear how the basic sense of the Head Movement Constraint follows from the ECP.

The elements of the discussion can be fit together into the following formal proof of the HMC given the principles of GB. The question is what X^0 movements are possible. First consider the range of conceivable "landing sites" (i.e. endpoints of the movement) for a moved X^0. Excluding one-bar–level projections, which are generally rather inert,[12] there are three possible types:

(55) (i) Substitution for a maximal projection (i.e. a specifier).
(ii) Adjunction to a maximal projection.
(iii) Substitution for or adjunction to an X^0 category.

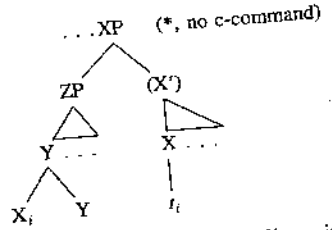
Following Chomsky (1986b), I assume that (55i) and (55ii) are impossible in general, by some version of "structure preservation" which entails that XP projections cannot take the positions of X^0 projections and vice versa.[13] This leaves only (55iii), movement to X^0 positions (which by X′ theory will be head positions) to consider. Here I will not distinguish between substitution and adjunction but simply note that adjunction will be the normal case, since, at least for the lexical categories, phrases are not usually generated with syntactically null heads.

Now, suppose that an X^0 "X" moves into an X^0 "Y" that properly governs XP, the projection of X. If Y properly governs XP, then it theta marks and hence selects XP. This is exactly case (54a) discussed above: X governs and is chain-coindexed with t, and the ECP is satisfied. Thus movement of an X^0 into a Y^0 which properly governs the XP headed by the X^0 is permitted.

Conversely, suppose that X moves somewhere else, to a Y which does not select the XP that X heads. There are two possibilities: either the landing site of X is within XP, or it is outside XP.[14] If it is within XP it is easy to see that X will not c-command its trace: since the trace of X is the head of XP, it is the only X^0 which is an immediate daughter of X′ or XP (by X-bar theory); hence Y must be contained in some maximal projection ZP properly contained in XP. X is also in ZP, which cannot contain the trace, and so

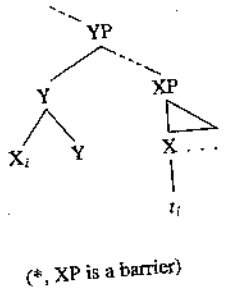
X does not c-command the trace. If it does not c-command, neither does it govern the trace. The configuration would be something like:

(56)



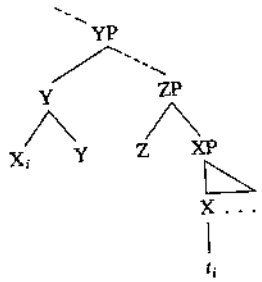
Suppose instead that the landing site of X is outside XP. Here there are two further cases to consider: either XP is selected or it is not. These two cases are schematized in (57a) and (57b):

(57) a.



(*, XP is a barrier)

b.



(*, ZP is a barrier)

If XP is not selected, it is an unselected maximal projection that dominates the trace but excludes the antecedent; hence it is a barrier between X and the trace (57a). Thus X does not govern *t*. If, on the other hand, XP is selected, then there must be some head Z which selects it, and (by assumption) Z is distinct from Y. Now normally (but see below) Z selects XP only if Z and XP are sisters within ZP, the projection of Z. By X-bar theory, Z is the only X^0 which is an immediate constituent of ZP; hence Y is not an immediate constituent of ZP. Instead, Y is either contained in the maximal projection WP (by assumption not equal to XP) contained in ZP (cf. (56)), or Y is not contained in ZP at all, as in (57b). In the first case, X will also be in WP, and WP is a maximal projection that does not contain *t;* hence X does not c-command *t;* hence X does not govern *t*. In the second case, X is also not contained in ZP. Now ZP is a maximal projection which contains *t* and excludes X, and whose head Z selects XP which contains *t;* thus ZP is a barrier between X and *t*. Once again, X does not govern *t*.

The preceding two paragraphs show that if X does not move to a head

position Y which selects XP, the projection of X, X will not govern its trace. It follows that the trace can never be antecedent-governed. Neither can it be lexically governed, since it is an X^0-level category and, as discussed above, X^0 categories never bear theta indexes. Therefore, the trace cannot be properly governed at all, and ECP is violated. Hence it is forbidden for the X^0 to move anywhere but to the Y^0 that selects its projection.

