Chapter 1

The Theory of Principles and Parameters

with Howard Lasnik

1.1 Introduction

Principles-and-parameters (P&P) theory is not a precisely articulated theoretical system, but rather a particular approach to classical problems of the study of language, guided by certain leading ideas that had been taking shape since the origins of modern generative grammar some 40 years ago. These ideas crystallized into a distinctive approach to the topic by about 1980. In the years since, many specific variants have been developed and explored. The empirical base of these inquiries has also greatly expanded as they have extended to languages of widely varying types and have engaged a much broader range of evidence concerning language and its use, also penetrating to far greater depth. In this survey we will not attempt to delineate the variety of proposals that have been investigated or to assess their empirical successes and inadequacies. Rather, we will pursue a particular path through the array of ideas and principles that have been developed, sometimes noting other directions that have been pursued, but without any attempt to be comprehensive; similarly, bibliographic references are far from comprehensive, usually indicating only a few studies of particular questions. The choice of a particular path should be regarded only as an expository device, an effort to indicate the kinds of questions that are being addressed, some of the thinking that guides much research, and its empirical motivation. We

This chapter, coauthored with Howard Lasnik, was originally published in *Syntax: An International Handbook of Contemporary Research*, edited by Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld, and Theo Vennemann (Berlin and New York: Walter de Gruyter, 1993). It appears here, with minor revisions, by permission of the publisher.

do not mean to imply that these particular choices have been well established in contrast to others, only some of which we will be able even to mention.

The study of generative grammar has been guided by several fundamental problems, each with a traditional flavor. The basic concern is to determine and characterize the linguistic capacities of particular individuals. We are concerned, then, with states of the language faculty, which we understand to be some array of cognitive traits and capacities, a particular component of the human mind/brain. The language faculty has an initial state, genetically determined; in the normal course of development it passes through a series of states in early childhood, reaching a relatively stable steady state that undergoes little subsequent change, apart from the lexicon. To a good first approximation, the initial state appears to be uniform for the species. Adapting traditional terms to a special usage, we call the theory of the state attained its *grammar* and the theory of the initial state *Universal Grammar* (UG).

There is also reason to believe that the initial state is in crucial respects a special characteristic of humans, with properties that appear to be unusual in the biological world. If true, that is a matter of broader interest, but one of no direct relevance to determining the properties and nature of this faculty of the mind/brain.

Two fundamental problems, then, are to determine, for each individual (say, Jones) the properties of the steady state that Jones's language faculty attains, and the properties of the initial state that is a common human endowment. We distinguish between Jones's *competence* (knowledge and understanding) and his *performance* (what he does with that knowledge and understanding). The steady state constitutes Jones's mature linguistic competence.

A salient property of the steady state is that it permits infinite use of finite means, to borrow Wilhelm von Humboldt's aphorism. A particular choice of these finite means is a particular language, taking a language to be a way to speak and understand, in a traditional formulation. Jones's competence is constituted by the particular system of finite means he has acquired.

The notion of "infinite use" requires further analysis. In the light of insights of the formal sciences in the 20th century, we distinguish two senses of this notion, the first relating to competence, the second to performance. In the first sense, a language specifies an infinite range of symbolic objects, which we call *structural descriptions* (SDs). We may think

of the language, then, as a finitely specified generative procedure (function) that enumerates an infinite set of SDs. Each SD, in turn, specifies the full array of phonetic, semantic, and syntactic properties of a particular linguistic expression. This sense of "infinite use" relates to Jones's linguistic competence: the generative procedure with its infinite scope.

The second sense of "infinite use" has to do with Jones's performance as he makes use of his competence to express his thoughts, to refer, to produce signals, to interpret what he hears, and so on. The language faculty is embedded in performance systems, which access the generative procedure. It is in this broader context that questions of realization and use of SDs arise, questions of articulation, intentionality, interpretation, and the like: How does Jones say X? What is Jones talking about? What does Jones take Smith to be saying or intending to convey? And so on. We might think of the SD as providing instructions to the performance systems that enable Jones to carry out these actions.

When we say that Jones has the language L, we now mean that Jones's language faculty is in the state L, which we identify with a generative procedure embedded in performance systems. To distinguish this concept of language from others, let us refer to it as *I-language*, where I is to suggest "internal," "individual," and "intensional." The concept of language is internal, in that it deals with an inner state of Jones's mind/brain, independent of other elements in the world. It is individual in that it deals with Jones, and with language communities only derivatively, as groups of people with similar I-languages. It is intensional in the technical sense that the I-language is a function specified in intension, not extension: its extension is the set of SDs (what we might call the *structure* of the I-language). Two distinct I-languages might, in principle, have the same structure, though as a matter of empirical fact, human language may happen not to permit this option. That is, it might turn out that the range of I-languages permitted by UG is so narrow that the theoretical option is simply not realized, that there are no distinct I-languages generating the same set of SDs. This seems, in fact, not unlikely, but it is not a logical necessity. When we use the term *language* below, we mean I-language.

In the earliest work in generative grammar, it was assumed that Jones's language generates an SD for each of the permissible phonetic forms for human language, a set to be specified by UG. Thus, Jones's language assigns a particular status to such expressions as (1), where *t* (trace) indicates the position in which the question word is construed.

(1) a. John is sleeping
b. John seems sleeping
c. what do you think that Mary fixed *t* (answer: the car)
d. what do you wonder whether Mary fixed *t* (answer: the car)
e. how do you wonder whether Mary fixed the car *t* (answer: with a wrench)
f. expressions of Swahili, Hungarian, etc.

In fact, some of the most instructive recent work has been concerned with the differences illustrated by (1d–e), both in some sense "deviant," but assigned a different status by Jones's language (sections 1.3.3, 1.4.1); and one might well learn about the languages of Jones and Wang by studying their reactions to utterances of Swahili.

Another notion that appears commonly in the literature is "formal language" in the technical sense: set of well-formed formulas; in a familiar variety of formal arithmetic, "$(2 + 2) = 5$" but not "$2 + = 2)\,5\,($". Call such a set an *E-language*, where *E* is to suggest "external" and "extensional." In the theory of formal languages, the E-language is defined by stipulation, hence is unproblematic. But it is a question of empirical fact whether natural language has any counterpart to this notion, that is, whether Jones's I-language generates not only a set of SDs but also a distinguished E-language: some subset of the phonetic forms of UG, including some but not all of those of (1). Apart from expository passages, the concept of E-language scarcely appears in the tradition of generative grammar that we are considering here. As distinct from the notions discussed earlier, it has no known status in the study of language. One might define E-language in one or another way, but it does not seem to matter how this is done; there is no known gap in linguistic theory, no explanatory function, that would be filled were such a concept presented. Hence, it will play no role in our discussion.

In the study of formal languages, we may distinguish *weak generation* of E-language from *strong generation* of the structure of the language (the set of SDs). The *weak generative capacity* of a theory of I-languages is the set of E-languages weakly generated, and its *strong generative capacity* is the set of structures strongly generated. In the study of natural language, the concepts of structure and strong generation are central; the concepts of E-language and weak generation at best marginal, and perhaps not empirically meaningful at all. Note that if E-languages do exist, they are at a considerably further remove from mechanisms and behavior than I-language. Thus, the child is presented with specimens of behavior in particular circumstances and acquires an I-language in some manner to be determined. The I-language is a state of the mind/brain. It has a certain structure (i.e., strongly generates a set of SDs). It may or may not also weakly generate an E-language, a highly abstract object remote from mechanisms and behavior.

In the terms just outlined, we can consider some of the classical problems of the study of language.

(2) a. What does Jones know when he has a particular language?
b. How did Jones acquire this knowledge?
c. How does Jones put this knowledge to use?
d. How did these properties of the mind/brain evolve in the species?
e. How are these properties realized in mechanisms of the brain?

Under (2a), we want to account for a wide variety of facts, for example, that Jones knows that

(3) a. *Pin* rhymes with *bin*.
b. Each expression of (1) has its specific status.
c. If Mary is too clever to expect anyone to catch, then we don't expect anyone to catch Mary (but nothing is said about whether Mary expects anyone to catch us).
d. If Mary is too angry to run the meeting, then either Mary is so angry that *she* can't run the meeting, or she is so angry that *we* can't run the meeting (compare: *the crowd is too angry to run the meeting*); in contrast, *which meeting is Mary too angry to run* has only the former (nondeviant) interpretation.
e. If Mary painted the house brown, then its exterior (not necessarily its interior) is brown.
f. If Mary persuaded Bill to go to college, then Bill came to intend to go to college (while Mary may or may not have).

The proposed answer to problem (2a) would be that Jones has language L generating SDs that express such facts as (3). Note that Jones has this knowledge whether or not he is aware of these facts about himself; it may take some effort to elicit such awareness, and it might even be beyond Jones's capacities. This is a question that falls within the broader context of performance systems.

The answer to problem (2b) lies in substantial part in UG. The correct theory of the initial state will be rich enough to account for the attainment

of a specific language on the basis of the evidence available to the child, but not so rich as to exclude attainable languages. We may proceed to ask as well how environmental factors and maturational processes interact with the initial state described by UG.

Problem (2c) calls for the development of performance theories, among them, theories of production and interpretation. Put generally, the problems are beyond reach: it would be unreasonable to pose the problem of how Jones decides to say what he does, or how he interprets what he hears in particular circumstances. But highly idealized aspects of the problem are amenable to study. A standard empirical hypothesis is that one component of the mind/brain is a *parser*, which assigns a percept to a signal (abstracting from other circumstances relevant to interpretation). The parser presumably incorporates the language and much else, and the hypothesis is that interpretation involves such a system, embedded in others.

It has sometimes been argued that linguistic theory must meet the empirical condition that it account for the ease and rapidity of parsing. But parsing does not, in fact, have these properties. Parsing may be slow and difficult, or even impossible, and it may be "in error" in the sense that the percept assigned (if any) fails to match the SD associated with the signal; many familiar cases have been studied. In general, it is not the case that language is readily usable or "designed for use." The subparts that are used are usable, trivially; biological considerations lead us to expect no more than that. Similarly, returning to problem (2b), there is no a priori reason to expect that the languages permitted by UG be learnable—that is, attainable under normal circumstances. All that we can expect is that some of them may be; the others will not be found in human societies. If proposals within the P&P approach are close to the mark, then it will follow that languages are in fact learnable, but that is an empirical discovery, and a rather surprising one.

Problems (2d–e) appear to be beyond serious inquiry for the time being, along with many similar questions about cognition generally. Here again one must be wary of many pitfalls (Lewontin 1990). We will put these matters aside.

A grammar for Jones is true if (or to the extent that) the language it describes is the one Jones has. In that case the grammar will account for such facts as (3), by providing a language that generates appropriate SDs. A true grammar is said to meet the condition of *descriptive adequacy*. A theory of UG is true if (or to the extent that) it correctly describes the initial state of the language faculty. In that case it will provide a descriptively adequate grammar for each attainable language. A true theory of UG meets the condition of *explanatory adequacy*. The terminology is intended to suggest a certain plausible pattern of explanation. Given an array of facts such as (3), we can give an account of them at one level by providing a grammar for Jones, and we can provide an explanation for them at a deeper level by answering problem (2b), that is, by showing how these facts derive from UG, given the "boundary conditions" set by experience. Note that this pattern of explanation, though standard, makes certain empirical assumptions about the actual process of acquisition that are by no means obviously true, for example, that the process is *as if* it were instantaneous. Such assumptions are indirectly supported to the extent that the explanations succeed.

Any serious approach to complex phenomena involves innumerable idealizations, and the one just sketched is no exception. We do not expect to find "pure instantiations" of the initial state of the language faculty (hence of UG). Rather, Jones will have some jumble of systems, based on the peculiar pattern of his experience. The explanatory model outlined deals specifically with language acquisition under the idealized conditions of a homogeneous speech community. We assume that the system described by UG is a real component of the mind/brain, put to use in the complex circumstances of ordinary life. The validity of this assumption is hardly in question. To reject it would be to assume either (1) that nonhomogeneous (conflicting) data are required for language acquisition, or (2) that the mind/brain does indeed have the system described by UG, but it is not used in language acquisition. Neither assumption is remotely plausible. Rejecting them, we accept the approach just outlined as a reasonable approach to the truth about humans, and a likely prerequisite to any serious inquiry into the complex and chaotic phenomenal world.

Furthermore, even if a homogeneous speech community existed, we would not expect its linguistic system to be a "pure case." Rather, all sorts of accidents of history would have contaminated the system, as in the properties of (roughly) Romance versus Germanic origin in the lexicon of English. The proper topic of inquiry, then, should be a theory of the initial state that abstracts from such accidents, no trivial matter. For working purposes (and nothing more than that), we may make a rough and tentative distinction between the *core* of a language and its *periphery*, where the core consists of what we tentatively assume to be pure

instantiations of UG and the periphery consists of marked exceptions (irregular verbs, etc.). Note that the periphery will also exhibit properties of UG (e.g., ablaut phenomena), though less transparently. A reasonable approach would be to focus attention on the core system, putting aside phenomena that result from historical accident, dialect mixture, personal idiosyncrasies, and the like. As in any other empirical inquiry, theory-internal considerations enter into the effort to pursue this course, and we expect further distinctions to be necessary (consider, for example, the phenomenon of *do*-insertion in English as in (1c–e), not on a par with irregular verbs, but not of the generality of fronting of question words).

The preceding remarks are largely conceptual, though not without empirical consequences. We now proceed along a particular path, in the manner indicated earlier, assuming further empirical risk at each point.

We assume that the language (the generative procedure, the I-language) has two components: a computational system and a lexicon. The first generates the form of SDs; the second characterizes the lexical items that appear in them. Many crucial questions arise as to how these systems interact. We will assume that one aspect of an SD is a system of representation, called *D-Structure*, at which lexical items are inserted. D-Structure expresses lexical properties in a form accessible to the computational system.

We assume further a distinction between *inflectional* and *derivational* processes of morphology, the latter internal to the lexicon, the former involving computational operations of a broader syntactic scope. These computational operations might involve *word formation* or *checking*. Consider for example the past tense form *walked*. The lexicon contains the root [walk], with its idiosyncratic properties of sound, meaning, and form specified; and the inflectional feature [tense], one value of which is [past]. One of the computational rules, call it *R*, associates the two by combining them (either adjoining [walk] to [tense], or conversely). We might interpret this descriptive comment in two ways. One possibility is that [walk] is drawn from the lexicon as such; then *R* combines it with [past]. A second possibility is that processes internal to the lexicon (*redundancy rules*) form the word *walked* with the properties [walk] and [past] already specified. The rule *R* then combines the amalgam with [past], checking and licensing its intrinsic feature [past]. In this case the lexicon is more structured. It contains the element [walk], as before, along with rules indicating that any verb may also intrinsically possess

such properties as [past], [plural], and the like. Similar questions arise about complex words (causatives, noun incorporation structures, compound nouns, etc.). As these topics are pursued with more precision, within more closely articulated theories, important and often subtle empirical issues arise (Marantz 1984, Fabb 1984, Baker 1988, Di Sciullo and Williams 1988, Grimshaw 1990).

The SD provides information (to be interpreted by performance systems) about the properties of each linguistic expression, including its sound and its meaning. We assume that the design of language provides a variety of symbolic systems (*levels of representation*) fulfilling these tasks, including the level of *Phonetic Form* (PF) and the level of *Logical Form* (LF), specifying aspects of sound and meaning, respectively, insofar as they are linguistically determined. Another is the level of D-Structure, which relates the computational system and the lexicon.

The level PF must satisfy three basic conditions of adequacy. It must be *universal*, in the sense that an expression of any actual or potential human language is representable within it. It must be an *interface*, in that its elements have an interpretation in terms of the sensorimotor systems. And it must be *uniform*, in that this interpretation is uniform for all languages, so as to capture all and only the properties of the system of language as such.

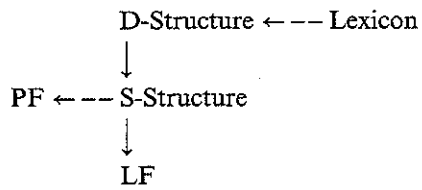
The same three conditions hold for LF. To capture what the language faculty determines about the meaning of an expression, it must be universal, in that any thought expressible in a human language is representable in it; an interface, in that these representations have an interpretation in terms of other systems of the mind/brain involved in thought, referring, planning, and so on; and uniform, in just the sense that the phonetic system is. We will put aside important questions concerning the nature of the LF interface: does it involve a conceptual system (Jackendoff 1983, 1990b), a use theory of meaning, a causal theory of reference, etc.? The conditions are more obscure than in the case of the phonetic analogue, because the systems at the interface are much less well understood, but there is nonetheless a wealth of evidence firm enough to allow substantive inquiry.

According to this conception, then, each SD contains three interface levels: the external interface levels PF and LF, and the internal interface level of D-Structure. The elements at these levels are further analyzed into features: phonological, selectional, categorial, and so on. In general,

each symbol of the representations is a feature set, in respects to be further specified.

A further assumption, developed in the *Extended Standard Theory* (EST), is that these levels are not related directly; rather, their relations are mediated by an intermediate level of *S-Structure*. Adopting this view, each SD is a sequence $(\pi, \lambda, \delta, \sigma)$, where π and λ are representations at the external interface levels PF and LF, δ is at the internal interface of computational system and lexicon, and σ is derivative. The first three levels meet empirical conditions imposed by the performance systems and the lexicon. The level of S-Structure must relate to these three levels in the manner specified in UG; we might think of it, informally, as the (presumably unique) "solution" to this set of conditions. In the subsequent discussion we restrict ourselves largely to the levels D-Structure, S-Structure, and LF, and the relations among them (syntax in a narrow sense). We are thus concerned primarily with the derivation from D-Structure to LF in (4).

(4)



Subtle questions arise as to how the relations among these levels are to be construed: specifically, is there an inherent "directionality," so that the relations should be construed as a mapping of one level to another, or is there simply a nondirectional relation? To formulate this as a real empirical issue is not a simple matter, and empirical evidence to distinguish such possibilities is not easy to come by. But interesting (and conflicting) arguments have been presented. Discrimination among these alternatives becomes particularly difficult if we adopt (as we will) the standard EST assumption, from the early 1970s, that representations may include *empty categories* (ECs): elements (feature sets) that are perfectly substantive from the point of view of the computational system, but that do not happen to be assigned an interpretation by the mapping from S-Structure to PF, though they may have indirect phonetic effects; thus, the contraction rules of English convert *want to* into the phonological word *wanna* when there is no trace intervening (*who do you wanna see* but not *who do you wanna see John* (Chomsky and Lasnik 1977)).

We will tentatively proceed on the assumption that the relations are, in fact, directional: D-Structure is mapped to S-Structure, which is (independently) mapped to PF and LF.

The earliest modern work in generative grammar borrowed standard ideas of traditional grammar, which recognized (I) that a sentence has a hierarchy of phrases (noun phrases, clauses, etc.) and that these (or their heads) enter into certain grammatical relations; and (II) that sentences belong to various grammatical constructions with systematic relations among them, some more "basic" than others (actives more basic than passives, declaratives more basic than interrogatives, etc.). Correspondingly, the earliest versions of UG provided two kinds of rules: (I) phrase structure rules generating SDs that express the hierarchy of phrases; and (II) transformational rules that form grammatical constructions from abstract underlying forms, with more transformations involved in formation of the less basic constructions (thus, only obligatory transformations apply to form active declaratives (*kernel sentences*), but some optional ones are involved in formation of passives, interrogatives, etc.). The phrase structure rules provide a "geometrical" account of grammatical relations, understood relationally; that is, subject is not a syntactic category like noun phrase or verb, but is understood as the relation subject-of holding of the pair (*John, left*) in *John left*, and so on (Chomsky 1951, 1965, 1975a). These notions were defined in such a way that the phrase structure rules (I) generate D-Structures (deep structures), each a *phrase marker* that represents hierarchy and relations. Transformations convert these objects into new phrase markers. In the later EST version, as noted, D-Structures are mapped to S-Structures by such derivations, and the latter are mapped independently to PF and LF.

The resort to phrase structure rules was also suggested by other considerations. The earliest work concentrated on what is now called generative phonology, and in this domain "rewriting rules" of the form $X \rightarrow Y$, where X is an expression "rewritten" as Y in the course of derivation, seems an appropriate device. If these rules are restricted to the form $XAY \rightarrow XZY$, A a single symbol and Z nonnull, then we have a system of rules that can form phrase structure representations in a natural way (*context-free* rules if X, Y are null). Further motivation derived from the theory of formal systems. Grammatical transformations as generative devices were suggested by work of Harris (1952), which used

formal relations among expressions as a device to "normalize" texts for the analysis of discourse.