This completes the proof of the Head Movement Constraint (43).

More or less. In fact, there are two "kinks" in the proof which are important to mention, since they point to situations in which what has been derived is not quite identical to (43). The first is that (43) mentions proper government, which, as a relationship between heads, is equivalent to theta role assignment in the framework of assumptions I have adopted. My definitions of government, however, now hinge on the broader notion of selection, rather than on theta role assignment. Thus, what we have really derived is the statement that X can incorporate into Y whenever Y selects XP, the projection of X. This means, for example, that Infl can move into the complementizer position and the verb can move into the Infl position, even though nonlexical Cs and Is are not proper governors (on most views). This switch to selection is probably an improvement, given the existence of structures like the following in English, where exactly this sort of movement seems to occur:

(58) a. Who **will** Brent help next? (cf. Brent **will** help Pete)
b. $[_{CP}$who will$_i$ $[_{IP}$Brent t_i $[_{VP}$help next]].

(59) a. Ellen **cooks** blistering hot curry. (cf. Ellen does **cook** curry)
b. $[_{IP}$Ellen $[$cook$_i$+PRES$]$ $[_{VP}t_i$ hot curry]].

The second kink in the proof involves the assumption that Z selects XP only if Z and XP are sisters within ZP, the projection of Z. This is the standard case, but there are two situations in which it can fail to be true. The first is when V selects its external argument, even though that argument is not in VP but rather in IP, governed by the I ((60a)). The second involves cases of movement, where XP is selected in its D-structure position but then moves to another position, say Comp, where it is governed by C ((60b)). Concrete instances of these cases are diagrammed below:

(60) a. . . . $[_{VP}V_1$ $[_{IP_1}$ $[_{NP_2}$**N**] I $[_{VP}V_2$ . . .]]]
b. . . . $[_{VP}V_1$ $[_{CP_1}$ $[_{XP_{i2}}$**X**] C $[_{IP}$NP I $[_{VP}V_2$ t_{i2} . . .]]]

where important selection relations are represented by cosubscripting with arabic numerals. In these circumstances, the maximal projection of the highlighted X^0 is selected, so it will not be a barrier between the matrix V and the X^0. In fact, IP and CP are not barriers between the two either: they are selected (by V_1), and their heads do not select the phrase containing the

highlighted X^0. Thus, government holds between V_i and the highlighted word, and that word will be able to incorporate from the indicated positions in (60) directly into V_i, skipping over the closer head (I or C), without violating the ECP. Here the special property of the nonlexical categories of not selecting their specifiers comes into play, allowing a case of Incorporation which violates the HMC. In fact, we will see (chapter 4) that this too is an improvement, that Incorporation is slightly freer in just these cases where it comes out of a category with a nonlexical head. Note that even here X^0 movement will be strictly bounded; in (60) the highlighted X^0 would not be able to incorporate into anything beyond V_i, since the VP it heads will be a barrier.

With these comments in mind, I conclude that the Head Movement Constraint, to the extent that it is true, follows entirely from the ECP. Thus, the fact that X^0 movement obeys the HMC means that the trace of X^0 movement in fact is subject to the ECP. Thus, X^0 movement is like other types of movement in this important way. In other words, incorporation does have property (42i), the first characteristic property of the Move-Alpha relationship, as well as (42ii).

If X^0 movement and XP movement are really governed by the same principles (like the ECP), a degree of parallelism is expected to show up overtly in paradigmatic patterns, as long as other factors can be controlled for. The greatest similarity would be expected between X^0s and adjunct XPs, since adjuncts, like X^0s, are not theta-marked. Therefore their traces, like those of X^0s, must be governed by the antecedent, and their distribution should be similar in certain ways. (Argument XPs will be somewhat different, because they are properly governed by the head that theta marks them, thereby satisfying the ECP in a way unavailable to X^0s and to adjuncts.) In fact, this is true. Consider the following sentences:

(61) a. In what manner did you fix the car *t*?
b. *In what manner did you leave [to fix the car *t*]?
d. *On what table did you buy [the books *t*]? (Huang (1982))