As for UG, the earliest versions assumed that it provided a format for rule systems and an evaluation metric that assigned a "value" to each generative procedure of the proper format. The crucial empirical condition on UG, then, is that the system provide only a few high-valued I-languages consistent with the kinds of data available to the child, perhaps only one. If UG is *feasible* in this sense, the fundamental problem (2b) can be addressed (Chomsky 1965).

This approach recorded many achievements, but faced a fundamental and recurrent problem: the tension between descriptive and explanatory adequacy. To achieve descriptive adequacy, it seemed necessary to enrich the format of permissible systems, but in doing so we lose the property of feasibility, so that problem (2b) is still unresolved. The conflict arises as soon as we move from the intuitive hints and examples of traditional grammar to explicit generative procedures. It was quickly recognized that the problem is inherent in the kinds of rule systems that were being considered. The most plausible approach to it is to try to "factor out" overarching principles that govern rule application generally, assigning them to UG; the actual rules of grammar can then be given in the simplest form, with these principles ensuring that they will operate in such a way as to yield the observed phenomena in their full complexity (Chomsky 1964, Ross 1967). The limit that might be reached is that rules are eliminated entirely, the "apparent rules" being deduced from general principles of UG, in the sense that the interaction of the principles would yield the phenomena that the rules had been constructed to describe. To the extent that this result can be achieved, the rules postulated for particular languages will be shown to be epiphenomena.

Such ideas were pursued with a good deal of success from the early 1960s, leading to the P&P approach, which assumed that the limit can in fact be attained: the hypothesis is that all principles are assigned to UG and that language variation is restricted to certain options as to how the principles apply. If so, then rule systems are eliminable, at least for the core of the language.

To illustrate, consider again (1c–e), repeated here.

(1) c. what do you think that Mary fixed *t*
d. what do you wonder whether Mary fixed *t*
e. how do you wonder whether Mary fixed the car *t*

The goal is to show that the question words move from the position of *t* by a general principle that allows movement quite freely, with the options, interpretations, and varying status determined by the interaction of this principle with others.

What is the status of the rules (I) (phrase structure) and (II) (transformational) under this conception? The transformational rules still exist, but only as principles of UG, freely applicable to arbitrary expressions. Such devices appear to be unavoidable in one or another form, whether taken to be operations forming derivations or relations established on representations. As for phrase structure rules, it appears that they may be completely superfluous. That would not be too surprising. With the advantage of hindsight we can see that, unlike transformational rules, they were a dubious device to begin with, recapitulating information that must be presented, ineliminably, in the lexicon. For example, the fact that *persuade* takes a noun phrase (NP) and clausal phrase (CP) as complements, as a lexical property, requires that there be phrase structure rules yielding V–NP–CP as an instantiation of the phrase XP headed by the verb *persuade*; and completely general properties require further that XP must be VP (verb phrase), not, say, NP. The apparent eliminability of phrase structure rules became clear by the late 1960s, with the separation of the lexicon from the computational system and the development of *X-bar theory* (section 1.3.2).

The issues can be sharpened by considering two properties that descriptive statements about language might have or lack. They may or may not be language-particular; they may or may not be construction-particular. The statements of traditional grammar are typically both language- and construction-particular, and the same is true of the rules of early generative grammar. Consider the rule analyzing VP as V–NP, or the rules fronting the question word in different ways in (ic–e). Spelled out in full detail, these phrase structure and transformational rules are specific to English and to these constructions. There are few exceptions to this pattern.

The P&P approach aims to reduce descriptive statements to two categories: language-invariant, and language-particular. The language-invariant statements are principles (including the parameters, each on a par with a principle of UG); the language-particular ones are specifications of particular values of parameters. The notion of construction, in the traditional sense, effectively disappears; it is perhaps useful for descriptive taxonomy but has no theoretical status. Thus, there are no such

constructions as Verb Phrase, or interrogative and relative clause, or passive and raising constructions. Rather, there are just general principles that interact to form these descriptive artifacts.

The parametric options available appear to be quite restricted. An assumption that seems not unrealistic is that there is only one computational system that forms derivations from D-Structure to LF; at some point in the derivation (S-Structure), the process branches to form PF by an independent phonological derivation (as in (4)). Options would then be restricted to two cases: (1) properties of the lexicon, or (2) the point in the derivation (4) from D-Structure to LF at which structures are mapped to PF (S-Structure) (Stowell 1986).

In the category (1), apart from Saussurean arbitrariness and some limited variety in the choice of substantive elements, we have options as to how nonsubstantive (functional) elements are realized (Borer 1984, Fukui 1986, Speas 1986) and variations in global properties of heads (e.g., do verbs precede or follow their complements?) (Travis 1984).

In the category (2) we find, for example, languages with overt movement of question phrase (English, Italian, etc.) and languages without overt movement (Chinese, Japanese, etc.). In these *in-situ* languages, with the question phrase in the position that would be occupied by a trace in languages with overt movement, there is good evidence that similar movement operations take place, but only in the mapping from S-Structure to LF, with no indication in the physical form itself; the branch point at which PF is formed from S-Structure precedes these operations in the derivation (4) from D-Structure to LF (Huang 1982, Lasnik and Saito 1984, 1992). Similarly, we find languages with overt manifestation of grammatical case (Greek, German, Japanese, etc.) and others with virtually no such manifestation (English, Chinese, etc.). But again, there is good reason to believe that the case systems are basically similar cross-linguistically and that the differences lie primarily in their phonetic realization (the mapping to PF).

The general expectation, for all constructions, is that languages will be very similar at the D-Structure and LF levels, as in the examples just discussed. It is unlikely that there are parameters that affect the form of LF representation or the computational process from S-Structure to LF; little evidence is available to the language learner bearing on these matters, and there would be no way for values to be determined with any reliability. Accordingly, any variations at the LF level must be reflexes of D-Structure parameter settings, or of variations in the mapping from

D-Structure to S-Structure to the extent that its properties are determined from inspection of PF forms. D-Structure, in turn, reflects lexical properties; these too appear to be limited in variety insofar as they affect the computational system. At the PF level, properties of the language can be readily observed and variation is possible within the fixed repertoire of phonetic properties and the invariant principles of universal phonetics. S-Structures are not constrained by interface conditions and can vary within the range permitted by the variation of the interface levels, the branch point to the PF mapping, and any independent conditions that may hold of S-Structure.

The principles that have been investigated fall into two general categories: principles that are applied to construct derivations (transformational operations and conditions on the way they operate), and principles that apply to representations (licensing conditions). The transformational operations are movement (adjunction, substitution), deletion, and perhaps insertion; we may think of these as instances of the general operation *Affect* α, α arbitrary (Lasnik and Saito 1984). Conditions of locality and others constrain the application and functioning of these operations. Licensing conditions at the external interface levels PF and LF establish the relation of language to other faculties of the mind/brain. D-Structure conditions specify the manner in which lexical properties are expressed in grammatical structures. That there should be S-Structure conditions is less obvious, but it seems that they may exist (see section 1.3.3).

The principles have further structure. There are natural groupings into *modules* of language (binding theory, θ-theory, Case theory, etc.). Certain unifying concepts enter into many or all modules: conditions of locality, "geometrical" properties defined on phrase markers, and so on. There are also certain general ideas that appear to have wide applicability, among them, principles of economy stating that there can be no superfluous symbols in representations (the principle of Full Interpretation, FI) or superfluous steps in derivations (Chomsky 1986b, chapters 2–4 of this book). As these principles are given an explicit formulation, they become empirical hypotheses with specific import and range.

The principle FI is assumed as a matter of course in phonology; if a symbol in a representation has no sensorimotor interpretations, the representation does not qualify as a PF representation. This is what we called the "interface condition." The same condition applied to LF also entails that every element of the representation have a (language-independent) interpretation. There can, for example, be no true expletives,

or vacuous quantifiers, at the LF level. The principle of economy of derivation requires that computational operations must be driven by some condition on representations, as a "last resort" to overcome a failure to meet such a condition. Interacting with other principles of UG, such economy principles have wide-ranging effects and may, when matters are properly understood, subsume much of what appears to be the specific character of particular principles.

The shifts in focus over the years alter the task of inquiry considerably and yield different conceptions of what constitutes a "real result" in the study of language. Suppose we have some collection of phenomena in a particular language. In the early stages of generative grammar, the task was to find a rule system of the permitted form from which these phenomena (and infinitely many others) could be derived. That is a harder task than the ones posed in pregenerative grammar, but not an impossible one: there are many potential rule systems, and it is often possible to devise one that will more or less work—though the problem of explanatory adequacy at once arises, as noted.

But this achievement, however difficult, does not count as a real result if we adopt the P&P approach as a goal. Rather, it merely sets the problem. The task is now to show how the phenomena derived by the rule system can be deduced from the invariant principles of UG with parameters set in one of the permissible ways. This is a far harder and more challenging task. It is an important fact that the problem can now be posed realistically, and solved in interesting ways in some range of cases, with failures that are also interesting insofar as they point the way to better solutions. The departure from the long and rich tradition of linguistic inquiry is much sharper and more radical than in early generative grammar, with problems that are quite new and prospects that appear promising.

Other traditional problems also assume a different form under a P&P approach. Questions of typology and language change will be expressed in terms of parameter choice (Lightfoot 1991). The theory of language acquisition will be concerned with acquisition of lexical items, fixing of parameters, and perhaps maturation of principles (Hyams 1986, Roeper and Williams 1987, Borer and Wexler 1987; Chien and Wexler 1991, Crain 1991, Pierce 1992). It might turn out that parsers are basically uniform for all languages: the parsers for English and Japanese would differ only in that parameters are set differently (Fong 1991). Other

issues would also require some rethinking, if this approach turns out to be correct.

Much of the most fruitful inquiry into generative grammar in the past years has pursued the working hypothesis that UG is a simple and elegant theory, with fundamental principles that have an intuitive character and broad generality. By dissolving the notion of construction and moving toward "rule-free" systems, the P&P approach carries this tendency considerably forward. A related assumption is that UG is "nonredundant," in the sense that phenomena are explained by interaction of principles in one particular way. Discovery that phenomena are "overdetermined" has commonly been taken to indicate a theoretical deficiency that should be overcome by new or refined principles. These working hypotheses have proven successful as a guide to inquiry, leading to the discovery of a vast range of empirical phenomena in widely varied languages and to forms of explanation that much exceed what could be contemplated not many years ago. These are rather surprising facts. The guiding ideas resemble those often adopted in the study of inorganic phenomena, where success has often been spectacular since the 17th century. But language is a biological system, and biological systems typically are "messy," intricate, the result of evolutionary "tinkering," and shaped by accidental circumstances and by physical conditions that hold of complex systems with varied functions and elements. Redundancy is not only a typical feature of such systems, but an expected one, in that it helps to compensate for injury and defect, and to accommodate to a diversity of ends and functions. Language use appears to have the expected properties; as noted, it is a familiar fact that large parts of language are "unusable," and the usable parts appear to form a chaotic and "unprincipled" segment of the full language. Nevertheless, it has been a fruitful working hypothesis that in its basic structure, the language faculty has properties of simplicity and elegance that are not characteristic of complex organic systems, just as its infinite digital character seems biologically rather isolated. Possibly these conclusions are artifacts reflecting a particular pattern of inquiry; the range of completely unexplained and apparently chaotic phenomena of language lends credibility to such skepticism. Still, the progress that has been made by the contrary stance cannot be overlooked.

The P&P approach is sometimes termed *Government-Binding (GB) Theory*. The terminology is misleading. True, early efforts to synthesize

current thinking in these terms happened to concentrate on the theories of government and of binding (Chomsky 1981a), but these modules of language stand alongside many others: Case theory, θ-theory, and so on. It may turn out that the concept of government has a kind of unifying role, but there is nothing inherent to the approach that requires this. Furthermore, insofar as the theories of government and binding deal with real phenomena, they will appear in some form in every approach to language; this approach has no special claim on them. Determination of the nature of these and other systems is a common project, not specific to this particular conception of the nature of language and its use.

1.2 The Lexicon

A person who has a language has access to detailed information about words of the language. Any theory of language must reflect this fact; thus, any theory of language must include some sort of lexicon, the repository of all (idiosyncratic) properties of particular lexical items. These properties include a representation of the phonological form of each item, a specification of its syntactic category, and its semantic characteristics. Of particular interest in this discussion are the s(-emantic) selection and thematic properties of lexical heads: verbs, nouns, adjectives, and pre- or postpositions. These specify the "argument structure" of a head, indicating how many arguments the head licenses and what semantic role each receives. For example, the verb *give* must be specified as assigning an agent role, a theme role, and a goal/recipient role. In (5) *John*, *a book*, and *Mary* have these respective thematic (θ-)roles.

(5) John gave a book to Mary

The association between assigned θ-roles and argument positions is to a large extent predictable. For example, "agent" is apparently never assigned to a complement. And to the extent that the association is predictable rather than idiosyncratic, it need not (hence, must not) be stated in particular lexical entries.

This conception of the lexicon is based on that developed in the 1960s (Chomsky 1965), but it departs from it in certain respects. There, subcategorization and selectional conditions played a central role. The former conditions state for a lexical head what phrasal categories it takes as complements—for example, that *kick* takes an NP complement. The latter conditions specify intrinsic semantic features of the complement(s)

and subject. In this case the NP complement of *kick* is [+concrete]. It was noted in section 1.1 that phrase structure rules are (largely) redundant with subcategorization, hence are (largely) eliminable. But now note that subcategorization follows almost entirely from θ-role specification. A verb with no θ-role to assign to a complement will not be able to take a complement. A verb with (obligatory) θ-roles to assign will have to occur in a configuration with enough arguments (possibly including complements) to receive those θ-roles. Further, at least in part, selectional restrictions will also be determined by thematic properties. To receive a particular θ-role, the inherent semantic features of an argument must be compatible with that θ-role.

These tentative conclusions about the organization of the lexicon raise important questions about the acquisition of lexical knowledge. Suppose that subcategorization (c-selection) is artifactual, its effects derived from semantic properties (s-selection). It is reasonable to ask whether this is a consequence of the acquisition procedure itself (Pesetsky 1982). Pesetsky (developing ideas of Grimshaw (1979)) suggests that this must be so. He compares the primitives of c-selection (syntactic categories such as NP, CP, etc.) with those of θ-theory ("agent," "patient," "goal," etc.) and argues that the latter, but not the former, meet what we may call the condition of *epistemological priority*. That is, they can plausibly be applied by the learner to provide a preliminary, prelinguistic analysis of a reasonable sample of data and thus can provide the basis for development from the initial state to the steady state. This is an attractive line of reasoning, but, given our current understanding of these issues, it is not conclusive. While it does seem correct that the primitives of c-selection do not have epistemological priority, it is not at all clear that those of s-selection do have such a status. Although the notion "agent of an action" is possibly available to the child in advance of any syntactic knowledge, it is less clear that the θ-theoretic notion "agent of a *sentence*" is. That is, before the child knows anything about the syntax of his or her language (beyond what is given by UG), can the child determine what portion of a sentence constitutes the agent? Further, the evidence available to the learner likely consists of sentences rather than simply individual verbs in isolation. But such sentences explicitly display c-selection properties: they exhibit verbs along with their complements. Thus, the child is simultaneously presented with evidence bearing on both s-selection (given that sentences are presented in context, and assuming that the relevant contexts can be determined) and c-selection. It

is reasonable to assume that both aspects of the evidence contribute to the development of the knowledge. Alongside the state of affairs outlined by Pesetsky, the converse situation with c-selection evidence in fact providing information about the meanings of verbs (Lasnik 1990, Gleitman 1990) might also obtain. For example, exposure to a sentence containing a clausal complement to an unfamiliar verb would lead the learner to hypothesize that the verb is one of propositional attitude.

This scenario is not necessarily in conflict with Pesetsky's initial point about the organization of lexical entries. The means by which knowledge is arrived at is not invariably reflected in the form that the knowledge ultimately takes. For example, Grimshaw (1981) argues that the acquisition of the syntactic category of a lexical item is based in part on the notion "canonical structural realization" (CSR). The CSR of a physical object is N, that of an action is V, and so on. In the absence of evidence, the child will assume that a word belongs to its CSR—that, say, a word referring to an action is a verb. As Grimshaw indicates, while such "semantic bootstrapping" might constitute part of the acquisition procedure, the resulting steady-state lexicon has no such requirement. Languages commonly have nouns, like *destruction*, referring to actions (as well as verbs, like *be*, that don't refer to actions).

Note that this consideration indicates that lexical entries contain at least some syntactic information, in addition to the phonological and semantic information that surely must be present. Grimshaw argues that further syntactic specification is needed as well, c-selection in addition to s-selection. To consider one example, Grimshaw observes that the semantic category "question" can be structurally realized as either a clause, as in (6), or an NP, as in (7).

(6) Mary asked [what time it was]

(7) Mary asked [the time]

The verb *ask* semantically selects a question. Grimshaw argues that it is also necessary to specify that it c-selects clause or NP in order to distinguish it from *wonder*, which only takes a clause (where * indicates deviance).

(8) Mary wondered [what time it was]

(9) *Mary wondered [the time]

Since, as suggested above, it is possible to reduce most of c-selection to s-selection, the question arises whether such reduction might somehow

be available in this instance as well. Pesetsky argues that it is. As we will see in section 1.4.3, NPs must receive abstract Case from a Case assigner while clauses need not. (Henceforth, we will capitalize the word *Case* in its technical usage.) Given this, Pesetsky proposes that the difference between *ask* and *wonder* need not be stated in terms of c-selection, but rather follows from a Case difference: *ask* assigns objective Case but *wonder* does not. In this regard, *wonder* patterns with adjectives, which also do not assign objective Case.

(10) Mary is uncertain [what time it is]

(11) *Mary is uncertain [the time]

Pesetsky presents further evidence for this Case-assigning distinction between verbs like *ask* and those like *wonder*. In English, generally only objective Case-assigning verbs can occur in the passive. Given this, (6) and (8) contrast in precisely the predicted fashion.

(12) it was asked what time it was

(13) *it was wondered what time it was

As Pesetsky notes, a descriptive generalization pointed out by Grimshaw now follows: among the verbs that s-select questions, some c-select clause or NP while others c-select only clause; none c-select only NP. There are Case-assigning differences among verbs, and these are relevant to c-selection of NP (because of the Case requirement of NPs), but not of clauses.

This reduction seems quite successful for a wide range of cases, but it is important to note that formal syntactic specifications in lexical entries have not been entirely eliminated in favor of semantic ones. Whether or not a verb assigns objective Case is, as far as is known at present, a purely formal property not deducible from semantics. While much of c-selection follows from s-selection, there is a syntactic residue, statable, if Pesetsky is correct, in terms of lexically idiosyncratic Case properties.

We will introduce further properties of the lexicon as required by the exposition.

1.3 Computational System

1.3.1 General Properties of Derivations and Representations

The generative procedure that constitutes the (I-)language consists of a lexicon and a computational system. In section 1.2 we outlined some

properties of the lexicon. We now turn to the computational system. Under the general assumptions of section 1.1, we consider the four levels of representation of the EST system and the relations that hold among them, focusing attention on "narrow syntax," that is, the derivation relating D-Structure, S-Structure, and LF.

D-Structure, LF, and PF are interface levels, which satisfy the general condition FI in a manner to be made precise. Each level is a symbolic system, consisting of atomic elements (primes) and objects constructed from them by concatenation and other operations. We take these objects to be phrase markers in the familiar sense (represented conventionally by trees or labeled bracketing). Each prime is a feature complex, though for orthographic convenience we will generally use conventional symbols. For concreteness, take categories to be as in (14), for nouns, verbs, adjectives, and pre- and postpositions, respectively.

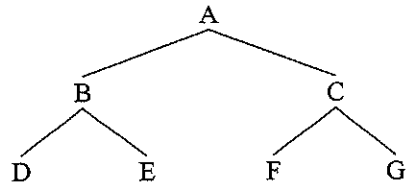
(14) a. N = [+N, −V]
b. V = [−N, +V]
c. A = [+N, +V]
d. P = [−N, −V]

The feature [+N] is the traditional substantive; the feature [+V], predicate.

The primes constituting the terminal string of a phrase marker are drawn from the lexicon; others are *projected* from these *heads* by operations of the computational system. Elements that project no further are *maximal projections*. In informal notation, XP is the maximal projection from the terminal category X; thus, NP is the maximal projection of its head N, and so on. See section 1.3.2.

The two basic relations of a phrase marker are *domination* and *linearity*. In the phrase marker (15) we say that B dominates D and E, C dominates F and G, and A dominates all other categories (nodes). Furthermore, B precedes C, F, and G; D precedes E, C, F, and G; and so on.

(15)



If X is a head, its "sister" is its *complement*; thus, if D and F are heads, then E and G are their complements in (15). We assume that ordering relations are determined by a few parameter settings. Thus, in English, a *right-branching* language, all heads precede their complements, while in Japanese, a *left-branching* language, all heads follow their complements; the order is determined by one setting of the *head parameter*. Examples below that abstract from particular languages are usually to be interpreted independently of the order given. Domination relations are determined by general principles (section 1.3.2).

One fundamental concept that applies throughout the modules of grammar is *command* (Klima 1964, Langacker 1969, Lasnik 1976, Reinhart 1976, Stowell 1981, Aoun and Sportiche 1981). We say that α *c-commands* β if α does not dominate β and every γ that dominates α dominates β. Thus, in (15) B c-commands C, F, G; C c-commands B, D, E; D c-commands E and conversely; F c-commands G and conversely. Where γ is restricted to maximal projections, we say that α *m-commands* β.

A second fundamental concept is *government* (Chomsky 1981a, 1986a, Rizzi 1990), a more "local" variety of command to which we return in section 1.4.1.

Given the language L, each SD is a sequence (π, λ, δ, σ), these being phrase markers drawn from the levels PF, LF, D-Structure, and S-Structure, respectively. The element δ reflects properties of the items selected from the lexicon as these are interpreted by the principles of UG, with the parameters fixed for L. The elements σ and λ are formed by successive application of operations of the computational system to δ; they will have the properties of δ, as modified by these operations. The PF representation π is a string of phonetic primes with syllabic and intonational structure indicated, derived by a computation from σ. We assume that the primes themselves are not modified in the course of the derivation from δ to λ.

A typical lexical entry consists of a phonological matrix and other features, among them the categorial features N, V, and so on; and in the case of Ns, Case and agreement features (person, number, gender), henceforth *φ-features*. In principle, any of these features may be lacking. In one case of particular interest, the entire phonological matrix is lacking. In this case the element is an EC (empty category). Among these ECs we have the elements *e* of (16), (17); we use * to indicate severe deviance, ? a weaker variety.

(16) a. John expected [*e* to hurt himself]
b. it is common [*e* to hurt oneself]

(17) **e* arrived yesterday ("he arrived yesterday")

We refer to the EC of (16) as *PRO*, an element that can be *controlled* by its *antecedent* (*John*, in (16a)) or can be arbitrary in interpretation, as in (16b). Possibly the latter is also a case of control by an EC occupying the same position as *for us* in (18) (Epstein 1984).

(18) it is convenient for us [for others to do the hard work]

If so, PRO is always controlled. See section 1.4.2.

The EC of (17) is a pronominal element, henceforth *pro*. It is not permitted in this position in English; the counterpart would be grammatical in Italian, a *null subject language*. On factors relevant to fixing the parameters, see Rizzi 1982, 1986a, Huang 1984, Borer 1984, Jaeggli and Safir 1989. This EC acts much in the manner of an ordinary pronoun, having reference fixed by context or by some antecedent in an appropriate position. The structural relations of (antecedent, *pro*) pairs are, furthermore, generally like those of (antecedent, pronoun) and unlike those of control. For example, in a null subject language we find the equivalents of (19a–b), analogous to the pair (19c–d) (*John* taken to be the antecedent of *pro*, *he*).

(19) a. the people that *pro* taught admired John
b. **pro* admired John
c. the people that he taught admired John
d. *he admired John

The behavior of *pro* and *he* is similar, while PRO can never appear in these positions.

A third type of EC, not drawn from the lexicon but created in the course of a derivation, is illustrated in (20).

(20) a. I wonder [who John expected [*e* to hurt himself]]
b. John was expected [*e* to hurt himself]

We refer to this EC as *trace* (*t*), a relational notion *trace-of* X, where X is the moved element serving as the antecedent *binding* the trace. Thus, *John* binds *e* in (20b) much as *e* binds the reflexive or as *they* binds the reciprocal in (21), in turn binding the reflexive.

(21) they expected [each other to hurt themselves]

In (20a) *e* is the trace of the NP *who*. The trace functions as a variable bound by *who*, understood as a restricted quantifier: 'for which *e*, *e* a person'. Here, *e* in turn binds *himself*, just as *each other* binds *themselves* in (21) and *Bill* binds *himself* in (22), with *Bill* substituting for the variable of (20a).

(22) John expected [Bill to hurt himself]

In (20a) both *e* and *himself* function as variables bound by the restricted quantifier, so that the LF form would be interpreted 'I wonder [for which *e*, *e* a person, John expected *e* to hurt *e*]'. Note that we are using the term *bind* here to cover the association of an antecedent with its trace quite generally, including the case of the (syntactic) binding of a variable by a quantifier-like element; and we also use the term, at LF, in the sense of quantifier-variable binding.

In (20b) the verb *was* is composed of the lexical element *be* and the inflectional elements [past, 3 person, singular]. Assume now that the process of composition adjoins the copula to the inflectional elements (raising). Recall that there are two interpretations of this process: (1) raising of *be* to the inflection position of the sentence to construct the combined form [*be* + inflections], or (2) raising of [*be* + inflections] (= *was*, drawn from the lexicon with its features already assigned) to the inflection position, where the features are checked. Either way, we have a second trace in (20b) = (23).

(23) John was e_2 expected [e_1 to hurt himself]

The EC e_2 is the trace of *be* or *was*; e_1 is the trace of *John*, binding *himself*. In each case the trace occupies the position from which its antecedent was moved. For concreteness, we adopt the checking theory (2), so that we have *was* raising in (23).

Raising of *was* to the inflection position is necessary to check inflectional properties. The same is true of the other inflected verbs, for example, *wonder* in (20a), which is [present, 1 person, singular]. Thus, a fuller (though still only partial) representation would be (24), where e_1 is the trace of *wonder*.

(24) I wonder e_1 [who John expected [e_2 to hurt himself]]

There is reason to believe that in English (24) is an LF representation, while the counterpart in other similar languages (e.g., French) is an S-Structure representation; (23) and its counterparts are S-Structure representations in both kinds of language (Emonds 1978, Pollock 1989).

Thus, English auxiliaries raise at S-Structure but main verbs raise only at LF, while the corresponding French elements all raise at S-Structure. English and French would then be identical in relevant respects at D-Structure and LF, while differing at S-Structure, with English (25a) (corresponding to the basically shared D-Structure) versus French (25b) (corresponding to the basically shared LF form).

(25) a. John often [kisses Mary]
b. Jean embrasse souvent [*t* Marie]
Jean kisses often Marie

Informally, the trace functions throughout as if the antecedent were in that position, receiving and assigning syntactic and semantic properties. Thus, *e* is in the normal position of the antecedent of a reflexive in both (20a) and (20b). And in (25b), the trace is the verbal head of VP, assigning a particular semantic role and grammatical Case to its nominal object.

Note that PRO and trace are quite different in their properties. Thus, an element that controls PRO is an independent argument in the sense of section 1.2, assigned an independent semantic role; but an element that binds a trace is not. Compare (16a) and (20b), repeated here:

(26) a. John expected [*e* to hurt himself]
b. John was expected [*e* to hurt himself]

In (26a) *John* is the subject argument of *expected*, exactly as in (22); the EC controlled by *John* has its independent function as subject of *hurt*. In (26b), in contrast, *John* has no semantic role other than what it "inherits" from its trace, as subject of *hurt*. Since the subject of *is expected* is assigned no independent argument role, it can be a nonargument (an *expletive*), as in (27).

(27) there is expected [to be an eclipse tomorrow]

Other differences of interpretation follow. Compare, for example, (28a) and (28b).

(28) a. your friends hoped [*e* to finish the meeting happy]
b. your friends seemed [*e* to finish the meeting happy]

In (28a) *your friends* and *e* are independent arguments, assigned their semantic roles as subjects of *hope* and *finish*, respectively; therefore, the EC must be PRO, controlled by *your friends*. But *seem* assigns no semantic role to its subject, which can again be an expletive, as in (29).

(29) a. it seems [your friends finished the meeting happy]
b. there seems [*e* to be a mistake in your argument]

Accordingly, the EC in (28b) must be trace, with its antecedent *your friends* receiving its semantic role as an argument as if it were in that position. We know further that the adjective *happy* modifies the subject of its own clause, not that of a higher clause. Thus, in (30) *happy* modifies *meeting*, not *your friends*; the sentence means that your friends hoped that the atmosphere would be happy when the meeting ends.

(30) your friends hoped [the meeting would finish happy]

In (28), then, *happy* modifies PRO in (a) and trace in (b). Example (28a) thus means that your friends had a certain wish: that they would be happy as the meeting ends. But (28b) has roughly the meaning of (29a), with *happy* modifying *your friends*.

Other differences of meaning also appear, as in (31a) and (31b) (Burzio 1986).

(31) a. one translator each was expected *t'* to be assigned *t* to the visiting diplomats
b. one translator each hoped PRO to be assigned *t* to the visiting diplomats

In (31a) neither *one translator each* nor its trace *t'* is in a position with independent argument status. Therefore, the argument phrase *one translator each* is interpreted as if it were in the position of the trace *t*, with the argument status of object of *assigned*; the meaning is that it was expected that one translator each would be assigned to the visiting diplomats (i.e., each diplomat would be assigned one translator). In (31b), in contrast, *one translator each* and PRO are independent arguments; it is PRO, not *one translator each*, that binds *t* and is interpreted as if it were in that position. The subject *one translator each* is thus left without an interpretation, very much as it is in the similar construction (32).

(32) one translator each hoped that he would be assigned to the visiting diplomats

Although the argument status of the antecedent of a trace is determined in the position of the trace, the antecedent may still have an independent semantic role in other respects. Compare the examples of (33).

(33) a. *it seems to each other [that your friends are happy]
b. your friends seem to each other [*t* to be happy]
c. it seems [that all your friends have not yet arrived]
d. all your friends seem [to have not yet arrived]

In (33a) *your friends* cannot bind the reciprocal *each other*, but it can in (33b), thus functioning in its overt position, not that of its trace. In (33c) and (33d) the overt positions are relevant for determining scopal properties: thus, only (33c) can mean that it seems that not all your friends have arrived, with *not* taking scope over *all*. We see, then, that scopal properties and argument status are determined in different ways for antecedent-trace constructions. Such facts as these ought to fall out as consequences of the theory of ECs and semantic interpretation. See section 1.4.2.

PRO and trace also differ in their syntactic distribution. Thus, in (34) we see the properties of control, with the antecedent and PRO functioning as independent arguments; but the properties of trace, with only one argument, cannot be exhibited in the analogous structures, as (35) illustrates.

(34) a. John asked whether [PRO to leave]
b. John expected that it would be fun [PRO to visit London]

(35) a. *John was asked whether [*t* to leave]
b. *John was expected that it would be fun [*t* to visit London]

In fact, trace and PRO do not overlap in their distribution; the facts should, again, fall out of the theory of ECs.

We also allow a fourth type of EC, one that has only the categorial features [±N, ±V], projecting in the usual way. They serve only as targets for movement, to be filled in the course of derivation. Since these elements have no semantic role, they will not satisfy the condition FI at D-Structure (as we will sharpen this below), and we may tentatively assume that they and the structures projected from them are inserted in the course of derivation, in a manner permitted by the theory of phrase structure. See section 1.4.3 for further comment.

If these kinds of EC are indeed distinct, then we expect them to differ in feature composition (Chomsky 1982, Lasnik 1989). Optimally, the features should be just those that distinguish overt elements. As a first approximation, suppose that overt NPs fall into the categories *anaphor* (reflexives, reciprocals), *pronoun*, and *r-expression* (*John*, *the rational square root of 2*, and other expressions that are "quasi-referential" in the internalist sense of section 1.1). We might assume, then, that we have two two-valued features, [anaphor] and [pronominal], with potentially four categories.

(36) a. [+anaphor, −pronominal]
b. [−anaphor, +pronominal]
c. [−anaphor, −pronominal]
d. [+anaphor, +pronominal]

An expression that is [+anaphor] functions referentially only in interaction with its antecedent; the reference of an expression that is [+pronominal] may be determined by an antecedent (but it does refer). Reflexives and reciprocals thus fall into category (36a) and pronouns into category (36b). The third category contains elements that refer but are not referentially dependent: r-expressions. The four ECs discussed above would have the typology of (37).

(37) a. Trace of NP is [+anaphor, −pronominal].
b. *Pro* is [−anaphor, +pronominal].
c. Trace of operator (variable) is [−anaphor, −pronominal].
d. PRO is [+anaphor, +pronominal].

Thus, trace of NP is nonreferential, *pro* has the properties of pronouns, and variables are "referential" in that they are placeholders for r-expressions. Controlled PRO falls into category (37d), hence all PRO if apparent uncontrolled PRO actually has a hidden controller (see (18)). We would expect, then, that trace of NP, *pro*, and variable would share relevant properties of overt anaphors, pronouns, and r-expressions, respectively. Such elements as English *one*, French *on*, German *man* might be partial overt counterparts to PRO, sharing the modal interpretation of arbitrary PRO and its restriction to subject position (Chomsky 1986b).

These expectations are largely satisfied, when we abstract away from other factors. Thus, the structural relation of a trace to its antecedent is basically that of an anaphor to its antecedent; in both cases the antecedent must c-command the trace, and other structural conditions must be met, as illustrated in (38), with the examples kept slightly different to avoid factors that bar the unwanted structures.

(38) a. i. John hurt himself
ii. John was hurt *t*
b. i. *himself thought [John seems to be intelligent]
ii. **t* thought [John seems that it is raining]
c. i. *John decided [himself left early]
ii. *John was decided [*t* to leave early]

These properties sharply restrict the options for movement of NPs: raising not lowering, object-to-subject but not conversely, and so on (Fiengo 1977).

Similar but not quite identical conditions hold of PRO. Thus, the C-Command Condition is illustrated by (39).

(39) a. John expects [PRO to hurt himself]
b. *[John's mother] expects [PRO to hurt himself]
c. *John expects [PRO to tell [Mary's brother] about herself]

In (39c) PRO is in a position to bind *herself* but the C-Command Condition requires that its antecedent be *John*, not *Mary*.

Similarly, variables share relevant properties of r-expressions, as expected.

(40) a. i. They think [John will leave tomorrow]
ii. I wonder who they think [*t* will leave tomorrow]
b. i. *it seems [John to be intelligent]
ii. *I wonder who it seems [*t* to be intelligent]
c. i. he thinks [John is intelligent]
ii. I wonder who [he thinks [*t* is intelligent]]
iii. John thinks [he is intelligent]
iv. I wonder who [*t* thinks [he is intelligent]]

In (40a) the name and the variable appear as Case-marked subjects of finite clauses, and the expressions are well formed, satisfying the Case-marking condition on r-expressions, to which we return directly. In (40b) the name and the variable appear as subjects of infinitives lacking Case, and the expressions are severely deviant. In (40ci) *he* is not referentially bound by *John* (we cannot take *he* to refer to John, as we may in (40ciii)); and in the parallel structure (40cii) *he* and the variable *t* are unrelated referentially (we cannot take *he* to be a variable bound by the operator *who*, which binds *t*, as we may in (40civ)). Again, many conditions on movement fall out as special cases.

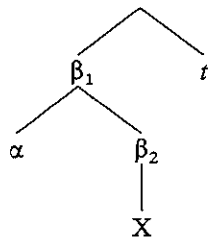
These ECs also have other features of overt expressions, specifically, ϕ-features. Thus, the trace in (20a) has the features [masculine, singular]; hence the choice of overt anaphor.

An EC lacking the typological features of (37) or ϕ-features is uninterpretable, hence impermissible at LF by the principle FI. Such an element, identified only by its categorial features (NP, V, etc.), may appear in the course of a derivation, but only as a position to be filled or otherwise eliminated.

It is an open question whether movement always leaves a trace, and whether, when it does, there are independent reasons for this. For the purposes of exposition, we tentatively assume that movement of an element α always leaves a trace and, in the simplest case, forms a *chain* (α, t), where α, the *head* of the chain, is the moved element and t is its trace. The chain is an *X-chain* if its head has the property X; we return to relevant choices of X. The elements subject to interpretation at the interface level LF are chains (sometimes one-membered), each an abstract representation of the head of the chain.

The movement operation (henceforth Move α) is an invariant principle of computation, stating that a category can be moved to a target position. We take the moved category and the target to be primes (lexical items, EC targets for movement, or projections from these minimal elements), with two options: either the moved category α replaces the target β (substitution), or it adjoins to it (adjunction), as in (41) (order irrelevant, t the trace of α, β_1 and β_2 two occurrences of β).

(41)



Any further constraints on movement will be derivative from other principles, including conditions on representations.

There are two natural interpretations of the elements formed by adjunction: we might assume that each occurrence of β in (41) is a category in its own right (Lasnik and Saito 1992) or that together they form a single category $[\beta_1, \beta_2]$ with the two occurrences of β as its *segments*

(May 1985, Chomsky 1986a). Empirical differences follow, as usual, as further theoretical structure is articulated.

The segment-category distinction requires a sharpening of the concepts of dominance and those derived from it (command, etc.). Let us say that the category $[\beta_1, \beta_2]$ in (41) *includes* X, *excludes* t, and *contains* α (and whatever is dominated by these elements). We restrict domination to inclusion. Thus, $[\beta_1, \beta_2]$ dominates only X. We say that a segment or category α *covers* β if it contains β, includes β, or $= \beta$. Defining the command relations as before, α c-commands t in (41), since it is not dominated (only contained) by β; but Y included in α does not. We carry over the properties of head and command to the postadjunction structure. Thus, if γ was the head of the preadjunction category β and c-commanded δ, then in the postadjunction structure $[\beta_1, \beta_2]$, γ remains the head and c-commands δ. Where no confusion arises, we will refer to the postadjunction category $[\beta_1, \beta_2]$ simply as β.

Substitution is constrained by a UG principle of *recoverability of deletion*, which requires that no information be lost by the operation; thus, α may substitute for β only if there is no feature conflict between them. The target of substitution will therefore always be an EC with the same categorial features as the moved category (the structure-preserving hypothesis of Emonds 1976). A similar property holds for adjunction, it appears (see section 1.3.3).

Move α permits multiple (*successive-cyclic*) movement, as in (42), derived from the D-Structures (43), with the targets of movement inserted.

(42) a. John seems [t' to have been expected [t to leave]]
b. I wonder [who John thought [t' Bill expected [t to leave]]]

(43) a. *e* seems [*e* to have been expected [John to leave]]
b. I wonder [*e* John thought [*e* Bill expected [*who* to leave]]]

In (42a) we have the chain (*John*, t', t) with the *links* (*John*, t') and (t', t); in (42b) the chain (*who*, t', t), also with two links. The *heads* of the chains are *John*, *who*, respectively.

We have so far assumed that the operation Move α forms a single link of a chain. Alternatively, we might assume that the operation is not Move α but rather Form Chain, an operation that forms the full chains of (42) from the D-Structures (43) in a single step. Within a richer theoretical context, the distinction may be more than merely notational (see chapter 3). We tentatively assume the more conventional Move α interpretation. The operation Move α satisfies narrow locality conditions.

Suppose that the position of the intermediate trace t in (42) is filled, as in (44), so that the chain must be formed with a single link, skipping the blocked position (occupied by *it*, *whether*, *whether*, respectively).

(44) a. *John seems that [*it* was expected [t to leave]]
b. ?what did John remember [*whether* Bill fixed t]]
c. *how did John remember [*whether* Bill fixed the car t]]

The chains (*John*, t), (*what*, t), (*how*, t) violate the locality conditions, and the expressions are deviant, though in strikingly different ways, facts that demand explanation in terms of properties of UG. Note that in case (44c) it is the PF form with *this interpretation*—that is, with *how* construed in the position of the trace—that is deviant; if *how* is construed with *remember*, there is no deviance. The single PF form has two distinct SDs, one sharply deviant, the other not.

Recall that each element must have a uniform, language-independent interpretation at the interface level LF (the principle FI). Some elements are arguments assigned specific semantic roles (θ-roles), such as agent and goal (see section 1.2); overt anaphors, PRO, and r-expressions (including variables) are all arguments. Expletives (e.g., English *there*, Italian *ci*) are assigned no θ-roles. Some elements (e.g., English *it*, French *il*, Italian *pro*) may ambiguously serve as arguments or expletives. By FI, expletives must be somehow removed at LF (section 1.3.3).

An argument must receive a θ-role from a head (θ-marking). An argument may also receive a semantic role (whether to be considered a θ-role or not is a theory-internal question that we put aside) by predication by an XP (see Williams 1980), possibly an open sentence (e.g., the relative clause of (45), with a variable position t).

(45) the job was offered to Mary, [who everyone agreed t had the best qualifications]

Other XPs (*adjuncts*, such as adverbial phrases) assign a semantic role to a predicate, a head, or another adjunct. As illustrated in (44b–c), movement of adjuncts and arguments has quite different properties (Huang 1982, Kayne 1984, Lasnik and Saito 1984, 1992, Aoun 1986, Rizzi 1990, Cinque 1990). A *θ-position* is a position to which a θ-role is assigned. The elements receiving interpretation at LF are chains. Hence, each argument chain (46) must contain at least one θ-position.