Again following Chomsky (1986b), I assume adjunct *wh*-phrases move through a position adjoined to VP on the way to Comp. Thus, fuller S-structures of the relevant portions of (61) are represented in (62):

(62) a. . . .[$_{VP}$*t'* [$_{VP}$fix [the car] *t*]
b. *. . .[$_{VP}$*t''* [$_{VP}$leave [$_{CP}$*t'* *e* [$_{IP}$PRO to [$_{VP}$*t*. . .]]]]]
c. *. . .[$_{VP}$*t'* [$_{VP}$buy [$_{NP}$the books *t*]]]

This paradigm is directly parallel to the (abstract) paradigm for X^0 movement illustrated in (54): extraction is okay when movement is out of the immediate category ((62a) and (54a)), but it is impossible when movement is out of an adjunct ((62b) and (54b)), or when the moved item is embedded one maximal projection more deeply ((62c) and (54c)). The account of this adjunct movement paradigm is, of course, the same as the ECP account of the X^0 movement paradigm. There is no barrier between *t'* and *t* in (62a) (VP does not exclude *t'*); CP is a barrier between *t''* and *t'* in (62b) because it is not selected; NP is a barrier between *t'* and *t* in (62c) because it contains the trace and a head distinct from *t'*.[15] Thus, government holds between the first trace and the second only in (62a), and this structure alone satisfies the ECP. (61) thus shows that there is in fact a deep similarity between the distribution of incorporation and that of XP movement. Therefore, unifying the two under the schema of Move-Alpha is not a theoretical trick but an expression of significant generalizations about the structure of natural language.

The last property of Move-Alpha which we expect to appear in Incorporation processes is that they should obey the Subjacency condition (see (42iii)). In fact, this requirement is fulfilled vacuously, because, as we have seen, the ECP induces a strictly stronger locality condition on X^0 movement already. Subjacency says that movement is degraded if it crosses more than one bounding node (roughly NP or S), but it is easy to see that whenever this happens the movement will also cross at least one barrier. Hence, an X^0 undergoing such a movement will never leave a properly governed trace. Thus, we can assume that Incorporation is subject to Subjacency, but this condition will always be redundant—just as it is for the *wh*-movement of adjuncts (Chomsky (1986b)) and for NP movement in passives and raising-to-subject constructions (cf. Marantz (1982b)).

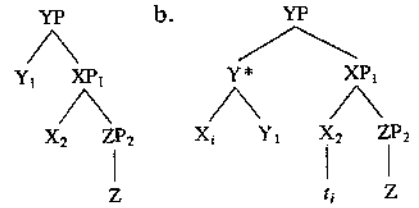
In conclusion, we have seen that Incorporation is nothing more and nothing less than a special case of the general transformational rule Move-Alpha. The main empirical consequence of this identification is that it makes it possible to derive the distribution of incorporation processes—described roughly by the Head Movement Constraint—in terms of the ECP, thereby capturing parallels with the distribution of *wh*-movement. This position will be seen in later chapters to have many desirable consequences; ultimately it will explain why only certain GF changing processes are possible. In what follows I will sometimes refer to the HMC for clarity and convenience, but it should be kept in mind that this is not a basic principle of UG, but rather a derived consequence.

2.2.4 The Government Transparency Corollary

The concepts defined in the last subsection are formulated such that they have a further consequence that will be of great importance: the conse-

quence that grammatical functions (appear to) change in incorporation structures. Consider again an abstract example such as (63b), and compare it to the parallel structure without incorporation in (63a), where theta coindexing is explicitly represented:

(63) a.



b.

In the last subsection we discussed (63a) and concluded that Y governs X in XP. It does not, however, govern ZP: XP is a barrier between Y and ZP because its head selects ZP.

There is an important difference in (63b), however. Here the lexical category Y* again governs the head of XP, allowing the trace in that position to be properly governed (technically by X). However, our principles—in particular definition (49)—imply that Y* ALSO GOVERNS ZP IN THIS CONFIGURATION. The reason for this involves the stipulation in (49ii) that a category whose head selects something containing Y is a barrier to government between Y and X only if that head is distinct from Y (or, if Y is not a head, from the head of the smallest category containing Y), where, for completeness, (64) makes explicit the sense of DISTINCTNESS which I have been assuming:

(64) X is **distinct** from Y only if no part of Y is a member of a (movement) chain containing X.

As observed before, this statement is superfluous for the cases we have considered so far. It is not superfluous with respect to a structure like (63b), however. Unlike in (63a), the potential barrier XP has as its head a trace whose antecedent is contained in the potential governor Y*; hence the heads are not distinct, and XP fails to be a barrier between Y* and ZP. This means that Y(*) comes to govern ZP by virtue of having incorporated the former governor of ZP. This result can be stated in the following terms:

(65) The GOVERNMENT TRANSPARENCY COROLLARY (GTC)
A lexical category which has an item incorporated into it governs everything which the incorporated item governed in its original structural position.

Thus, X^0 movement will automatically change the government properties of a structure in the way described in (65), simply because it, like all movement, induces a grammatical dependency between two distinct nodes. The expression "government transparency" is used because intuitively (65) says that an XP becomes transparent/invisible for the purposes of government when its head is incorporated.

The GTC is of fundamental importance because it explains why GF changing phenomena as characterized in section 1.1 are inherently associated with Incorporation. Take again the example of noun incorporation in Mohawk:

(66) a. *Ka-rakv ne* [*sawatis hrao-**nuhs**-a?*].
3N-white DET John 3M-**house**-SUF
'John's house is white.'
b. *Hrao-**nuhs**-rakv ne* [*sawatis t*].
3M-**house**-white DET John
'John's house is white.'

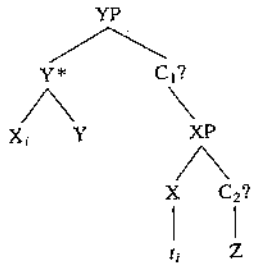
Here the unincorporated sentence (66a) has (63a) as part of its S-structure, while the incorporated sentence (66b) has (63b) as part of its S-structure, where the verb *-rakv* 'white' is "Y," the noun *-nuhs-* 'house' is "X," and the NP *sawatis* 'John' is "ZP." Note that there is a peculiar shift in verbal agreement between (66a) and (66b): in (66a) the verb has neuter agreement matching its thematic argument 'house'; whereas in (66b) it has masculine agreement, matching the possessor of its argument 'John.' Suppose (as is standard) that a verb can only agree with an NP which it governs. Then the GTC accounts for the agreement shift: in the unincorporated structure (66a) the verb does not govern the possessor and hence cannot agree with it; if, however, the intervening head is incorporated as in (66b), the verb does govern the possessor, and agreement between the two becomes possible. In other words, the possessor comes to have a canonical property of Mohawk objects as a side effect of Incorporation. This gives the appearance of Possessor Raising—one of the core GF changing processes of 1.1.2.[16] Recall that grammatical function names in GB can be defined relative to a particular subtheory of the framework, because of the framework's modular structure. Thus, we can say that 'John' changes from a possessor to an object of the matrix verb with respect to government, even though it does not change GFs with respect to X-bar theory (the standard sense of GFs in the work of Chomsky). More generally, we predict that a phrase stranded by Incorporation will behave like an object of the higher verb with respect to the government theory module and modules directly dependent on it (notably case theory), but it will not change status with respect to

X-bar theory and modules dependent on it. Thus, it will look like GFs change, but only partially so, with "changed" GFs in general retaining some of the characteristics of their original function.

Since the Government Transparency Corollary will be the mainstay of my explanation of the so-called grammatical function changing phenomena, and since some government configurations are more complex than those illustrated in (63) (e.g. (60)), I finish this section by showing that the GTC can in fact be proved formally in full generality.

Assume that X governs Z and X incorporates into Y, where X and Y are X^0 level categories and Z is any category. We wish to prove that the derived word Y* (=X+Y) governs Z. Let *t* stand for the trace of X—clearly this also governs Z—and let XP be the maximal projection of *t* (orginally of X). By the ECP, we know that X must govern *t*; therefore Y* governs *t* as well. These relationships are shown in (67):

(67)



In order to show that Y* governs Z, we must show two things: that Y* c-commands Z and that there is no barrier between Y* and Z. The first of these is immediate: Y* governs *t* and *t* governs Z; therefore Y* c-commands *t* and *t* c-commands Z. It follows that Y* must c-command Z.