(46) $(\alpha_1, \ldots, \alpha_n)$

Furthermore, α_n, the position occupied by α_1 at D-Structure, must be a θ-position. The reason lies in the interpretation of D-Structure as a grammatical realization of lexical properties. Accordingly, θ-marking must take place at D-Structure: an element, moved or not, will have at LF exactly the θ-marking properties (assigning and receiving θ-roles) that it has at D-Structure. From the same consideration, it follows that nothing can move into a θ-position, gaining a θ-role that was not assigned to it at D-Structure. Thus, a chain can have no more than one θ-position, though any number of semantic roles may be assigned in this position. In (47), for example, *the wall* receives a semantic role from both *paint* and *red*.

(47) we painted the wall red

The theory of Case (section 1.4.3) requires that every argument have abstract Case (possibly realized overtly in one or another way, depending on specific morphological properties of the language). Hence, an argument chain (46) must have one and only one θ-position (namely, α_n) and at least one position in which Case is assigned (a *Case position*). Following Joseph Aoun, we might think of the function of Case as to make an argument chain *visible* for θ-marking. The Last Resort condition on movement (see section 1.1) requires that movement is permitted only to satisfy some condition, in particular, to satisfy visibility (hence, FI). Once an element has moved to a Case position, it can move no further, all relevant conditions now being satisfied. It follows, then, that every argument chain must be headed by a Case position and must terminate in a θ-position (the *Chain Condition*).

Note that these conclusions hold only for arguments other than PRO, an anomaly to which we return in section 1.4.3. On the status of chains headed by expletives with regard to the Chain Condition, see section 1.3.3.

We have so far considered chains that originate from an NP argument position of D-Structure. These fall into the two types illustrated in (42), repeated here.

(48) a. John seems [*t'* to have been expected [*t* to leave]]
b. I wonder [who John thought [*t'* Bill expected [*t* to leave]]]

In (48a) we have, among others, the argument chain (*John*, *t'*, *t*) and in (48b) the operator-variable chain (*who*, *t'*, *t*).

Chains may also originate from non-NP positions. One case, already mentioned, is the movement of a lexical category (*head movement*), as in

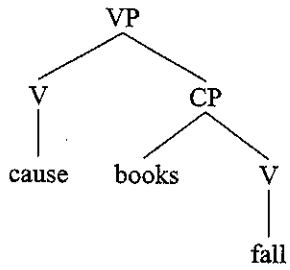
(23), (24), repeated here, illustrating the raising of V to the inflectional positions.

(49) a. John was *t* expected to hurt himself
b. I wonder *t* who John expected to hurt himself

Here we have the chains (*was*, *t*) and (*wonder*, *t*), the latter an LF chain for English.

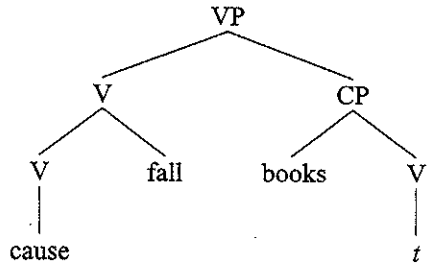
Head movement is also involved in formation of compound words in many languages. Suppose we were to form a causative verb meaning 'cause-to-fall' from the underlying D-Structure (50) by adjoining *fall* to *cause*.

(50)



This operation yields the structure (51), *t* the trace of *fall*.

(51)



See Baker 1988. Here *cause* is the head of a two-segment verbal category, if we assume a segment theory of adjunction.

A second kind of chain originating from a non-NP position arises from movement of nonarguments (adjuncts, predicates), as in (52).

(52) a. [to whose benefit] would that proposal be *t*
b. [how carefully] does he expect to fix the car *t*
c. [visit England], he never will *t*
d. [as successful as Mary], I don't think that John will ever be *t*

In each case the bracketed nonargument is the antecedent of the trace; the chains, then, are ([*to whose benefit*], *t*), ([*how carefully*], *t*), ([*visit England*], *t*), ([*as successful as Mary*], *t*), respectively. The questioned element in (52a) is really *who*; the rest is "carried along" because *who* cannot be extracted from the D-Structure position (53) ("pied-piping"; Ross 1967).

(53) that proposal would be [to who + POSSESSIVE benefit]

The natural interpretation reflects the D-Structure form; the meaning is 'for which person *x*, that proposal would be to *x*'s benefit'. There is evidence that the LF form should indeed be construed in something like this manner (see section 1.3.3). Case (52b) might be interpreted similarly; thus, the interpretation would be 'for what degree *x*, he expects to fix the car [*x* carefully]'. We might, then, argue that these are not really cases of movement of adjunct phrases as such, but rather of the question elements *who, how*, with the adjunct phrase carried along. We might conclude further that operator movement is the only kind of movement to which adjunct phrases are subject, unlike arguments, which can form argument chains. The conclusion is supported by the observation that although adjuncts can typically appear in many sentence positions, they are not interpreted as if they had moved from some more deeply embedded position (Saito 1985). Thus, (54a) is not given the interpretation of (54b), as it would be if *carefully* in (54a) had been moved from the D-Structure position of *carefully* in (54b).

(54) a. carefully, John told me to fix the car
b. John told me to [fix the car carefully]

This suggests that (52b) might also be regarded as a kind of pied-piping, with the moved element *how* carrying along the larger phrase *how carefully*. See chapters 3 and 4.

Within the theory of empty categories and chains, we can return to the question of directionality of interlevel relations raised in section 1.1. As noted there, such questions are obscure at best, and become even more subtle under the assumptions of trace theory. Consider again the S-Structure representations (42) derived from the D-Structure representations (43) (repeated here).

(55) a. John seems [*t'* to have been expected [t to leave]]
b. I wonder [who John thought [*t'* Bill expected [*t* to leave]]]

(56) a. *e* seems [*e* to have been expected [John to leave]]
b. I wonder [*e* John thought [*e* Bill expected [*who* to leave]]]

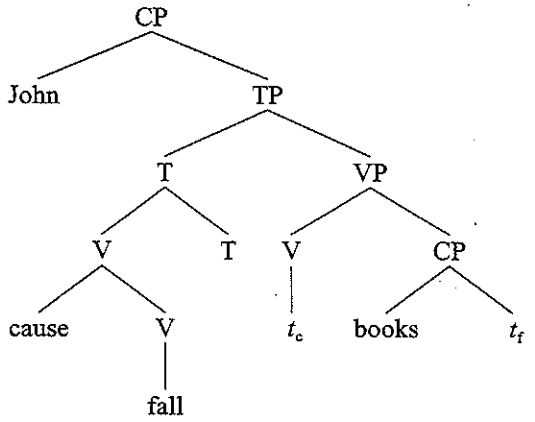
We now ask whether (55a–b) are derived from (56a–b), respectively, by movement of *John, who*; or whether D-Structure is derived from S-Structure by algorithm (Sportiche 1983, Rizzi 1986b), so that D-Structure is, in effect, a derived property of S-Structure; or whether there is simply a nondirectional relation between the paired expressions. These are alternative expressions of the relation between S-Structure and the lexicon. All three approaches are "transformational" in the abstract sense that they consider a relation between a "displaced element" and the position in which such an element is standardly interpreted; and in the case of (55b), the position in which it would be overt at S-Structure in languages of the Chinese-Japanese variety (see section 1.1). Such displacement relations are a fundamental feature of human language, which must be captured somehow. Apparent differences among alternative formulations often dissolve, on inquiry, to notational questions about how this property is expressed; similar questions arise with regard to apparent differences between "multilevel" approaches and "unilevel" alternatives that code global properties of phrase markers in complex symbols (Chomsky 1951, Harman 1963, Gazdar 1981). In the present case the empirical distinguishability of the approaches turns on highly theory-internal considerations. We will continue to adopt the derivational approach of section 1.1. We assume that this is, at root, a question of truth and falsity, though a subtle one.

To see some of the problems that arise, consider the locality conditions on Move α. A general condition, illustrated in (44), is that the target of movement must be the closest possible position, with varying effects depending on the kind of movement involved. The condition is very strict for head movement, which cannot pass over the closest c-commanding head (the *Head Movement Constraint* (HMC), a special case of more general principles; see section 1.4.1). Thus, in (57) formation of (b) from the D-Structure (a), raising *will* to the clause-initial position, satisfies the HMC; but raising of *read* to this position, crossing the possible target position occupied by *will*, violates the HMC, yielding the sharply deviant interrogative expression (57c).

(57) a. John will read the book
b. will John *t* read the book
c. *read John will *t* the book

But the locality relations expressed in the step-by-step computation might not be directly expressed at the output levels. That is, a derivation may satisfy the HMC in each step, but the output may appear to indicate that the condition is violated. Consider again the formation of a causative verb meaning 'cause-to-fall' by adjoining *fall* to *cause*, as in (51). Recall that a verb must also be raised to the inflection position. Hence, the newly formed category *cause-fall* must now raise to this position, forming the structure (58) (where TP is tense-headed phrase, t_f is the trace of *fall*, and t_c is the trace of *cause-fall*).

(58)



Here we have two chains: (*cause-fall*, t_c) and (*fall*, t_f). Each step of chain formation satisfies the strict locality condition. But the resulting chain headed by *fall* does not. In the S-Structure, the chain (*fall*, t_f) violates the HMC, because of the intervening head t_c, a possible target of movement that is "skipped" by the chain. The form should thus be as deviant as (57c), but it is well formed. The locality conditions are satisfied stepwise in the derivation, but are not satisfied by the output chain. Modifications required under nonderivational approaches are not entirely straightforward.

1.3.2 D-Structure

The computational system forms SDs that express the basic structural facts (syntactic, phonological, and semantic) of the language in the form of phrase markers with terminal strings drawn from the lexicon. We are assuming that such properties of natural language as "displaced elements" are expressed by multiple representational levels, each simple in form

and with simple operations such as Move α relating them. Each level captures certain systematic aspects of the full complexity. The relation of the computational system to the lexicon is expressed at the internal interface level of D-Structure. D-Structure is mapped to LF, the interface with conceptual and performance systems; at some point (S-Structure), perhaps varying somewhat from language to language, the derivation "branches" and an independent mapping (phonology) forms the PF representation that provides the interface with the sensorimotor systems. See (4).

The earliest attempts to develop generative grammar in the modern sense postulated a single level of syntactic representation, formed by rules of the form (59), where A is a single symbol and X, Y, Z are strings (X and Y possibly null), S is the designated initial symbol, and there is a set of designated terminal symbols that are then mapped by other rules to phonetic forms.

(59) $XAY \rightarrow XZY$

The symbols were assumed to be *complex*, consisting of two kinds of elements: categorial and structural. Categorial elements were NP, V, and so on. Structural elements were features that coded global properties of phrase markers; for example, NP-VP agreement in *the men are here* is coded by the [+plural] feature assigned to S and "inherited" by NP and VP through application of the rule [S, +plural] → [NP, +plural] [VP, +plural] (Chomsky 1951). Subsequent work "factored" the complexity into two components, restricting the symbols to just their categorial part (phrase structure rules forming phrase markers) and adding transformational rules to express global properties of expressions (Chomsky 1975a, Lees 1963, Matthews 1964, Klima 1964). A later step restricted the recursive part of the generative procedure to rules of the form (59) and separated the lexicon from the computational system (Chomsky 1965). This provided a two-level system: phrase structure rules and lexical insertion form D-Structure and transformations form the derived phrase markers of *surface structure*, then subjected to phonetic interpretation. The *Standard Theory* assumed further that only D-Structures are subjected to semantic interpretation, a position elaborated in *Generative Semantics* (Lakoff 1971). The *Extended Standard Theory* (EST) proposed that surface structure determines crucial elements of semantic interpretation (Jackendoff 1972, Chomsky 1972). Later work led to the four-level conception of EST outlined earlier, and the P&P

approach, which dispenses entirely with rule systems for particular languages and particular constructions.

Separation of the lexicon from the computational system permits simplification of the rules (59) to context-free, with X, Y null. Thus, instead of (59), we have the context-free rules (60).

(60) a. $A \rightarrow Z$
b. $B \rightarrow l$

Here A, B are nonterminal symbols, Z is a nonnull string of nonterminal symbols or grammatical formatives, and l is a position of lexical insertion. B is a nonbranching *lexical* category, and Z contains at most one lexical category. Z of (60a) is therefore as in either (61a) or (61b), where C_i is a nonlexical category, X and Y are strings of nonlexical categories, and L is a lexical category.

(61) a. $A \rightarrow C_1 \ldots C_n$
b. $A \rightarrow XLY$

These moves exposed the crucial redundancy in phrase structure rules already discussed (sections 1.1, 1.2): the form of Z in (60a) depends on inherent properties of lexical items. Further redundancies are also immediately apparent. In (60b) the properties of the lexical category B are completely determined by the lexical element inserted in l. Considering the possible forms in (61), we observe further that in (61b) the properties of A are determined by L: thus, if L is N, A is NP; if L is V, A is VP; and so on. The rule is *endocentric*, with the *head* L of the construction *projecting* the dominating category A. Suppose we assume that the rules (61a) are also endocentric, taking A to be a projection of one of the C_is (an expression of ideas developed in structural linguistics in terms of discovery procedures of constituent analysis (Harris 1951)). We now have rules of the form (62).

(62) a. $X^n \rightarrow ZX^mW$
b. $X^0 \rightarrow l$

Here n is typically $m + 1$ and X^i is some set of categorial features (see (14)); and X^0 is a lexical category. The element inserted in position l determines the features of X^i and, to a substantial extent, the choices of Z and W. At this point phrase structure rules are largely eliminated from particular languages; they are expressed as general properties of UG, within the framework of *X-bar theory*.

A further proposal restricts the rules (62a) to the forms (63).

(63) a. $X^n \rightarrow ZX^{n-1}$
b. $X^m \rightarrow X^mY$
c. $X^1 \rightarrow X^0W$

For n maximal, we use the conventional symbol XP for X^n; $n = 0$ is often dropped, where no confusion arises. To form a full phrase marker, each X^0 is replaced by a lexical element with the categorial features of X.

Suppose that $n = 2$ and $m = 1$ or 2 in (63), so that the possible rule forms are (64).

(64) a. $X^2 \rightarrow X^2Y$
b. $X^2 \rightarrow ZX^1$
c. $X^1 \rightarrow X^1Y$
d. $X^1 \rightarrow X^0W$

The nonterminal elements are X^1, X^2 (conventionally, X′, X″, or $\overline{X}$, $\overline{\overline{X}}$), $X^2 = XP$. Assume further that Z, Y are single symbols. We call Z the *specifier* (Spec) of X^2, the elements of W the *complements* of X^0, and Y in (64a) an *adjunct* of X^2. The status of Y in (64c) is ambiguous, depending on further articulation of the theory; let us tentatively classify it as an adjunct. Note that the notions specifier, complement, and adjunct are functional (relational), not categorial; thus, there is no categorial symbol *Spec*, but rather a relation specifier-of, and so on.

This is essentially the system of Chomsky 1981a, and the basis for further concepts defined there. We continue with these assumptions, turning later to modifications required under alternatives.

Muysken (1982) proposes that the bar levels are determined by the feature system [projected, maximal]. Thus, X^0 = [X, −projected, −maximal], X^1 = [X, +projected, −maximal]; X^2 = [X, +projected, +maximal]. Note that this approach permits a distinction between adjunction structures formed at D-Structure and by adjunction operations. See also Jackendoff 1977, Stowell 1981, Speas 1986, Fukui 1986, Baltin and Kroch 1989.

With the move to X-bar theory, the phrase structure system for a particular language is largely restricted to specification of the parameters that determine the ordering of head-complement, head-adjunct, and specifier-head. Choices above are typical for a *head-initial* language. The rules (62)–(64) themselves belong to UG (order aside), not to particular grammars. As discussed in sections 1.1 and 1.2, the elimination of

phrase structure rules has always been a plausible goal for linguistic theory, because of their redundancy with ineliminable lexical properties. If X-bar theory can be sustained in its most general form, choice of items from the lexicon will determine the D-Structure phrase markers for a language with parameters fixed.

Items of the lexicon are of two general types: with or without substantive content. We restrict the term *lexical* to the former category; the latter are *functional.* Each item is a feature set. Lexical elements head NP, VP, AP, and PP, and their subcategories (adverbial phrases, etc.). At D-Structure and LF, each such XP must play its appropriate semantic role, satisfying FI, as discussed earlier. The heads of these categories have (1) categorial features; (2) grammatical features such as ϕ-features and others checked in the course of derivations, continuing to assume one of the interpretations of morphological structure discussed in section 1.1; (3) a phonological matrix, further articulated by the mapping to PF; (4) inherent semantic and syntactic features that determine *s(emantic)-selection* and *c(ategorial)-selection*, respectively. Thus, *persuade* has features determining that it has an NP and a propositional complement, with their specific θ-roles. As discussed in section 1.2, c-selection is at least in part determined by s-selection; if the determination is complete, we can restrict attention to s-selection. We may now assume that a complement appears at D-Structure only in a θ-position, θ-marked by its head. Since the computational rules can add no further complements, it follows that at every level, complements are θ-positions, in fact, θ-marked the same way at each level (the *Projection Principle*). The Projection Principle and the related conditions on θ-marking provide a particular interpretation for the general condition FI at D-Structure and LF.

Functional items also have feature structure, but do not enter into θ-marking. Their presence or absence is determined by principles of UG, with some parameterization. Each functional element has certain selectional properties: it will take certain kinds of complements, and may or may not take a specifier. The specifiers typically (though perhaps not always) are targets for movement, in the sense discussed earlier. Hence, they have no independent semantic role at all. As suggested in section 1.3.1, we may assume them to be inserted in the course of derivation, unless some general condition on D-Structure requires their presence.

We assume that a full clause is headed by a complementizer C, hence is a CP, satisfying X-bar theory. C may have a specifier and must have a complement, a propositional phrase that we assume to be headed by another functional category I (inflection), which has the obligatory complement VP. Hence, a clause will typically have the form (65) (Bresnan 1972, Fassi Fehri 1980, Stowell 1981, Chomsky 1986a).

(65) $[_{CP}$ Spec $[_{C'}$ C $[_{IP}$ Spec $[_{I'}$ I VP]]]]

Specifiers are typically optional; we assume this is true of [Spec, CP]. The *Extended Projection Principle* (EPP) states that [Spec, IP] is obligatory, perhaps as a morphological property of I or by virtue of the predicational character of VP (Williams 1980, Rothstein 1983). The specifier of IP is the *subject of* IP; the nominal complement of VP is the *object of* VP. We take these to be functional rather than categorial notions; for different views, see Bresnan 1982, Perlmutter 1983. By the Projection Principle, the object is a θ-position. The subject may or may not be; it may be filled by an expletive or an argument at D-Structure. [Spec, IP] is therefore a potential θ-position. An actual or potential θ-position is an *A-position*; others are *$\overline{A}$-positions* (A-bar positions). As matters stand at this point, complement and subject ([Spec, IP]) are A-positions, and [Spec, CP] and adjunct positions are $\overline{A}$-positions. A chain headed by an element in an A-position is an *A-chain*; a chain headed by an element in an $\overline{A}$-position is an *$\overline{A}$-chain.* The distinction between A- and $\overline{A}$-positions, and between A- and $\overline{A}$-chains, plays a central role in the theory of movement and other modules of grammar. We return to some problems concerning these notions.

Recall the two interpretations of the syntactic rule *R* that associates lexical items with their inflectional features: word formation by adjunction, or checking (see section 1.1). If we adopt the former approach, it follows that the operation *R* must apply in the D- to S-structure derivation, because it "feeds" the rules of the phonological (PF) component. The checking alternative does not strictly imply that morphological properties must be determined by S-Structure, but we will assume that this is nevertheless true. It follows that the inflected head of VP must have its features assigned or checked by I at S-Structure, either through lowering of I to V or through raising of V to I (see sections 1.3.1, 1.3.3). In the lowering case the S-Structure chain is deficient. There must therefore be an LF operation that raises the adjunction structure [V–I] to replace the trace of the lowered I, voiding the potential violation and providing an LF similar to what we find in a language with raising at S-Structure (on some empirical consequences, see chapter 2). At LF, then, V will always be at least as high as I in (65).

The [V–I] complex may also raise further to C. In *V-second* languages such as Germanic generally, V raises to C and some other phrase raises to [Spec, CP] in the main clause (Den Besten 1989, Vikner 1990). The same phenomenon appears more marginally in English questions and some other constructions. We assume these to have the form illustrated in (66), *who* being in [Spec, CP], *has* raising to C and leaving the trace *t*, t_w being the trace of *who*.

(66) $[_{CP}$ who has $[_{IP}$ John t $[_{VP}$ met $t_w]]]$

By virtue of the general properties of X-bar theory, the only options in the pre-IP position, introducing a clause, are YP–X^0 or X^0; X^0 may be null and commonly must be in embedded clauses if [Spec, CP] is nonnull (the *Doubly Filled Comp Filter*; see Keyser 1975). We assume that in general, overt movement of the question words is to the [Spec, CP] position, and the same is true of other constructions.

Structures of the form (65) may also appear in embedded position, as in the indirect question (67a) or the declarative clauses (67b).

(67) a. (I wonder) $[_{CP}$ who C $[_{IP}$ John has met $t_w]]$
b. i. (I believe) $[_{CP}$ that $[_{IP}$ John has met Bill]]
ii. (I prefer) $[_{CP}$ for $[_{IP}$ John to meet Bill]]
iii. (it was decided) $[_{CP}$ C $[_{IP}$ PRO to meet Bill]]

In (67a) and (67biii) the C head of CP is null; in (67bi) it is *that*; and in (67bii) it is *for*. The head of IP is [+tense] in (67a), (67bi); it is [−tense] in (67bii–iii). [Spec, CP] is unfilled in (67b), but it can be realized in other embedded constructions, for example, (67a), the relative clause (68a), or the complex adjectival clause (68b), where there is good reason to believe that Op is an empty operator in [Spec, CP]. C is empty in both cases and *t* is the trace of Op.

(68) a. the man $[_{CP}$ Op C $[_{IP}$ John met $t]]$
b. Mary is too clever $[_{CP}$ Op C $[_{IP}$ PRO to catch $t]]$

The embedded clauses of (68) are predicates, open sentences with a variable position. In (68a) Op could be *who*, also semantically vacuous in this case. As a matter of (nontrivial) empirical fact, FI at LF includes the property of *strong binding*: every variable must have its range fixed by a restricted quantifier, or have its value determined by an antecedent. Since the operators in (68) are vacuous, the value of the variable must be fixed by the antecedents *man*, *Mary*, the choice being determined by locality conditions on predication.

These properties suffice to explain such examples as (3c), repeated here as (69a), the *if*-clause having the form (69b).

(69) a. if Mary is too clever to expect anyone to catch, then we don't expect anyone to catch Mary
b. Mary is too clever $[_{CP}$ Op C $[_{IP}$ PRO to expect [anyone to catch $t]]]$

The embedded CP is a typical case of long (successive-cyclic) movement, analogous to (70) with *who* in place of Op.

(70) (I wonder) [who he expected [them to catch t]]

The variable must not be bound by *anyone* or PRO in (69b), just as it must not be bound by the elements *them* or *he* in (70); we return to the operative principle of binding theory in sections 1.3.3, 1.4.2. By the strong binding condition, the variable must therefore have *Mary* as its antecedent. Furthermore, PRO must be arbitrary, for if it is bound by *Mary* (as in *Mary is too clever* [*PRO to catch Bill*]), then the variable will be bound by PRO, violating the principle just illustrated. We therefore have the interpretation (69a). Note that the account assumes crucially that binding is based upon an equivalence relation; see section 1.4.2.

On the same assumptions, we can reduce the problem of explaining the deviance of (71) to that of the deviance of overt operator movement, as in the analogous example of (72).

(71) a. *the man [you met people that caught t]
b. *Mary is too clever [to meet [people that caught t]]

(72) *who did John meet people that caught t

In all cases the locality conditions on movement are violated. See section 1.4.1.

We have assumed so far that embedded infinitivals are CPs, as in (67bii–iii) or (73).

(73) I wonder who he decided $[_{CP}$ C [PRO to catch $t]]$

In such cases the embedded subject must be PRO if the C head is empty and must be an overt NP if it is the Case-assigning element *for*, with dialectal variation. But there are other propositional phrases in which neither PRO nor the Case-assigning complementizer *for* can appear, for instance, (74).

(74) a. John believes [Bill to be intelligent]
b. John considers [Bill intelligent]
c. that gift made [Bill my friend for life]

Thus, in (74a) we cannot have *for Bill* or PRO instead of *Bill*. Similarly, in such constructions as these, the embedded subject can be trace, unlike the infinitival CPs. Compare:

(75) a. Bill is believed [*t* to be intelligent]
b. *Bill was decided [$_{CP}$ [*t* to be intelligent]]

In general, the embedded subject of (74) behaves very much as if it were an object of the verb of the main clause (the *matrix* verb), though it is not a θ-marked complement of the verb, but rather the subject of an embedded clause. Constructions of the form (74a) are rather idiosyncratic to English; in similar languages (e.g., German), the corresponding expressions have the properties of (67bii–iii), (73), and so on.

The embedded clause of (74a) contains I, hence IP; there is no evidence for any further structure. To account for the differences from the embedded CP infinitivals, we must assume either that the embedded clause is just IP, or that there is an EC complementizer that assigns Case, like *for* (Kayne 1984). On the former assumption, which we will pursue here, the embedded subject is governed by the matrix verb, a relation that suffices to assign Case, license trace, and bar PRO, as in verb-object constructions. Note that the question whether (75a) is a raising construction (like *John seems* [*t to be intelligent*]) or a passive construction (like *his claims were believed t*) does not arise, these concepts having been discarded as taxonomic artifacts (section 1.1). The construction is formed by Move α as a "last resort," the Case-assigning property of the verb having been "absorbed" by the passive morphology. In the examples of (74b–c) there is no overt functional head. Assuming the phrase boundaries indicated, either there is an EC I, or the embedded phrases are projections of their predicates, so-called *small clauses* (Stowell 1978, 1981). Either way, *Bill* is the subject of the embedded clause, behaving as in (74a) and unlike the subject of an embedded CP.

We have so far considered two functional categories: I and C. A natural extension is that just as propositions are projections of functional categories, so are the traditional noun phrases. The functional head in this case is D, a position filled by a determiner, a possessive agreement element, or a pronoun (Postal 1966a, Brame 1981, 1982, Abney 1987).

The phrases *that picture of Bill* and *John's picture of Bill* would therefore have the forms (76).

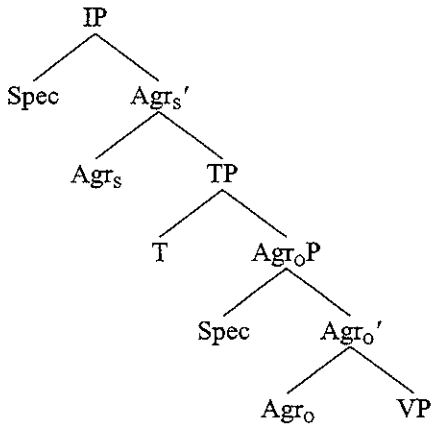
(76) a. [$_{DP}$ that [$_{NP}$ picture of Bill]]
b. [$_{DP}$ John Poss [$_{NP}$ picture of Bill]]

In (76a) [Spec, DP] is missing; in (76b) it is filled by the "subject" of the DP, *John*, to which the affix Poss is adjoined by a phonological operation. The D head is *that* in (76a) and Poss in (76b) (in some languages—for instance, Turkish—manifesting visible agreement with the subject; see Kornfilt 1985). Noun phrases in the informal sense are thus similar in internal structure to clauses (possibly even containing a "complementizer" position; Szabolcsi 1987). We might expect, then, to find N-raising to D, analogous to V-raising to I; see Longobardi 1994. There are numerous other consequences, which we cannot pursue here. We will use the informal notation Noun Phrase for DP or NP, unless confusion would arise.

We might ask whether these considerations generalize to other major categories, so that AP and VP are also complements of a functional element, even in V–VP or Modal–VP constructions. If so, a natural choice would be an element involved in Case assignment and agreement (call it *Agr*, a collection of φ-features). Such possibilities suggest a reconsideration of the functional element I, which has the strange property of being "double-headed" in the version of X-bar theory we are considering, assuming that T(ense) and Agr are independent heads. Following Pollock (1989), let us assume that T and Agr head separate maximal projections. Assuming that VP (and AP) is a complement of Agr, we now have the structure [Spec–T–Agr–VP] for the phrase we have called IP (now a term of convenience only), with T having AgrP as its complement, and VP, AP being complements of the Agr head of AgrP. Pollock argues on different grounds for the same structure: [Spec–T–Agr–VP]. In this structure the specifier of IP is not commanded (c- or m-commanded) by Agr, hence not governed by it. Hence, if (as we assume throughout) the operative relations among elements are based on such local relations, there would be no natural expression of subject-verb agreement. There is other evidence to suggest that the order should be Agr–T (Belletti 1990), where Agr is involved in subject agreement and nominative Case assignment. The proper reconciliation of these conflicting proposals may be that there are two Agr elements in IP, each a collection of φ-features, one involved in subject agreement and subject Case, the other in object agreement and object Case. Thus, the full structure will be (77), where

Agr$_S$ and *Agr*$_O$ are informal notations to distinguish the two functional roles of Agr, *Spec* indicates a functional role as before, and IP = AgrP.

(77)



Here we omit a possible [Spec, TP]. Embedded in this structure there might also be a phrase headed by the functional element *Negation*, or perhaps more broadly, a category that includes an affirmation marker and others as well (Pollock 1989, Laka 1990). We might proceed to assume that Case and agreement generally are manifestations of the Spec-head relation (Koopman 1987, Mahajan 1990; also see section 1.4.3 and chapters 2, 3).

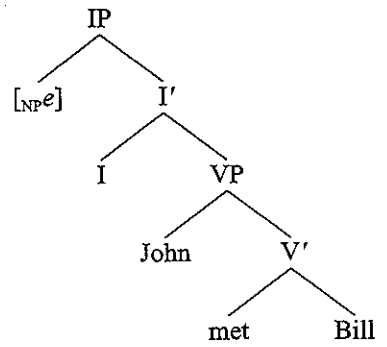
The status of [Spec, IP] is anomalous in several respects. One is that it may or may not be a θ-position, depending on lexical choices. Thus, in (78) the subject of *hurt* is a θ-position occupied by the trace of the argument *John*, taken to be the agent of *hurt*; but the subject of *seems* is a non-θ-position, which can also be occupied by the expletive *it*.

(78) a. John seems [*t* to have hurt himself]
b. it seems [that John has hurt himself]

[Spec, IP] is also the only position in which θ-role is not assigned within the m-command domain of a lexical head.

Such idiosyncratic properties would be eliminated if we were to assume that a thematic subject originates from a position internal to VP, then raising to [Spec, IP]. Collapsing the inflectional nodes to I for convenience, the D-Structure underlying *John met Bill* would then be (79) (Kitagawa 1986, Kuroda 1988, Sportiche 1988, Koopman and Sportiche 1991).

(79)



The subject and object are now θ-marked within the m-command domain of the verb *met*, within VP. On the present assumptions, *John* is [Spec, VP] and raises to [Spec, IP] to receive Case and produce a visible chain. By LF, *met* will have raised to I. If V raises to I at S-Structure and its subject raises to [Spec, IP] only at LF, we have a VSO language (at S-Structure). If the θ-role assigned to subject (the *external* θ-role, in the sense of Williams 1980) is in part compositionally determined (Marantz 1984), then these properties might be expressed internal to VP, as properties of the paired elements (subject, V').

The assumptions sketched out here provide a certain version of a "universal base hypothesis," a notion that has been explored from various points of view. If they are on the right track, typological variation should reduce to the ordering parameters and properties of functional elements. As discussed earlier, we expect that D-Structure and LF vary little in essential properties, D-Structure reflecting lexical properties through the mechanisms of X-bar theory and the parametric options for functional elements, and LF being the outcome of an invariant computational process that maps D-Structure to S-Structure and then to LF. A further proposal is that there is a uniform structural representation of θ-roles: thus, agent is typically associated with [Spec, VP], theme or patient with complement to V, and so on. This appears more plausible as evidence mounts questioning the existence of ergative languages at the level of θ-theory (Baker 1988, Johns 1987). See section 1.2.

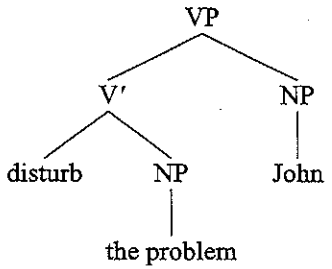
We have so far kept to the assumption of Chomsky 1981a that all internal θ-roles (all θ-roles apart from the role of subject) are assigned to sisters of the head. This assumption has repeatedly been questioned and has largely been abandoned. To mention a few cases, Kayne (1984) proposes that all branching is binary (yielding "unambiguous paths"). If so,

some internal θ-roles will be assigned to nonsisters. Kayne suggests, for example, that double-object verbs have the structure in (80), in which case *give* will θ-mark NPs properly contained within its complement.

(80) give [Mary books]

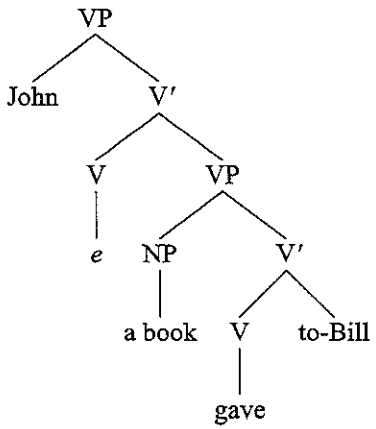
Similar ideas have been pursued in other studies as well. Belletti and Rizzi (1988) argue that the underlying structure of "psych-verb" constructions such as *the problem disturbed John* is (81), where the sister of *disturb* is assigned the θ-role theme (as usual), then raising to [Spec, IP], while the sister of V′ receives the θ-role experiencer (see also Pesetsky 1995, Bouchard 1991).

(81)



Larson proposes that double-object verbs such as *give* enter into D-Structures of the form (82) (Larson 1988, 1990; for an opposing view, see Jackendoff 1990a).

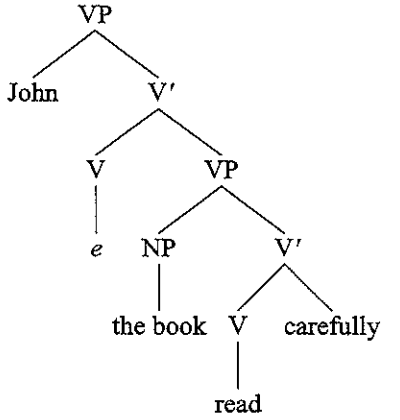
(82)



V raises to the empty main verb position of the higher *VP shell*, yielding *John gave a book to Bill*. Alternatively, operations similar to those yielding the passive construction could "absorb" the Case of *Bill*, forcing it to

raise to the subjectlike position of *a book*, which in turn becomes an adjunct, yielding *John gave Bill a book*. In (82) the direct object *a book*, though θ-marked as theme by the verb, is not its sister. Larson also indicates that adverbs are the innermost complements of V. Thus, the structure underlying *John read the book carefully* would be (83).

(83)



In this case the sister of the verb is an adverb that is not θ-marked at all, and the sole internal θ-role is assigned to a nonsister (*the book*).

With such modifications, the notion "θ-position" is still well defined, but "A-" and "$\overline{\text{A}}$-position" are not. These notions are formally quite different in character. A particular occurrence of a category in a phrase marker is, or is not, a θ-position, depending on whether it is θ-marked in that phrase marker. The notion "A-position," however, depends upon "potential θ-marking," which is to say that it presupposes an equivalence relation among phrase markers: an A-position is one that is θ-marked in the equivalent position of some member of the equivalence class. This is not an entirely straightforward notion, and with modifications of the sort just outlined, it becomes unspecifiable in any way that will bear the considerable theoretical burden that has been laid on the A versus $\overline{\text{A}}$ distinction, which enters crucially into large areas of current work.

The intuitive content of the distinction to be captured is reasonably clear. θ-positions and specifiers of inflectional elements share a range of structural properties; other non-θ-marked positions ([Spec, CP], elements adjoined to XP, non-θ-marked positions governed by a head) share a different range of structural properties. These are the former

A- and $\overline{A}$-positions, respectively. There are various proposals as to how to capture this distinction in terms of natural classes, and how to extend and sharpen it (e.g., for [Spec, DP]).

One approach (see chapter 3) is based on the observation that certain functional elements are, in effect, features of a head, in that they must be adjoined to this head to check its inherent features (alternatively, to assign these inherent features to it). Tense and the Agr elements are features of V in this sense, but C is not. Given a lexical head L, we say that a position is *L-related* if it is the specifier or complement of a feature of L. The L-related positions are the former A-positions, with the exception of non-θ-marked elements such as *carefully* in (83). But this exception will not be problematic if independent considerations block movement of such elements to any L-related position (raising). If economy considerations permit raising only when it is required (i.e., only Last Resort movement), then the issue will not arise; see sections 1.1, 1.3.1.

Along these lines, one might reconstruct something like the A versus $\overline{A}$ distinction. The account now relies on properties of occurrences of a category in a phrase marker, without reference to equivalence classes of phrase markers. Other uses of these notions, as in binding theory, appear to fall into place without too much difficulty. We leave the matter with these informal indications of a direction to explore, merely noting here that certain concepts that serve as foundations for much current work were originally defined on the basis of assumptions that have been widely abandoned and therefore must be reconstructed in some different way. With these qualifications, we will continue to use the notions with their intuitive content, as is standard in current technical work.

1.3.3 Derived Syntactic Representations

We have adopted the EST assumption that the derivations from D-Structure to PF and LF have a common part: D-Structure is mapped to S-Structure by Affect α, and the derivation then branches into two independent paths, one forming PF, the other forming LF (the *PF component* and the *LF component*, respectively). These are the two external interface levels. Since our concern here is syntax in the narrow sense, we restrict ourselves to the computation from D-Structure to LF.

The part of this derivation that maps S-Structure to LF is sometimes trivial, but whenever structural properties relevant to meaning are not already expressed at S-Structure, this mapping is substantive. Following Chomsky (1977), May (1977), we assume that scope of operators is

structurally represented at LF in terms of c-command. For interrogative operators, as will be discussed below, movement to an appropriate scope position takes place sometimes between D-Structure and S-Structure and sometimes between S-Structure and LF. Movement of quantifiers (May's "quantifier raising," QR) is generally an S-Structure to LF operation. The examples of "inversely linked" quantification discussed by May, as in (84), clearly indicate that S-Structure configuration does not suffice.

(84) everybody in some Italian city likes it.

Here *some Italian city* has wide scope, even though at S-Structure it is contained within the universally quantified NP. The correct interpretation is structurally represented in (85), with the entire subject NP having undergone QR, and the existential expression having raised still further.

(85) $[_{IP}$[some Italian city]$_i$ $[_{IP}$[everybody in t_i]$_j$ $[_{IP}$ t_j likes it]]]

See May 1977, 1985, for further motivation for QR.

Since it is an interface level, there are further requirements on LF. Given FI, every element of the LF representation of an expression must be subject to interpretation at the interface. As noted in section 1.1, this entails that there should be no true expletives in an LF representation. In such expressions as (86), then, the expletive element *there* must somehow be eliminated in the mapping from S-Structure to LF.

(86) there is a man in the room

One possibility that can be readily dismissed is that the expletive is simply deleted. The EPP demands that a clause have a subject at every syntactic level. Deletion of *there* would violate this requirement at LF. The expletive also appears to have φ-features that enter into agreement with the inflected verb. In (86) those features are [3 person, singular]; in (87) they are [3 person, plural].

(87) There are men in the room

A strong form of recoverability of deletion would presumably prevent the deletion of an element with φ-features. Given that *there* must be eliminated and cannot be deleted, the remaining possibility is that it is the target of a movement operation, with the *associate* of the expletive (*a man* in (86) and *men* in (87)) moving to the position of the expletive. Whether it is construed as substitution or adjunction, we may assume that this operation produces a new element combining the relevant features of the expletive and its associate: [*there*, *a man*] in (86), [*there*, *men*]

in (87). Let us call this an *amalgamated expletive*, leaving open its exact form.

We now have an account for the apparently anomalous rightward agreement in these cases, that is, the fact that the inflected verb agrees with the NP that follows it: *is* and *are* cannot be interchanged in (86), (87). The LF movement analysis directly predicts this paradigm. *There* must be replaced, but the phrase amalgamating with it must be nondistinct from it in features. If the operation is substitution, this requirement will follow from the recoverability condition. If the operation is adjunction, it will follow from a feature-matching requirement. Alternatively, we might assume that *there* lacks ϕ-features and that the overt agreement is an S-Structure reflex of agreement at the LF level between the inflected verb and the amalgamated expletive, its agreement features provided by the associate. Note further that one of the central properties of these constructions—that there *is* an argument associated with the expletive—also follows, since FI demands that the expletive be replaced.