To show that there is no barrier between Y* and Z, let us assume that there is a barrier—call it C—between them, and show that this leads to a contradiction. Suppose, then, that C exists. Then it must be a maximal projection that dominates Z and excludes Y. Furthermore, it must either not be selected, or it must have a head which is a "closer governor." Our strategy will be to show that under either assumption it will be either a barrier between Y* and *t* or a barrier between *t* and Z, which is inconsistent with our hypotheses (cf. (67)).

Assume first that C is not selected. Now since *t* is an X^0 and *t* c-commands Z, the first maximal projection containing *t*—namely XP—must contain Z. Hence, both XP and C contain Z. Since we assume that phrases cannot overlap (except at surface structure), it follows that C dominates XP, C equals XP, or XP dominates C. If C dominates or is equal to

XP (i.e. "C_1" in (67)), then it contains *t*. Furthermore, we have already assumed that it excludes Y* and is not selected; hence C is a barrier between Y* and *t*. But this is impossible, since Y* must govern *t* by the ECP. If, on the other hand, XP properly dominates C, then C must exclude *t*, since C is a maximal projection and *t* is the head of XP ("C_2" in (67)). Yet C also dominates Z and is not selected; hence C is now a barrier between *t* and Z. This too is impossible, since by assumption *t* governs Z. Hence, there cannot be an adjunct type barrier between Y* and Z.

Assume then that C is a minimality condition barrier. Then the head of C must be distinct from Y* and must select some maximal projection WP which dominates or is equal to Z. As above, the fact that WP and XP both contain Z implies that WP dominates XP, WP is equal to XP, or XP dominates WP. If WP dominates or is equal to XP, then it contains *t*. Then C excludes Y* and its head selects WP containing *t*, so C is a barrier between Y* and *t*. Again, this is inconsistent with the fact that Y* governs *t*. The only remaining possibility of having a barrier between Y* and Z is if XP dominates WP. Suppose that XP also dominates C. Then if the head of C is distinct from *t*, C will be a barrier between *t* and Z, contrary to the assumption that *t* governs Z. Thus, either the head of C is not distinct from *t*, or XP is equal to C—in which case the head of C is equal to *t*, since each phrase has a unique head (by X-bar theory). Therefore, the head of C is not distinct from *t*; furthermore *t* is not distinct from Y*, since Y* by hypothesis contains X, the antecedent of *t*. Therefore, the head of C is not distinct from Y*. But this contradicts the assumption that C is a minimality condition barrier between Y* and Z.

All possible assumptions about the nature of C as a barrier between Y* and Z lead to contradictions. Therefore there must be no barrier between Y* and Z. We have already established that Y* c-commands Z. Thus, Y* governs Z. This completes the proof of the Government Transparency Corollary, valid for all situations.

The last segment of the proof above shows clearly how the transparency for government of phrases whose heads have been incorporated depends crucially on the assumption that the head of a minimality condition barrier must be distinct from the potential governor in the sense of (64). Certainly, this is something of an innovation in the theory of government.[17] I believe, however, that the innovation is both minimal and rather natural. As discussed in the last subsection, we can view government as the relation of being "directly theta-connected" to some Y, a grammaticalization of the core notion of being an argument of Y. There are two ways in which one can fail to be directly theta-connected to Y: one can be not theta-connected at all (the adjunct-type barrier); or one can be more directly

theta-connected to something else, Z (the Minimality Condition–type barrier). If, however, Z is not distinct from Y, being theta-connected to Z can itself intuitively count as being theta-connected to Y, in a slightly extended notion of "connected." Thus, it is not particularly surprising that a notion such as (64) should be relevant to the definition of government. The empirical evidence that government relationships do change in just the way that the GTC describes will be overwhelming: nearly all of grammatical function changing will be understandable in these terms.