From an S-Structure corresponding to (86), then, we derive the LF representation (88), *t* the trace of *a man*.

(88) [there, a man] is *t* in the room

Since the expletive occupies an A-position at S-Structure ([Spec, IP]), the LF movement forming the amalgamated expletive is A-movement. It follows that the relation between the associate and its trace meets the narrow conditions on A-movement. We now have an account for the fact that in the overt expression, the expletive and its associate conform to the locality requirements of A-chains. This follows from the fact that at LF, they are amalgamated to form an A-chain. We therefore have expletive-associate relations of the kind illustrated, but not those of (89), analogous to (90).

(89) a. **there* seems that *a man* is in the room
b. **there* seems that John saw *a man*
c. **there* was thought that [pictures of *a man* were on sale]

(90) a. **a man* seems that *t* is in the room
b. **a man* seems that John saw *t*
c. **a man* was thought that [pictures of *t* were on sale]

Note that the locality condition on the expletive-associate pair is that of A-movement, not binding, which is permissible in the analogue to (90c).

(91) *we* thought that [pictures of *each other* were on sale]

We return in section 1.4.3 to some problematic features of this analysis.

In section 1.3.1 we alluded to an approach to Case in terms of visibility for θ-marking. Expletives appear to contradict the principle, since they are not θ-marked but appear only in positions to which Case is assignable—in fact, only in a subset of such positions (subjects), but this follows from the fact that D-Structure complements are present only if they have a semantic role (typically, a θ-role). Thus, we find (92a) with nominative *there* and (92b) with accusative *there*, but (92c) is impossible.

(92) a. I believe [there is a man here]
b. I believe [there to be a man here]
c. *I tried [there to be a man here]

But now these facts fall neatly under the visibility approach. At LF we will have (93), where *t* is the trace of *a man* and EA is the amalgamated expletive.

(93) a. I believe [[$_{EA}$ *there*, *a man*] is *t* here]
b. I believe [[$_{EA}$ *there*, *a man*] to be *t* here]
c. *I tried [[$_{EA}$ *there*, *a man*] to be *t* here]

When an expletive is in a Caseless position at S-Structure, its associated argument will necessarily be in that position at LF and will, as a consequence, be invisible for θ-marking.

The analysis just sketched suggests that Case is checked at LF even though manifest at S-Structure; that is, it suggests that conditions requiring checking or assignment of Case are LF conditions, not S-Structure conditions, despite appearances. The same conclusion is suggested by the general approach to Case in terms of visibility, which links Case assignment to θ-theory. As discussed earlier, there is a preference on general conceptual grounds for interface conditions rather than S-Structure conditions. The various considerations so far adduced point in the same direction, but serious problems arise in trying to pursue this course. We return to the topic in section 1.4.3.

Turning to the S-Structure representation, with parameters fixed this is determined (presumably uniquely) by the choice of D-Structure and LF representations. S-Structure is unlike the three basic levels (D-Structure, PF, LF) in that it satisfies no constraints external to the computational system. It would therefore be reasonable to expect that conditions involving the interface (in particular, conditions bearing on the semantic

interpretation of SDs) should be restricted to the interface levels themselves, not applying at S-Structure. Nevertheless, there may be conditions of UG that must be satisfied at the S-Structure level.

There is some cross-linguistic variation in the character of S-Structure; in particular, functional elements vary in the ways they are articulated at S-Structure and hence are realized overtly. Languages may also differ, as noted, with regard to the placement of S-Structure in the derivation of LF from D-Structure, that is, the point of branching to PF. One well-studied case concerns the application of Move α that determines the scope of a question phrase (commonly called the "*wh*-phrase," by historical accident), moving it to the periphery of the proposition.

In English-type languages the effects of the movement operation are visible, yielding the S-Structure form (94), where *t* is the trace of *what*.

(94) a. what do you want [John to give *t* to Bill]
b. what do you want [John to give *t* to whom]

In a *multiple* question such as (94b), only one of the question phrases moves by S-Structure.

In the counterpart to (94a) in a Chinese-type language, the analogue to *what* is "in situ" at S-Structure, occupying the position of the trace in (94). We assume, following Huang 1982 and much subsequent work, that the phrase is moved to clause-peripheral position at LF, yielding an LF form resembling (94). More generally, in both types of language all question phrases will have moved to scopal position under this operation in the course of the derivation, within the LF component if not before (Higginbotham and May 1981, Aoun, Hornstein, and Sportiche 1981).

The D-Structure forms are therefore alike in relevant respects in English- and Chinese-type languages, as are the LF forms, the standard expectation (see section 1.1). But the S-Structure forms differ, depending on whether the operation that places the question phrase in the position that determines scope applies before or after the branching to the PF component at S Structure. One type of language (English, French, etc.) employs *overt movement* of a question phrase in the course of derivation of S-Structure from D-Structure, feeding the phonological component; another type (Chinese, Japanese, etc.) leaves all question phrases in situ at S-Structure. Both types of language employ *covert movement* within the LF component for any in-situ question phrase. A third type of language (e.g., Polish) has overt movement of all question phrases.

D-Structure and LF representations are again similar to the other two language types, but the S-Structures differ (Lasnik and Saito 1984).

Given a narrow theory of parametric variation of the sort discussed, these three language types should differ in properties of functional features. Cheng (1991) argues that mood (interrogative, declarative, etc.) must be indicated at S-Structure in the pre-IP position, hence by choice of either C or [Spec, CP]; the head of CP and its specifier thus serve as "force indicators" in something like the Fregean sense. If the lexicon contains an element Q (marking *yes-no* questions), then this element will suffice to identify an expression as an interrogative whether or not it contains an in-situ question phrase. There is no need, then, for the question phrase to raise to [Spec, CP] at S-Structure. Lacking the element Q, a language must employ overt movement of a question phrase to [Spec, CP] to be identified as an interrogative at S-Structure.

Suppose further that economy principles favor operations that do not feed the PF component over others that do; hence, if operations need not be overt to satisfy some condition, they will be assigned to the LF component, applying as "late" in the derivation as possible, at the point where they are forced by LF conditions (in the case under discussion, conditions of scope). These assumptions lead us to expect two basic categories of language in the simplest case: (1) languages with a Q element and the question phrase in situ (Chinese, Japanese); and (2) languages lacking a Q element and with a single question word in [Spec, CP] (English, German). At LF all question phrases will have moved, so that the quasi quantifier can be interpreted with its scope determined and a bound variable heading an argument chain. Other typological differences should then be reducible to internal morphology of the question phrase —for instance, languages of the Polish-Hungarian type with multiple fronting of question phrases at S-Structure (though perhaps not to [Spec, CP]; see Cheng 1991). On assumptions such as these, there are conditions that must be satisfied by S-Structure representations.

Overt and covert movement might have different properties. Huang (1982) proposed that the bounding conditions on overt movement are relaxed in the LF component so that we have such pairs as (95a) in English and (95b) in Chinese.

(95) a. *who do you like [books that criticize *t*]
b. ni xihuan [piping shei de shu]
you like [criticize who REL book]

Both expressions have the interpretation 'for which person x, you like books that criticize x', but only (95b) is well formed. The English example (95a) violates a locality condition on movement (*Subjacency*); its Chinese counterpart is free from this constraint (for varying approaches, see, among others, Huang 1982, Lasnik and Saito 1984, Nishigauchi 1986, Fiengo et al. 1988, Watanabe 1991).

A similar phenomenon is found in multiple questions in English-type languages. Thus, English (96a) is well formed with the interpretation (96b) expressed in the LF form (96c).

(96) a. who [t likes books that criticize whom]
b. for which persons y, x, [x likes books that criticize y]
c. [whom_j, who_i] [t_i likes books that criticize t_j]

We have assumed that overt movement, as in (94) or (96a), places the question phrase in the position [Spec, CP]. Possibly covert movement, not required for mood specification, may adjoin the question phrase to IP, treating it like a quantifier phrase assigned scope by QR. Typically, such question phrases as *who*, *whom* share semantic and distributional properties of quantifier phrases, and might be composed of an indefinite quantifier, a *wh*-feature, and the restriction on the quantifier (Chomsky 1964, Kuroda 1965, Nishigauchi 1986, Kim 1990, Watanabe 1991). Accordingly, *who* would be composed of [some x, *wh*-, x a person]; and so on. It would not then be surprising if such question phrases were to share properties of the indefinite quantifier, adjoining to IP in the LF component by QR, though it remains to explain why they move so freely, unlike QR, which is typically clause-bound.

In English-type languages, relative clauses are formed in much the same manner as interrogatives: an operator phrase, which may be either an EC operator Op or morphologically identical to a question phrase, is moved to [Spec, CP], leaving a trace that functions as a variable, as in (97).

(97) a. the people [who John expected to meet t]
b. the people [Op (that) John expected to meet t]

In either case, the relative clause is an open sentence functioning as a predicate (see (68)). In these constructions, movement is in the overt (pre-S-Structure) syntax, as shown in (97a), and satisfies the bounding conditions on overt movement, as illustrated in (98).

(98) a. *the man [who you like books that criticize t]
b. *the man [Op (that) you like books that criticize t]

While Chinese and Japanese have question words in situ, relative clauses show the properties of overt movement (Huang 1982, Watanabe 1991, Ishii 1991). These observations suggest that relative clauses require overt movement. The reason might be that predication must be established at S-Structure (Williams 1980). If so, we have another example of an S-Structure condition. It would remain to extend the analysis to languages that form relatives with in-situ pronouns (resumptive pronouns) and full NP heads in the position of the variable above (Sells 1984, Demirdache 1991).

These considerations extend to other constructions with EC operators, such as the complex adjectivals discussed in section 1.3.2 ((68)–(69)), with the locality properties of overt movement (repeated here).

(99) a. Mary is too clever [$_{CP}$ Op C [$_{IP}$ PRO to expect [anyone to catch t]]]
b. *Mary is too clever [$_{CP}$ Op C [$_{IP}$ PRO to meet [anyone who caught t]]]

Given the locality properties, the open sentences functioning as predicates must have been formed by overt movement, pre-S-Structure.

Some semantic properties of linguistic expressions appear to be determined by S-Structure configurations, independently of operations of the LF component. Let P be such a property. Then two accounts are possible.

(100) a. P holds at S-Structure.
b. P holds at LF under *reconstruction*, that is, with the moved phrase treated "as if" it were in the position of its trace.

If the former is correct, then the property P involves a condition on S-Structure. There are various ways of construing the notion of reconstruction.

A good deal of insight into these questions derives from the principle of binding theory—call it *Command*—stipulating that a pronoun cannot c-command its antecedent (see sections 1.3.2, 1.4.2). We can formulate this as a requirement that an r-expression α must be *A-free*, that is, not c-commanded by a pronoun in an A-position linked to α in the binding-theoretic sense. Thus, in (101a) and (101b) *John* is A-free; the pronoun (*him*, *his*) does not c-command *John* and can take *John* as its antecedent.

But in (101c) *he* c-commands *John* and must be assigned reference in some other way.

(101) a. John thought Mary took a picture of him
b. [his mother] thought Mary took a picture of John
c. he thought Mary took a picture of John

The principle Command applies to r-expressions generally, hence to variables as well as *John*, as we see in (102), analogous to (101), with the trace of *who* in the position of *John* in (101).

(102) a. the man who [*t* thought Mary took a picture of him]
b. the man who [[his mother] thought Mary took a picture of *t*]
c. the man who [he thought Mary took a picture of *t*]

In (102a) and (102b) the pronoun does not c-command *t*. Even if the pronoun and variable are referentially linked, the variable is A-free, though $\overline{\text{A}}$-bound by its operator. The variable and the pronoun can now be construed as variables bound ($\overline{\text{A}}$-bound) by *who*. The interpretations are 'the man *x* such that *x* thought Mary took a picture of *x*', 'the man *x* such that *x*'s mother thought Mary took a picture of *x*', respectively; the deviance of (102b), if any, is slight (Chomsky 1982, Higginbotham 1983, Lasnik and Stowell 1991).

But in (102c) *he* c-commands *t* and therefore cannot be linked to this variable or it will not be A-free; (102c) therefore cannot have the interpretation 'the man *x* such that *x* thought Mary took a picture of *x*'. There is nothing "wrong" with this interpretation; in fact, it is the interpretation of (102a). But it cannot be assigned to (102c), by virtue of Command (the property of *strong crossover*; Postal 1971, Wasow 1972, Lasnik 1976).

The principle Command also enters into the explanation of the meaning of the complex adjectivals of (99), as discussed earlier (see (68)–(69)).

We now ask at what level Command applies. Consider the examples (103).

(103) a. you said he liked [the pictures that John took]
b. [how many pictures that John took] did you say he liked *t*
c. who [*t* said he liked [how many pictures that John took]]

In (103a) *he* c-commands *John* and cannot take *John* as antecedent; in (103b) there is no c-command relation and *John* can be the antecedent of *he*. In the multiple-question-phrase construction (103c) *John* in fact cannot be the antecedent of *he*. It must be, then, that *he* c-commands *John* at the level of representation at which Command applies; the binding properties of (103c) are those of (103a), not (103b).

Returning to the two options of (100), we seem to be led here to adopt the first: that Command applies at S-Structure, before the bracketed question phrase is moved to preclausal position at LF, at which point (103c) would be formally similar to (103b), not (103a). Alternatively, we could assume, in the face of examples such as these, that the second option, reconstruction, holds for LF raising but not overt movement. More simply, we could dispense with both options, rejecting the tacit assumption that LF movement formed (104) from (103c), *t'* the trace of the LF-moved phrase.

(104) [[how many pictures that John took] who] [*t* said he liked *t'*]

Recalling that LF movement does not meet the strict locality conditions of S-Structure movement, we might reject the assumption that the entire NP is pied-piped when *how many* is raised to the scopal position, assuming rather that *how many* is extracted from the NP, yielding an LF form along the lines of (105), *t'* the trace of *how many*.

(105) [[how many] who] [*t* said he liked [*t'* pictures that John took]]

The answer, then, could be the pair (*12*, *Bill*), meaning that Bill said he liked 12 pictures that John took. But in the LF form (105), *he* c-commands *John* so that Command applies as in (103a). Pursuing such lines as these, we would not be led to adopt the assumption that Command applies at S-Structure, leaving us with the preferable option that conditions involving interpretation apply only at the interface levels. A further consequence would be that (103b–c) have somewhat different forms at LF; the empirical effect is unclear (Hornstein and Weinberg 1990).

Other constructions illustrate the process of reconstruction and are thus consistent with the restriction of the conditions on interpretation to the LF level. Consider (106).

(106) a. they said he admires John's father
b. who [*t* said he admires John's father]
c. (guess) whose father [they said he admires *t*]

In (106a) and (106b) *he* c-commands *John* and cannot take *John* as its antecedent, given Command. In (106b) *he* does not c-command *t*, so both

can be taken as variables bound by *who*, yielding the interpretation 'for which person x, x said x admires John's father'. In (106c) *he* does not c-command *who*, but it cannot be taken as a variable bound by *who*, even though this interpretation would leave t A-free. The complement of *guess* is interpreted as (107) with *he* unbound, analogous to (106a).

(107) for which person x [they said he admires x's father]

Thus, we have reconstruction: treatment of [*whose father*] as if the phrase were in the position of its trace t in (106c) (Chomsky 1977, Freidin and Lasnik 1981).

Questions proliferate quickly with further inquiry. Consider, for example, such constructions as (108), formed by successive-cyclic movement of the question phrase from the position of t, to the position of t', to [Spec, CP] of the matrix clause.

(108) a. [which picture of himself] did *John* say [t' that *Bill* liked t best]
b. [which pictures of each other] did *they* say [t' that *we* liked best]

Barss (1986) observes that the anaphor can take either of the italicized NPs as its antecedent. But an anaphor can only be bound by the closest c-commanding subject, as we see in the corresponding expressions (109), without *wh*-movement.

(109) a. John said [that Bill liked [that picture of himself] best]
b. they said [that we liked [those pictures of each other] best]

Here the antecedents must be *Bill*, *we*. In (108) the same binding condition requires that each of the traces be "visible," the question phrase being interpreted for binding as if it were in one or the other of these positions (*chain binding*).

Another problematic example is (110a), with the interpretation (110b) and, on our current assumptions, the LF representation (110c) (Higginbotham 1980, 1983).

(110) a. guess which picture of which boy [they said he admires t]
b. for which boy x, which picture y of x, [they said he admires y]
c. [[which boy]$_i$ [which picture of t_i]]$_j$ [they said he admires t_j]

Reconstruction in the manner of (106c) and (107) does not yield a structure barred by Command. Nevertheless, *he* cannot be construed as an occurrence of the bound variable x.

The formal property entering into reconstruction here seems to be that the pair (r-expression α, pronoun β) are referentially disconnected at LF

if there is a γ such that γ contains α and β c-commands γ or its trace. But that principle, applying at S-Structure, yields incorrect results for (103), barring binding of the pronoun in (103b). The discrepancy suggests that the problem with (110) lies elsewhere.

The problems are more general. Consider (111).

(111) a. the claim that John was asleep, he won't discuss t
b. the claim that John made, he won't discuss t

Case (111a) is analogous to (110); case (111b) to (103b). On our current assumptions, the pronoun must not take *John* as antecedent in (111a) or (111b); the conclusion is correct for (111a) but not for (111b). Still further complications arise when we consider differences between these examples of $\overline{\text{A}}$-movement and "scrambling" constructions in which the normal subject-object order is inverted.

We leave the topic in this unsettled state. For further discussion of these and related matters, from various points of view, see Lakoff 1968, Reinhart 1976, 1983, Van Riemsdijk and Williams 1981, Higginbotham 1980, 1983, Langendoen and Battistella 1982, Barss 1986, Freidin 1986, Lebeaux 1988, Saito 1989, and chapter 3.

Consideration of LF $\overline{\text{A}}$-movement also suggests that there is an S-Structure condition licensing *parasitic gap* (PG) constructions such as (112a), interpreted as (112b).

(112) a. which book did you file t [without my reading e first]
b. for which x, x a book, you filed x without my reading x first

Licensing of PGs by $\overline{\text{A}}$-chains is quite general, but those formed by LF movement do not license PGs, as illustrated in (113), with the S-Structure (113a) and the LF form (113b).

(113) a. *who [t filed which book [without my reading e]]
b. *[[which book]$_j$ who$_i$] [t_i filed t_j [without my reading e]]

The interpretation cannot be 'for which book x, who filed x without my reading x'. PG constructions, then, provide some evidence for the existence of S-Structure conditions.

The condition that licenses PGs must also account for the fact that these constructions are licensed by $\overline{\text{A}}$-chains but not A-chains. Thus, the A-chain (*the book*, t) of (114) does not license the PG e, unlike the $\overline{\text{A}}$-chain (*which book*, t) of (112a), with the same t-e relation.

(114) *the book was filed t [without my reading e first]

For further discussion, see Taraldsen 1981, Engdahl 1983, 1985, Chomsky 1982, 1986a, Kayne 1984, Longobardi 1985, Browning 1987, Cinque 1990.

Note that even the acceptable PGs are somewhat awkward; as in earlier cases discussed, we are interested in the relative deviance of various constructions, which is quite clear and demands explanation. The general literature on PGs regularly uses for illustration such pairs as (115), where the first is completely grammatical and the second sharply deviant, but these cases do not suffice to show that $\overline{A}$-chains license PGs while A-chains do not, because (115b) is ruled out for independent reasons of control theory, as illustrated in (116) (Lasnik and Uriagereka 1988).

(115) a. the book that you filed [without PRO reading *e*]
b. *the book that was filed [without PRO reading *e*]

(116) a. the book that you filed [without PRO thinking]
b. *the book that was filed [without PRO thinking]

The question of S-Structure conditions also arises in connection with elements lexically identified as affixes (e.g., pronominal clitics, verbal inflections, Case features). Since these properties are commonly overt at PF, they must be manifested at S-Structure (Lasnik 1981; we omit here the possibility that rules of the PF component might be rich enough to handle the phenomenon). As indicated earlier, the question becomes rather subtle if we assume the checking interpretation of inflectional features. Suppose again that English *walked* is inserted into D-Structure with the properties [walk], [past], the latter being checked and licensed by a syntactic rule *R* that joins [past] and *walked*. Suppose further that such functional elements as [tense] lack phonological matrices and are thus invisible at PF. We need not then assume that *R* is a lowering rule adjoining [past] to *walked*, to be reversed at LF; an alternative possibility is that the D- and S-Structures are alike, with *R* raising the verb to the inflectional position at LF, mirroring the process that is overt with auxiliaries and in French-type languages (for theory-internal arguments bearing on the matter, see chapters 2 and 3). The same question arises with regard to Case marking. Even if it is overt, the conceptual possibility remains that elements enter the computational system with their Case features already indicated, these being checked only at the LF level. Any apparent S-Structure requirement for Case would have to be satisfied in some other way. See section 1.4.3 and chapter 3.

Other theory-internal considerations suggest that empty categories must be licensed at S-Structure, in particular, traces in argument chains (Lasnik and Saito 1984, 1992; see section 1.4.1). If the relation of predication holding between an XP and its (syntactic) subject must satisfy S-Structure conditions, as suggested earlier, it is also natural (though not necessary) to suppose that licensing of an EC subject of predication should also take place at this level. Thus, according to Rizzi's theory, the null subject parameter reduces to properties of the system of the verbal inflection: in Italian, "strong" agreement (Agr) licenses *pro* subject; in French or English, the "weaker" Agr does not. We might expect, then, that this condition must be satisfied by the S-Structure configuration.

The plausibility of this assumption is enhanced by consideration of properties of expletive *pro*. Consider the D-Structures (117).

(117) a. *e* was stolen a book
b. *e* seems [*e'* to be a book missing]

In a null subject language, the expressions can surface in this form, with *e* being expletive *pro* and *e'* its trace; here *pro* is licensed by strong Agr. But in a non–null subject language, *e* must be replaced by S-Structure, either by an overt expletive or by raising of *a book* to fill this position, as in (118).

(118) a. i. ?there was stolen a book
ii. a book was stolen *t*
b. i. there seems [*t* to be a book missing]
ii. a book seems [*t* to be *t'* missing]

Some S-Structure property, it appears, must ensure that the options of (118) have been taken by the S-Structure level, not in the LF component. The problem becomes more severe if we adopt the strong version of FI that requires that expletives be replaced at LF (sections 1.3.1, 1.3.3). Then the S-Structure forms of (117) will appear at LF essentially as the (ii) forms of (118). It would follow, then, that the relevant distinctions must be established at S-Structure: *pro* is licensed at S-Structure, permitting (117) in Italian but not English. For alternative analyses, see chapters 3, 4.

It has also been proposed that some of the conditions that have been assumed to apply at LF actually apply within derivations from S-Structure to PF (Jaeggli 1980, Aoun et al. 1987). It cannot be that the conditions apply at the level of PF representation itself, because at

the interface level PF we have only phonetic features with no further relevant structure. The assumption would be, then, that these conditions apply either at S-Structure or at some level intermediate between S-Structure and PF.

We have assumed so far that X-bar theory applies at D-Structure, its properties being "carried over" to S-Structure and LF by the computational processes. Suppose that X-bar theory applies at S-Structure as well. Van Riemsdijk (1989) argues that on this assumption, movement need not be restricted to minimal and maximal phrases (X^0 and XP), as so far tacitly assumed. Movement of X′ ($=X^1$) could be allowed, to be followed by a process of "regeneration" that forms a proper X-bar structure at the S-Structure level in a minimal way. On this analysis, (119) would be derived by movement of the N′ category *Lösung*, followed by generation of *eine* to satisfy X-bar theory at S-Structure, *eine* being a "spelling out" of the ϕ-features of *Lösung*.

(119) [eine Lösung] hat er [eine bessere *t*] als ich
a solution has he a better (one) than I

If X-bar theory applies at S-Structure, Emonds's structure-preserving hypothesis for substitution (section 1.3.1) follows in essentials, since conflict of categorial features will violate X-bar-theoretic principles. A similar conclusion will also hold for adjunction. Suppose, for example, that an X^0 element is adjoined to the YP Z, forming (120).

(120) $[_{YP}\ X^0\ YP]]$

This structure violates X-bar theory, which requires that X^0 head an X′ structure. Adjunction of XP to YP, however, would yield a structure consistent with X-bar theory. Adjunction of X^0 to Y^0 yields a two-segment category $[Y^0, Y^0]$, with an internal structure "invisible" to X-bar theory. Pursuing this line of thinking, it may be possible to derive a version of the structure-preserving hypothesis for adjunction: essentially, the condition that a category can be adjoined only to a category of the same bar level.

1.4 Modules of Language

1.4.1 Government Theory

We have referred several times to the notion of *government*, a more "local" variety of command (section 1.3.1). We assume tentatively that the relevant notion of command is c-command. The concept of government has entered extensively into the study of the various modules of grammar. Hence, slight modifications in formulation have wide-ranging empirical consequences (see, among others, Aoun and Sportiche 1981, Chomsky 1981a, 1986a, Kayne 1984, Lasnik and Saito 1984, 1992, Rizzi 1990).

We say that α governs β if α c-commands β and there is no category γ that "protects" β from government by α. γ protects β in this sense if it is c-commanded by α and either (121a) or (121b) holds.

(121) a. γ is a *barrier* dominating β.
b. γ *intervenes* between α and β.

Government is *canonical* if the linear order of (α, β) accords with the value of the head parameter (Kayne 1984). We speak of "X-government" when the governor has the property X. There are two main categories of government to be considered: *antecedent government* of α by an antecedent of α, and *head government* of α by a head. We refer to these categories as *proper government*.

To make the concept of locality precise, we have to spell out the notions "barrier" and "intervene" in (121). Consider the two in turn.

We take a barrier to be an XP that is not a complement, putting aside now the ambiguous status of noncomplements to V under the various ramifications of Kayne's unambiguous path theory (section 1.3.2). Thus, in (122) the bracketed expressions are all XPs, but only those subscripted *B* are barriers for the elements they contain.

(122) a. I wonder which book [John told the students [that [they should read *t*]]]
b. ??I wonder which book [John met [someone $[_B$ who read *t*]]]
c. *I wonder how [John met [someone $[_B$ who [fixed the car *t*]]]]
d. ??I wonder which book [John left New York $[_B$ before he read *t*]]
e. *I wonder how [John left New York $[_B$ before he fixed the car *t*]]

In each case the trace indicates the position of extraction, under the intended interpretation: thus, (122e) asks how John fixed the car, not how he left New York. If we extract from within a barrier, the trace left behind will not be antecedent-governed; otherwise, it will be. When extraction crosses a barrier, the expression is deviant, indicating that antecedent

government is a condition on properly formed chains. In (122a) no barriers are crossed and the sentence is fully grammatical. In the other cases a barrier is crossed and the sentences are deviant. The violations are more severe in cases (122c) and (122e), illustrating a characteristic difference between argument and adjunct extraction.

It appears that not only a complement but also its specifier is exempt from barrierhood. Belletti and Rizzi (1981) observe that the process of *ne*-cliticization in Italian extracts *ne* from the object of the verb but not from its subject. The object, the complement of the verb, is not a barrier to government; the clitic *ne* thus governs the trace left by *ne*-extraction from the object, as required. But the trace of *ne*-extraction from the subject will not be antecedent-governed: the subject is not a complement, hence is a barrier, whether government is based on c-command or m-command. Hence, we have (123a) but not (123b).

(123) a. *pro* ne-ho visto [molti *t*]
I of.them-have seen many
'I have seen many of them'
b. *[molti *t*] ne-sono intelligenti]
many of.them-are intelligent

But now consider (124b), derived from the D-Structure (124a).

(124) a. *pro* ritengo [$_{\alpha}$[molti ne] intelligenti]
I believe many of.them intelligent
b. ne-ritengo [[molti *t*] intelligenti]
of.them-I.believe many intelligent
'I believe many of them (to be) intelligent'

Here the complement α of *ritengo* is a small clause. The phrase [*molti ne*] is the specifier of the small clause, hence is not a complement. But extraction is nevertheless permitted. We return to other illustrations of the same point.

We conclude, then, that XP is not a barrier if it is the complement of a head H or the specifier of the complement of H. The configuration of properties is not surprising, given that the head typically shares the features of its maximal projection and agrees with its specifier, so there is an indirect agreement relation between a maximal projection and its specifier. The same observation suggests that we generalize the property further: if α is the complement of H, then the daughters of α (its specifier and its head) are not barriers. When the head is an X^0, the question of

extraction from it does not arise, but it could arise in other configurations. Suppose that in a small clause (125), YP = XP, with XP being the head of YP and NP its specifier (the subject of the predicate XP).

(125) V [$_{YP}$ NP XP]

In (124a), then, α = YP = AP, and its head is the AP *intelligenti*. We have already seen that the specifier is not a barrier. Example (126) illustrates the fact that the same is true of the head.

(126) whom does he consider [$_{AP}$ Bill [$_{AP}$ angry at *t*]]

The status of (126) is no different from that of *whom is he angry at*. Thus, neither the complement nor the head of a complement is a barrier. Similarly, in (127) the main verb phrase of the embedded clause is not a barrier, and its VP head is also not a barrier, so that *who* extracts freely.

(127) I wonder [who [John [$_{VP}$[$_{VP}$ met *t*] [last night]]]]

Note that in the case of the small clause (126) as well as (127), we might also appeal to the segment theory of adjunction (section 1.3.1), requiring that a barrier be a category, not a segment, and taking the heads to be segments, hence not possible barriers.

We have dealt in a preliminary way with case (a) of (121); consider now case (b), with the configuration (128), where γ intervenes between α and β.

(128) ... α ... γ ... β ...

Recall that α c-commands the intervening element γ, which we assume further to c-command β; thus, left-to-right order in (128) expresses the c-command relation. Two cases of intervention have been explored; following Rizzi (1990), let us call them *rigid minimality* and *relativized minimality*.

(129) a. Rigid: γ is a head H (α arbitrary).
b. Relativized: γ is of the same "type" as α.

Rigid minimality can be restated in terms of barriers, taking the category immediately dominating γ to be a barrier. To spell out the concept of relativized minimality, we must characterize the relevant types. These are given in (130).

(130) a. If α is a head, γ is a head.
b. If α is in an A-position, then γ is a specifier in an A-position.
c. If α is in an $\overline{\text{A}}$-position, then γ is a specifier in an $\overline{\text{A}}$-position.

Recall that the concepts A- and $\overline{\text{A}}$-position are not properly defined in current theory; we suggested a way to approach the problem at the end of section 1.3.2 and continue to assume it here.

The three basic cases of relativized minimality are illustrated in (131) for heads, A-positions, and $\overline{\text{A}}$-positions, respectively, γ in capitals (see (44), (57)).

(131) a. *how fix [John WILL [*t* the car]]
b. *John seems [that [$_{IP}$ IT is certain [*t* to fix the car]]]
c. *guess [$_{CP}$ how [John wondered [WHY [we fixed the car *t*]]]]

In conventional terminology, case (131a) illustrates the *Head Movement Constraint* (HMC); case (131b) *superraising*; and case (131c) the *Wh-Island Constraint*. As the structure indicates, (131c) is to be understood as expressing John's puzzlement as to how we fixed the car, not as a query about how he wondered.

In (131a) *will* intervenes between *fix* and its trace, and both *fix* and *will* are heads. In (131b) *it* intervenes between *John* and its trace, both *it* and *John* are in A-positions, and *it* is the specifier of IP. In (131c), *why* intervenes between *how* and its trace, both *why* and *how* are in $\overline{\text{A}}$-positions, and *why* is the specifier of CP. In all three cases the expression is severely deviant.

We noted earlier that adjuncts and arguments behave somewhat differently with regard to extraction from barriers (see (122)). The same is true in case (130c) of intervention: compare (131c) (adjunct extraction) with (132) (argument extraction).

(132) ??guess [$_{CP}$ what [John wondered [*why* [we fixed *t*]]]]

While unacceptable, (132) is a much less serious violation than (131c).

These observations have a wide range of descriptive adequacy, but fall short of a satisfactory explanatory principle. We return to the question at the end of this section.

We have discussed some of the properties of the first case of proper government: antecedent government. Let us turn now to the second case: head government. Throughout the modules of grammar, we find relations (H, XP), where H is a head and XP a phrase with some property assigned (or checked) by H. These relations meet locality conditions that are typically narrower than either variety of command and have therefore often been considered to fall under the category of government. We noted earlier that government by a verb suffices to assign Case, bar

PRO, and license trace (section 1.3.2). In all cases the relation is narrower than command.

In Case theory we find that a verb V can assign (or check) the Case of an XP only if the XP is in a local relation to V. The verb *find* assigns accusative Case to *the book* in (133) but not in (134).

(133) a. we found the book
b. we found [$_{AP}$ the book incomprehensible]

(134) a. we found [$_{CP}$ that [$_{IP}$ the book was incomprehensible]]
b. we found the answer [$_{\alpha}$ when the book arrived]

In (133) no barrier protects *the book* from government by *find*. The same is true of (134a), but here the intervening head C^0 (=*that*) bars government of *the book* by *find*. In (134b) α is a barrier. In (134), then, *the book* must receive Case in some other way. If the construction in which it appears is infinitival, it will not receive Case at all, and the construction is ungrammatical, as in (135).

(135) a. *we tried [$_{CP}$ *e* [$_{IP}$ the book to win a prize]]
b. *we found John [$_{\alpha}$ when the book to arrive]

In (135a) the intermediate head C (=*e*) bars government of *the book*, as in (134a). It is natural to suppose, then, that government enters crucially into Case theory.

The positions to which a verb can assign Case are also, typically, those in which a trace can appear, suggesting that government by a verb can license trace. Thus, alongside (133), (134), and (135), we have (136) and (137).

(136) a. the book was found *t*
b. the book was found [$_{AP}$ *t* incomprehensible]
c. the book was believed [*t* to be incomprehensible]
d. the book seems [*t* to be incomprehensible]

(137) a. *the book was found [$_{CP}$ that [$_{IP}$ *t* was incomprehensible]]
b. *the book was tried [$_{CP}$ *e* [$_{IP}$ *t* to win a prize]]

Turning to PRO, we find a similar configuration. PRO cannot appear in governed positions, those in which, with the proper form of the verb, Case can be assigned or trace licensed.

(138) a. *we found PRO
b. *we found [$_{AP}$ PRO incomprehensible]

PRO is also excluded from positions that are governed but in which Case cannot be assigned, as in (139).

(139) a. *they expressed the belief [$_{IP}$ PRO to be intelligent]
b. *we expected [there to be found PRO]
c. *it was believed [PRO to be intelligent]
d. *it seems [PRO to be intelligent]

As discussed in section 1.3.2, we assume that the verb *believe* in English takes an IP, not a CP, complement. Thus, PRO is governed by *belief* in (139a) and *believed* in (139c), though no Case marking is possible. The constructions are barred. Thus, (139a) does not mean that they expressed the belief that someone or other is intelligent, with arbitrary PRO, or that they expressed the belief that they are intelligent, with PRO bound by *they*. Similarly, (139c) does not mean that it was believed that someone or other is intelligent; the phonetic form can only be interpreted with *it* raised, leaving a trace in the position of PRO. And (139b) does not mean that we expected there to be found someone or other, with arbitrary PRO.

A locality relation between a head and an XP also is found in θ-theory. Thus, a verb θ-marks only XPs within the VP that it heads. On the assumptions of section 1.3.2, the verb θ-marks the specifier of the VP and sisters of V′, relations that do not strictly fall under government theory, along with the complement, which does.

A closer look at head government shows that C (=C^0), whether overt or null, behaves rather differently from other heads we have considered. Thus, PRO is not barred from positions governed by C, as illustrated in (140).

(140) we decided [$_{CP}$ *e* [$_{IP}$ PRO to leave at noon]]

Similarly, C does not appear to license trace. Thus, we find that XPs move fairly freely, including VP and CP, but IP does not.

(141) a. [$_{VP}$ admit that he was wrong], John never will t_{VP}
b. [the claim t_{CP}] was made [$_{CP}$ that John was wrong]
c. *[$_{IP}$ Bill will visit tomorrow], I think [that t_{IP}]

C also does not license trace of subject. Thus, although C governs the trace in (142), extraction is barred; as is well known, languages have various special devices to overcome the problem (see below).

(142) *who did you say [$_{CP}$ that [$_{IP}$ *t* left yesterday]]

Properties of C are further illustrated in (143).

(143) a. *John was decided [$_{CP}$ *e* [$_{IP}$ *t* to leave at noon]]
b. *we decided [$_{CP}$ *e* [$_{IP}$ John to leave at noon]]
c. we decided [$_{CP}$ *e* [$_{IP}$ PRO to leave at noon]]

If the head *e* of CP were to license the trace in (143a), raising of *John* to the main clause subject position would be permitted. Note that *e* does not intervene between *John* and its trace if we adopt the notions of relativized minimality (it does under the assumptions of rigid minimality). Examples (143b) and (143c) illustrate the fact that *e* does intervene between the matrix verb and the embedded subject, blocking a government relation between them. Thus, in (143b) *John* cannot receive Case from a matrix verb, and in (143c) PRO is allowed, neither the matrix verb nor C properly governing it. Thus, C functions as an intervening head, but not a proper governor, licensing trace.

Similarly, while other X^{0}s typically raise, head-governing and thus licensing the trace left behind, that is not true of C. We find V-raising to V or I, N-raising to V (noun incorporation), I-raising to C (V-second), and so on, but we do not find C-raising to the matrix verb that governs it (e.g., incorporation into a higher verb of a verb that has been raised to V-second position). These facts too would follow from failure of C to properly govern.

C also differs from other heads with respect to barrierhood. Recall that a head typically frees a complement and its daughters (specifier and head) from barrierhood. But the situation is different in the case of C. Consider the following observations of Torrego (1985), who notes the contrast between (144) and (145) in Spanish.

(144) a. [$_{\alpha}$ de que autora] [no sabes [$_{CP}$[$_{\beta}$ qué traducciones t_{α}]
by what author don't you know what translations
[t_{β} han ganado premios internacionales]]]
have won international awards
b. *esta es la autora [$_{CP}$[$_{\alpha}$ de la que] C [$_{IP}$[$_{\beta}$ varias
this is the author [[by whom] several
traducciones t_{α}] han ganado premios internacionales]]
translations have won international awards]

In (144a) CP is the complement of *sabes* and is therefore not a barrier; its specifier β is also not a barrier, and antecedent government is not blocked, so extraction is permitted. In (144b), however, extraction is

blocked; even though β is the specifier of the complement of C, it is a barrier blocking antecedent government. A plausible conclusion is that C does not free its complement (or the daughters of the complement) from barrierhood, unlike other X^0s that we have considered, though pursuit of this issue takes us into complexities that we will ignore here.

C is unlike other heads that we have considered in other respects as well. Unlike inflectional elements, it is not a feature of the verb; thus, its specifier is not L-related, and is therefore an $\overline{A}$-position, not an A-position as are other specifiers (section 1.3.2). C also lacks the semantic content of some other heads.

In general, a good first approximation is that the proper governors are restricted to the lexical features (lexical categories, inflectional features of the verb, and perhaps others) and that only proper governors free their complements from barrierhood.

We have seen that C does not suffice as the required head governor of a subject trace. In (143a) the null complementizer *e* failed to license the trace of A-movement. The same failure is observed with an overt C in the similar configuration (145).

(145) *John is important [$_{CP}$ (for) [$_{IP}$ *t* to leave at noon]]

The paradigm with $\overline{A}$-movement (as opposed to A-movement) of the subject is less straightforward. While (142) is unacceptable, it becomes perfectly well formed if the overt complementizer is absent.

(146) who did you say [$_{CP}$[$_{IP}$ *t* left yesterday]]

In the approach outlined above, the question is how the subject trace is head-governed. Suppose there is a null complementizer and the movement of *who* proceeded successive-cyclically via the Spec of the lower CP. Then the representation would be as in (147).

(147) who did you say [$_{CP}$ *t'* *e* [$_{IP}$ *t* left yesterday]]

Spec-head agreement takes place between *t'* and *e* in this configuration. We tentatively suggest that this agreement provides *e* with features allowing it to license the trace *t*. The ungrammaticality of (142) (commonly called the *that-trace effect*), on the other hand, indicates that such feature sharing is not possible with the overt complementizer *that*. Note too that there is no derivation similar to that in (147) available for (143) since, quite generally, movement to an A-position cannot proceed through [Spec, CP]. Such "improper movement" results in an illicit A-bound

variable, as in constructions that fall under the principle Command discussed in section 1.3.3 (see also section 1.4.2).

One concern of some of the early literature on proper government (Huang 1982, Lasnik and Saito 1984) was the absence of *that*-trace effects with adjuncts. Thus, (148) is good with or without *that*.

(148) why do you think [(that) John left *t*]

Since adjuncts, like subjects, are not complements, the question arises how their traces are head-governed. When *that* is absent, the same mechanism is available as we posited for (147). But when *that* is present, no such mechanism exists, as demonstrated by (142) (see Rizzi 1990). The framework of Lasnik and Saito was slightly different so that the technical problem was actually apparent lack of *antecedent* government, but their solution can carry over under present assumptions. They suggest that as a consequence of the Projection Principle, argument traces must be licensed (γ-marked, in their terminology) at S-Structure, while adjunct traces are licensed only at LF. (142) will thus be ruled out at S-Structure while (148) will not be. Then in the LF component, *that*, being semantically empty, can be eliminated. The resulting configuration will allow government of the adjunct trace in just the same way that it allowed government of the subject trace in (147), if the head government requirement holds at LF.