2.2.5 The Place of Morphology

So far I have emphasized the syntactic side of Incorporation; now it is time to turn to the morphological side. Thus, the last general issue about the framework to be addressed is how the theory of morphology relates to the theory of syntax. This has been a topic of lively debate in recent years: see Anderson (1982), Pranka (1983), Fabb (1984), Sproat (1985b), Williams and DiSciullo (to appear), Sadock (1985), and Marantz (1984, 1985) for a variety of views. The view which I will adopt is similar to that of Sadock (1985) and especially Marantz: I claim that morphology is in effect another subtheory, roughly on a par with the established subtheories of principles of government-binding theory enumerated in 2.1.3 (see Baker (1984, 1986)). As such, "morphology theory" (as we may call it) can be characterized as the theory of what happens when a complex structure of the form $[_{Z^0}X + Y]$ is created. In this way, it is parallel to (say) the binding theory, which is the theory of structures of the form $[NP_i \ldots NP'_i]$, where the subscript is a referential index. Morphology theory's responsibility is twofold: first, it determines whether a structure dominated by an X^0 level category is grammatical or not in a given language; second, if the structure is well-formed it assigns it a phonological shape. Thus, morphology theory may include whatever principles, universal or particular, determine the level ordering effects of Siegel (1974) and Allen (1978); principles of the strict (phonological) cycle; principles of morphological subcategorization and feature percolation such as those of Lieber (1980); and/or whatever else in this general domain proves relevant. Probably, morphology theory also has at its disposal a simple list of forms in order to deal with phonological exceptions and suppletions of various kinds.

All or many of the various functions listed above have for the last fifteen years generally been restricted to the lexicon by various strengthenings of Chomsky's (1970) "lexicalist hypothesis." I use the term "lexicon" in a specific sense, however, as a level of grammar at which the inherent properties of items are represented, in particular, those properties which are

atomic from the point of view of other levels (cf. Fabb (1984), Williams and DiSciullo (to appear)). Morphology theory, on the other hand, is like the other subtheories in that it is freed from inherent association with any one level of description—although it certainly may contain principles which make specific reference to a given level. In this way, it can be compared to (for example) government theory, which includes both the definition of government, relevant to all syntactic levels, and the ECP, which holds specifically at LF. Thus, many of the constraints of morphology theory may have uniform consequences for the combination of two X^0s into a new zero-bar–level category, regardless of where in the grammar that combination occurs. In particular, from this perspective the same morphological principles may apply when two morphemes come together in the lexicon in the standard way, and when the same morphemes come together in the syntax as a result of Incorporation.

In fact, this seems to be the usual case in natural language. To take a simple example, consider the morpheme *-ir* in Chichewa. As we saw in 1.1.3, this is the characteristic morpheme of the applicative construction in this language, which I propose to analyze as Preposition Incorporation (1.2, chapter 5). It appears in structures like the following:

(68) a. *Msangalatsi a-ku-yend-a ndi ndodo.*
entertainer SP-PRES-walk-ASP with stick
'The entertainer walked with a stick.'

b. *Msangalatsi a-ku-yend-**er**-a ndodo.*
entertainer SP-PRES-walk-**APPL**-ASP stick
'The entertainer walked with a stick.'

(69) a. *Mbalame zi-ma-uluk-a ndi mapiko.*
birds SP-HAB-fly-ASP with wings
'Birds fly with (using) wings.'

b. *Mbalame zi-ma-uluk-**ir**-a mapiko*
birds SP-HAB-fly-**APPL**-ASP wings
'Birds fly with (using) wings.'

Here the highlighted applicative morpheme in the (b) sentences is associated with assigning a semantically transparent instrumental thematic role to the postverbal NP; the same role is canonically assigned to [NP, PP] in this and other languages (cf. (68a), (69a)). The Uniformity of Theta Assignment Hypothesis implies that this morpheme is an independent constituent at D-structure; hence the (b) sentences are derived by (P) Incorporation. The verb and the affix thus come together in the syntax in these sentences. Now compare the following sentences from the same language:

(70) a. *Mkango u-ku-yend-**er**-a anyani.*
lion SP-PRES-walk-**APPL**-ASP baboons
'The lion is inspecting the baboons.'
b. *Mkango u-ku-yend-a ndi anyani.*
lion SP-PRES-walk-ASP with baboons
*'The lion is inspecting the baboons.'
(OK 'The lion is walking with the baboons.')

(71) a. *Mtolankhani a-ku-thamang-**ir**-a chiphadzuwa.*
journalist SP-PRES-run-**APPL**-ASP beauty
'The journalist ran toward/pursued the beautiful woman.'
b. *Mtolankhani a-ku-thamang-a ndi chiphadzuwa.*
journalist SP-PRES-run-ASP with beauty
*'The journalist ran toward/pursued the beautiful woman.'
(OK 'The journalist ran with the beautiful woman.')

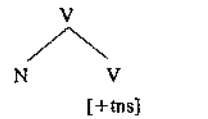
The verbs in the (a) sentences contain a recognizable morpheme identical in shape to the applicative morpheme. Yet in these cases there is no consistent theta role associated with its appearance—at least not a prepositional theta role—as comparison with the corresponding (b) sentences shows. Rather, the theta role assigned to the postverbal NP in these sentences must be listed in the lexicon as an idiosyncratic property of the forms *-yend-er-* and *-thamang-ir-*. Thus, the UTAH and the Projection Principle imply that the two morphemes in these words must not be independent constituents at any syntactic level. The verbal affix in these structures is hence a simple derivational transitivizing affix, which combines with verbs in the lexicon.

Whether or not one identifies the affix of (70), (71) with that of (68), (69) synchronically, the two share a property that must be captured by the grammar: both have two forms, *-ir-* and *-er-*, as the examples show. Which form appears is determined in both cases by a rule of vowel harmony: the tense /i/ form appears after verb stems whose last vowel is tense (/i/, /u/, or /a/); the lax /e/ form, after verb stems whose last vowel is lax (/e/ or /o/). This rule of vowel harmony is a general one in Chichewa. Thus, the same morphophonological principle determines the shape of combinations formed in the lexicon and the shape of combinations formed in the syntax. If morpho(phono)logical principles like this must be rooted in one level of the grammar, generalizations such as this will be lost. Thus, this situation points toward the view that morphology is simply the theory of structures dominated by an X^0 level node, independent of how or where this structure is formed, since such a view explains these similarities without duplicating rules or principles. Examples like this one from Chichewa will be plentiful in the chapters that follow (see also Baker (1986)).

A further virtue of this approach to the relationship between morphology and syntax is that it allows principles which are fundamentally morphological to determine syntactic structure in various ways. In this way, morphology theory is again parallel to other subtheories such as case theory and binding theory, whose requirements may force or forbid certain applications of Move-Alpha (see Chomsky (1981)). This can occur in a variety of ways.

The most important effect that morphology theory has on syntax is filtering out certain impossible incorporations. Thus, Move-Alpha can apply freely, but if it generates an X^0 level structure which morphology rules ill-formed or to which it fails to assign a phonological shape, the structure as a whole will be ungrammatical. Thus, Incorporation processes need not be absolutely productive, since a (semi-)idiosyncratic gap in a morphological paradigm will suffice to block Incorporation from taking place in those cases. Moreover, this gives us a way of answering certain questions about language variation. For example, it can be a consistent morphological property of a language that it has no productive compounds of the form:

(72)



English, in fact, has this property (apart from a few back formations derived from deverbal compounds; see Selkirk (1982)). Now, if the morphology component of a language rules out structures like (72) derived in the lexicon, it will also rule out such structures derived in the syntax, thereby making Mohawk-type Noun Incorporation impossible in the language. This gives the seeds of an explanation of what it means to say that English lacks Noun Incorporation but Mohawk has it, without claiming that there is an explicit rule of Noun Incorporation which a given language can either have or lack.

We can extend these ideas to explain why adjunction to a X^0 category is normally possible for X^0 movement but not for XP movement. It is a common principle of morphology to block syntactic phrases inside a word. Thus, one cannot normally form English compounds such as "eat-lunch-in-parks-hater," meaning 'one who hates eating lunch in parks', because of some such principle. This could be expressed roughly as:

(73) *X^0

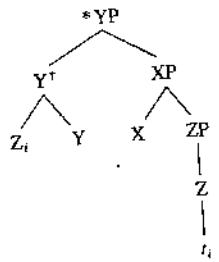
X^n, where n is greater than 0

This morphological well-formedness condition, which blocks the creation of impossible compounds in the lexicon, will also block the same structure from being formed in the syntax. This rules out adjunction to X^0 as a landing site for XP movement. This then has the consequence that "phrase incorporation" will generally not be found in natural language, a positive result (e.g. 3.2 below).[18]

This filtering function of morphology can take place in the opposite way as well. Lieber (1980) and Williams (1981a) argue that affixes are specified for all the same types of features as independent words are, including category. I accept this conclusion (for a range of cases) in a strong way when I assume that elements which appear as affixes on the surface can head phrases and assign theta roles just like normal words at the level of D-structure. The difference between affixes and words then, following Lieber, is simply that affixes must attach to a word—clearly a morphological requirement. If an item is specified as being an affix, but is generated independently at D-structure in accordance with the UTAH, that item will have to undergo X^0 movement to adjoin to some other X^0; failure to do so will result in a structure which violates a principle of morphology theory. This idea is developed in 3.4 below. Thus, morphology theory makes Incorporation obligatory in some cases and forbidden in others, even though the movement process is itself, as always, formally optional. This property of morphology theory is thus crucial to the program of eliminating explicit GF changing rules from the grammar.

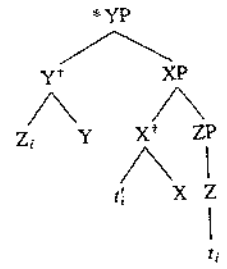
Finally, we can appeal to morphology theory to close one remaining gap in our derivation of the Head Movement Constraint from the ECP. In 2.2.3, it was shown that a structure such as (74) is ruled out by the ECP:

(74)



However, a priori another possible derivation could result in the same impossible surface string as (74) without violating the ECP: namely having Z undergo a type of "successive cyclic movement," passing through a position adjoined to X. This would yield:

(75)



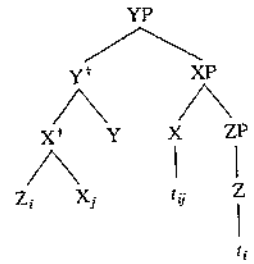
This derivation can plausibly be ruled out by morphology theory. Move-Alpha cannot in general move part of a word to some other place in the string; this part of the old lexicalist hypothesis still seems true. This can be captured by an obvious principle of morphology theory such as:

(76) *$[_{X^0} \ldots t_i \ldots]$

In other words, a trace can never be nonexhaustively dominated by a zero-level category, meaning that there are no traces inside words. This principle, of independent value, will rule out structure (75), since the category X^1 violates the constraint. Now, the HMC does truly follow from the ECP.[19]

It should be mentioned here that one kind of "successive cyclic movement" is still available to "Z," so that it can appear farther from its initial trace than is usually possible: the whole derived category X^1 can incorporate into its governor Y, yielding a structure such as:

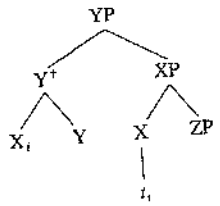
(77)



Here no morphological principles are violated. Moreover, since X^+ is coindexed with the trace of Z we may assume that when it moves it will leave a copy of this index behind on its trace. Hence, the (original) trace of Z continues to be properly governed after the second incorporation, and ECP is satisfied. In fact, we will find that sentences with substructures such as (77) are indeed attested.

Thus, the view of morphology as a semi-independent system of principles rather than as part of the lexicon proper has several attractive consequences for capturing morphophonological generalizations and for constraining the grammar. This perspective in turn makes Incorporation analyses of linguistic phenomena possible from the morphological point of view, since the complex word structures that X^0 movement generates in the syntax have the same morphological status as lexically formed structures. Thus, in a typical case of Incorporation such as:

(78)



the X^0 movement simultaneously causes a morphological change—creating a new zero level structure Y^+—and a syntactic change—creating a chain relationship between two nodes, thereby causing apparent GF changes (by the GTC). Thus an Incorporation analysis can explain the fundamental link between grammatical function changing and morphology, thereby providing a grasp on the questions raised in 1.1.3.

2.2.6 Prospectus

In this chapter, I have put forth a framework of explicit assumptions, definitions, and principles relevant to establishing a full theory of GF changing and incorporation processes. The discussion has been rather abstract and deductive, emphasizing the interconnections of certain leading ideas. It has therefore been based on a mere handful of schematic and illustrative examples, and results have in some cases been anticipated without a full range of evidence. In the chapters that follow, the emphasis shifts to concrete analyses of specific constructions in a variety of languages, using the

concepts developed here and providing the full range of evidence. Specifically, in the next five chapters I endeavor to show that for each of the GF changing processes considered in its own right there is strong empirical evidence for an analysis in terms of X^0 movement, and that in each case this analysis explains otherwise mysterious properties of the construction.