In the examples we have been considering, an adjunct trace is possible in a situation in which a subject trace is not. We also find (nearly) the opposite state of affairs. (149), with movement of the adjunct *how*, is completely impossible, whereas (150), with movement of a subject, is much less severely deviant.

(149) *how do you wonder [whether John said [Mary solved the problem *t*]]

(150) ??who do you wonder [whether John said [*t* solved the problem]]

In both examples the initial trace is appropriately governed, in the manner just discussed. The difference between (149) and (150) must lie elsewhere.

Consider the structures of the examples in more detail. We assume that *whether* occupies the Spec of the CP in which it appears.

(151) *how do you wonder [$_{CP}$ whether [$_{IP}$ John said [$_{CP}$ *t'* *e* [$_{IP}$ Mary solved the problem *t*]]]]

(152) ??who do you wonder [$_{CP}$ whether [$_{IP}$ John said [$_{CP}$ t' e [$_{IP}$ t solved the problem]]]]

Lasnik and Saito argue that not just initial traces, but also intermediate traces, must be appropriately governed. But the intermediate trace t' is not antecedent-governed in either (151) or (152). In the case of (152), Lasnik and Saito argue, the intermediate trace antecedent-governs the initial trace t and then is deleted in the LF component. Such a derivation is not possible for (151) if, as they suggest, all licensing of adjunct traces is at the level of LF. Thus, if t' is present in the LF representation of (151), t will be properly governed but t' will not be. And if t' is not present at the LF level, then t will not be antecedent-governed. Either way, then, the representation contains a trace that is not properly governed.

We have just seen how (149) and (150) can be distinguished in terms of proper government. In (149) there will inevitably be an "offending trace," but there need not be one in (150). However, although (150) is much better than (149), it is not perfect, and that fact remains to be explained. Evidently, *wh*-movement is not permitted to bypass an intermediate [Spec, CP], as it did in both (151) and (152). This is one consequence of the subjacency constraint on movement proposed by Chomsky (1977) as a partial unification of several earlier constraints on movement, including those of Chomsky (1964) and Ross (1967). Subjacency violations are characteristically less severe than proper government violations, all else equal. Another property of subjacency that distinguishes it from proper government was alluded to in section 1.3.3. Subjacency constrains overt movement, but apparently does not constrain covert movement between S-Structure and LF. This is seen in the following near-minimal pair, repeated from (95a), (96a):

(153) *who do you like [books that criticize t]

(154) who [t likes books that criticize whom]

The S-Structure position of *whom* in (154) is the LF position of the trace of *whom* after LF raising, which yields a structure that is, in relevant respects, identical to the S-Structure (and LF) representation of (153). Yet the two examples contrast sharply in grammaticality. Similarly, as discussed by Huang (1982), in languages with interrogative expressions in situ, such as Chinese, the LF movement of those expressions is not constrained by Subjacency. (155) (=(95b)) is the Chinese analogue of (153), but it is acceptable, much like (154).

(155) ni xihuan [piping shei de shu]
you like [criticize who REL book]

While LF movement seems not to conform to Subjacency, it does respect the proper government requirement. The following Chinese example allows LF movement of *sheme* 'what' into the higher clause, but does not allow such movement for *weisheme* 'why'.

(156) ni xiang-zhidao [Lisi weisheme mai-le sheme]
you wonder Lisi why bought what

(156) can mean (157) but not (158).

(157) what is the thing such that you wonder why Lisi bought that thing

(158) what is the reason such that you wonder what Lisi bought for that reason

The trace of the LF movement of *weisheme* to the higher clause will not be properly governed, under the operation that yields the barred interpretation (158).

Having reviewed some aspects of the theory of movement, let us return to the basic concept of government that enters crucially into this and apparently other modules of grammar. We noted that government is a "local" form of command, tentatively taking the operative notion to be c-command. Two elements of locality were introduced: government is blocked by certain barriers and by an intervening category (the Minimality Condition). The Minimality Condition has two variants: Rigid and Relativized Minimality. We kept to the latter, following Rizzi (1990). For the theory of movement, we took the relevant forms of government, proper government, to be antecedent government and head government by a lexical head or its features (the verbal inflections).

As discussed earlier, these ideas have considerable descriptive adequacy but lack the generality and clarity that we would hope to find in an explanatory theory of language (see section 1.1). In particular, the basic and appealing intuition that lies behind the principle of Relativized Minimality is not really captured by the mechanisms proposed, which list three arbitrary cases and add unexplained complexity (the role of specifier for two of the cases); see (130).

The basic intuition is that the operation Move α should always try to construct "the shortest link." If some legitimate target of movement is already occupied, the cost is deviance (see Rizzi 1990, 22–24; also

chapter 3). We may regard this as part of the general principle of economy of derivation. Conditions quite independent of Relativized Minimality require that only heads can move to head positions, and only elements in A-positions to A-positions. Furthermore, again for independent reasons, XPs can move only to specifier positions, and α can move only to a position that c-commands it. Hence, the special properties listed in (130) can be eliminated from the formulation of the condition, which reduces to (159).

(159) Minimize chain links.

If this approach is viable, we can eliminate the intervention condition of (121) in favor of a general condition on economy of derivations, restricting the definition of government to (160).

(160) α governs β if α c-commands β and there is no barrier for β c-commanded by α.

We want government to be constrained by the same locality condition that appears in binding theory and elsewhere. Thus, an antecedent α binds an anaphor β just in case it is the *local* binder; that is, there is no γ bound by α and binding β (see section 1.4.2). Similarly, α governs β only if there is no γ governed by α and governing β. This condition is now satisfied for antecedent government, by the economy condition (159). But an analogue still has to be stipulated for head government. That raises the question of whether the head government condition is, in fact, superfluous (Frampton 1992; also see chapter 3). We will proceed on the assumption that it is required, noting the problematic aspect of this assumption.

To make this intuitive account more precise and descriptively more accurate, we have to explain in what sense a "cost" accrues to failure to make the shortest move, and why violation of the economy condition is more severe for adjuncts than arguments, as noted throughout. Adapting mechanisms just discussed, we might suppose that when a chain link is formed by Move α, the trace created is assigned * if the economy condition (159) is violated as it is created (a version of the γ-marking operation of Lasnik and Saito 1984, 1992).

Note further that only certain entities are legitimate LF objects, just as only certain entities are legitimate PF objects (e.g., a [+high, +low] vowel, or a stressed consonant, is not a legitimate PF object, and a derivation that yields such an output fails to form a proper SD). We therefore need some notion of legitimate LF object. Suppose that the chain C of (161) is a legitimate LF object only if C is *uniform* (see Browning 1987).

(161) $C = (\alpha_1, \ldots, \alpha_n)$

The only other legitimate LF objects are operator-variable constructions (α, β), where α is in an $\overline{A}$-position and β heads a legitimate (uniform) chain.

Uniformity is a relational notion: the chain C is *uniform with respect to P* (UN[P]) if each α_i has property P or each α_i has non-P. One obvious choice for the relevant property P is L-relatedness, which we have suggested to ground the distinction between A- and $\overline{A}$-positions; see section 1.3.2. A chain is UN[L] if it is uniform with respect to L-relatedness. Heads and adjuncts are non-L-related and move only to non-L-related positions; hence, the chains they form are UN[L]. An argument chain consists only of L-related positions, hence is UN[L]. The basic types—heads, arguments, adjuncts—are therefore uniform chains, legitimate objects at LF.

Taking this as a first approximation, we now regard the operation of deletion, like movement, as a "last resort" principle, a special case of the principle of economy of derivation (make derivations as short as possible, with links as short as possible): operations in general are permissible only to form a legitimate LF object. Deletion is impermissible in a uniform chain, since these are already legitimate. Deletion in the chain C of (161) is, however, permissible for α_i in an $\overline{A}$-position, where $n > i > 1$ and α_n is in an A-position—that is, the case of successive-cyclic movement of an argument. In this case a starred trace can be deleted at LF, voiding the violation; in other cases it cannot.

An expression (an SD) is a Subjacency violation if its derivation forms a starred trace. It is an Empty Category Principle (ECP) violation if, furthermore, this starred trace remains at LF; hence, ECP violations are more severe than Subjacency violations, which leave no residue at LF. Note that the concept ECP is now a descriptive cover term for various kinds of violations that are marked at LF, among them, violations of the economy principle (Relativized Minimality).

We continue to assume that traces must be properly governed: both antecedent- and head-governed by a lexical feature (i.e., not C). To unify the account, let us say that a trace is marked * if it fails either of these conditions. Thus, a trace will be marked ** if it fails both, or if it fails

one along with the economy condition, and it will be marked *** if it fails all three, with multiple starring indicating increased deviance. We have failure of antecedent government in the case of movement over a barrier, or in the case of lowering in violation of the C-Command Condition; unless the offending trace deletes, the violation remains at LF. We speculated earlier that only proper governors free their complement from barrierhood. It will follow, then, that IP (the complement of C) will be free from barrierhood only if C has a lexical feature: that will happen if V–I raises to C.

Government now is the special case of local c-command when there is no barrier. Subjacency violations fail the economy condition that requires chain links to be minimal. There is generally further deviance if the violation leaves a residue in the LF representation. Traces must be properly governed (head- and antecedent-governed), requiring raising rather than lowering, with deviance if raising crosses a barrier. The special properties of C, manifest in many respects as we have seen, impose further constraints on extraction of subjects. Deletion, like movement, is driven by FI: the requirement that derivations must form legitimate LF objects. The guiding principle is economy of derivations and representations: derivations contain no superfluous steps, just as representations contain no superfluous symbols. See chapters 2 and 3 for further discussion.

1.4.2 Binding Theory

Among the imaginable anaphoric relations among NPs, some are possible, some are necessary, and still others are proscribed, depending on the nature of the NPs involved and the syntactic configurations in which they occur. For example, in (162) *him* can be referentially dependent upon *John* (can take *John* as its antecedent), while in (163) it cannot.

(162) John said Mary criticized him

(163) John criticized him

That is, (163) has no reading in which *him* refers to John, in the way that *himself* in (164) does.

(164) John criticized himself

Apparently, a pronoun cannot have an antecedent that is "too close" to it. Note that in (162), where antecedence is possible, a clause boundary

intervenes between pronoun and antecedent. There is no such boundary between pronoun and antecedent in (163).

As we have seen in section 1.3.3, distance in this sense does not always suffice to make antecedence possible. Consider (165), where a clause boundary intervenes between *he* and *John*, yet an anaphoric connection is impossible.

(165) he said Mary criticized John

Importantly, it is not the linear relation between pronoun and name that inhibits anaphora. This is evident from consideration of (166), in which *he* once again precedes *John*, yet anaphora is possible.

(166) after he entered the room, John sat down

Similarly, in (167) *his* can take *John* as its antecedent.

(167) his boss criticized John

The generalization covering (165)–(167) is approximately as in (168).

(168) A pronoun cannot take an element of its (c-command) domain as its antecedent.

The *c-command domain* of an element is the minimal phrase containing it. Thus, in (165) the domain of the pronoun is the entire sentence. Since, trivially, the putative antecedent is included in that domain, the anaphoric interpretation is inconsistent with the generalization (168). In (166), on the other hand, the domain of the pronoun is the adverbial clause, which does not include the antecedent *John*. Similarly, in (167) the domain of the pronoun is the subject NP, *his boss*, which does not include *John*.

There are a number of ways that the generalization in (168), which relates aspects of the structure and meaning of an utterance, might be expressed in the theory. One way is in terms of a constraint (171) on *binding*, a structural relation defined in (169), and *freedom* defined in (170).

(169) α *binds* β if α c-commands β and α, β are coindexed.

(170) If β is not bound, then β is free.

(171) An *r-expression* (fully referential expression—not a pronoun or an anaphor) must be free.

The fundamental relation in this approach, *coindexation*, is a symmetric one. For an alternative in terms of an asymmetric relation, *linking*, see Higginbotham 1983, 1985. Consider how (171), often called *Condition C* of the binding theory, will treat the examples in (165)–(167). Representation (172), for sentence (165), will be excluded, while representations (173) and (174), for (166) and (167), respectively, will be allowed.

(172) *he_i said Mary criticized $John_i$

(173) after he_i entered the room, $John_i$ sat down

(174) his_i boss criticized $John_i$

Note that according to (171), (175) is permitted if $i \neq j$.

(175) he_i said Mary criticized $John_j$

Hence, if (171) is truly to play a role in capturing the generalization in (168), an interpretation must be provided for the indexing in (175) that explicitly precludes the impossible interpretation. (176) suffices in this case.

(176) If the index of α is distinct from the index of β, then neither α nor β is the antecedent of the other.

Shortly we will see reason to strengthen this constraint on interpretation of contraindexation.

Returning now to the phenomenon in (163), given that there, too, we found a constraint on antecedence, it is reasonable to suppose that (176) should again play a role in the account. Evidently, all that is necessary is that the configuration (177) be allowed and (178) prohibited.

(177) $John_i$ criticized him_j

(178) *$John_i$ criticized him_i

(171) will not be effective in excluding (178), since that constraint is limited to circumstances where the bindee is an r-expression, while in (178) the bindee is a pronoun. Further, we do not want to generalize (171) to include pronouns as bindees, since that would incorrectly preclude antecedence in (162) by disallowing representation (179).

(179) $John_i$ said Mary criticized him_i

As noted earlier, there is a locality effect involved in this paradigm. A pronoun is clearly able to be within the domain of its antecedent, hence, is allowed to have a binder, but must not be "too close" to it. (180) is a rough statement of the necessary constraint (*Condition B* of the binding theory).

(180) A pronoun must be free in a local domain.

The precise nature of the relevant local domain remains to be specified. The examples under consideration suggest that the local domain is approximately the minimal clause containing the pronoun. We will limit our attention here to purely structural approaches. See Williams 1989 for an account in terms of θ-roles, and Reinhart and Reuland 1993 for one based on predication.

Note that, as predicted, a pronoun can have an antecedent in its clause just as long as that antecedent does not c-command it. (181) is a permissible representation.

(181) $John's_i$ boss criticized him_i

Anaphors, such as reciprocals and reflexives, require antecedents that bind them. In this, their behavior is quite different from that of pronouns, which *may* have binding antecedents, but *need* not. Additionally, at least in English and a number of other languages, the antecedent of an anaphor must be local to the anaphor. In particular, we have (182), *Condition A* of the binding theory.

(182) An anaphor must be bound in a local domain.

Under the null hypothesis that the "local domain" is the same for Condition A and Condition B, we predict complementarity between pronouns and anaphors. This prediction is confirmed to a substantial degree. The ill-formed (178) becomes grammatical, if its bound pronoun is replaced by an anaphor, as in (183).

(183) $John_i$ criticized $himself_i$

Conversely, the well-formed (179) becomes bad, if its pronoun is replaced by an anaphor.

(184) *$John_i$ said Mary criticized $himself_i$

All that remains for this rough approximation is to specify the interpretation for coindexation. That is, we must guarantee that (183) cannot mean that John criticized Harry. The necessary principle of interpretation is not entirely obvious. For the moment, let us assume (185), temporarily leaving open the precise import of the notion "antecedent."

(185) If the index of α is identical to the index of β, then α is the antecedent of β or β is the antecedent of α.

We now have three syntactic constraints, repeated as (186A–C), and the two principles of interpretation (176) and (185).

(186) A. An anaphor must be bound in a local domain.
B. A pronoun must be free in a local domain.
C. An r-expression must be free.

Before considering further the precise nature of the local domain involved in Conditions A and B, we return briefly to the semantic import of indexing relations. Earlier, we hinted that (176) would need to be strengthened. Consider, in this regard, representation (187).

(187) after John_i walked in, John_j criticized him_i

This representation is fully consistent with the only relevant syntactic conditions, Conditions B and C. Neither occurrence of *John* is bound, and *him* is free in its clause. According to (176), $John_j$ cannot be the antecedent of him_i, but $John_i$ is an appropriate antecedent. It is thus unclear why (187) does not have the interpretation (and status) of (188), where coreferential interpretation for the two occurrences of *John* contributes only a minor degree of deviance.

(188) after John_j walked in, John_j criticized himself_j

Given the sharp contrast between (188) and (187) on the relevant interpretation, the extreme deviance of (187) cannot be attributed to repetition of the name, but rather must stem from the relation between the second occurrence of the name and the pronoun. We must rule out (intended) coreference between these two NPs, even when the second does not take the first as its antecedent. We achieve this result by strengthening (176) to (189).

(189) If the index of α is distinct from the index of β, then α and β are noncoreferential.

(185) must now be modified, in corresponding fashion, to (190).

(190) If the index of α is identical to the index of β, then α and β are coreferential.

Consider the contrast between the mildly deviant (191) and the severely degraded (192), both on the relevant interpretation involving only one individual.

(191) ?after John walked in, John sat down

(192) *John criticized John

Condition C excludes representation (193) for (192), while permitting (194).

(193) $*\text{John}_i$ criticized John_i

(194) John_i criticized John_j

(189) now correctly guarantees noncoreference for the two NPs in (194). But now consider (191). On the desired interpretation, the two occurrences of *John* cannot be contraindexed, since (189) would demand noncoreference for such a representation. Coindexation, too, would be problematic under (185), since (185) demands antecedence in one direction or the other, yet a name, being fully referential in its own right, presumably cannot have an antecedent. This problem does not arise once (190) is substituted for (185).

Thus far we have limited our attention to anaphoric relations among singular NPs. Certain complications arise when we extend the scope of the investigation to plurals. The configurations giving rise to noncoreference effects, by the mechanisms outlined above, seem to give rise to *disjoint reference* effects as well (Postal 1966a). Just as a coreferential interpretation of the two NPs is markedly degraded in (195), so is overlap degraded in (196).

(195) he likes him

(196) they like him

Correspondingly, (197), whose NPs lexically demand coreference, is bad, and (198), whose NPs lexically demand overlap in reference, is substantially degraded also.

(197) *I like me

(198) ?*we like me

This suggests that (189) should be further strengthened.

(199) If the index of α is distinct from the index of β, then α and β are disjoint in reference.

In (195)–(198) Condition B excludes coindexing. (199) then demands disjoint reference of the necessarily contraindexed NPs. But a problem

arises for pronouns not in configurations subject to Condition B. Consider (200) and (201).

(200) they think he will be victorious

(201) we think I will be victorious

In contrast with (197) and (198), (200) and (201) allow an interpretation where the reference of the second NP is included in the reference of the first. The result is that (200) is ambiguous and (201) is grammatical. But given the two principles of interpretation (190) and (199), there is now no possible representation available for these examples. Neither (202) nor (203) will yield a consistent interpretation for (201).

(202) we_i think I_j will be victorious

(203) we_i think I_i will be victorious

By (199), in representation (202) *we* and *I* must be disjoint in reference, but this is inconsistent with the lexical meanings of the two pronouns. And by (190), in representation (203) the two pronouns must be coreferential, which is again inconsistent with their lexical meanings. Note further that it will not do to weaken (190) so that it only demands overlap in reference, rather than coreference. This is so since in (204), for example, coreference is clearly demanded between the subject pronoun and the object reflexive, but under the hypothesized weakening, overlap should suffice.

(204) $they_i$ praised $themselves_i$

Evidently, we require a richer set of notational possibilities than we have seen so far. At least three circumstances—coreference, disjoint reference, and overlap in reference—must be accommodated. But the purely binary distinction provided by coindexing versus contraindexing straightforwardly allows for only two. To overcome this limitation, one notational device sometimes used is an index that is not a simple integer, but rather is a set of integers (Sportiche 1985). (It might seem tempting to take cardinality of index to correspond to cardinality of the referent of the NP. But such a move has no formal basis and faces insurmountable difficulties. See Higginbotham 1985, Lasnik 1989.) In accord with this convention, *free* is redefined as follows:

(205) β is free with respect to α if either α does not c-command β or the intersection of the indices of α and β is null.

Correspondingly, we modify interpretive rule (199).

(206) If the intersection of the index of α and the index of β is null, then α and β are disjoint in reference.

The problematic contrast between (198) and (201) is now straightforwardly handled. By Condition B, *me* in (198) must be free, as in (207a) or (207b).

(207) a. $we_{\{i\}}$ like $me_{\{j\}}$
b. $we_{\{j,k\}}$ like $me_{\{i\}}$

(206) then demands of these representations that the subject and object be disjoint in reference. In (201), on the other hand, Condition B is irrelevant. The indices of subject and object are therefore permitted to overlap (though still not to be identical, given (190), which we maintain).

(208) $we_{\{i,j\}}$ think $I_{\{i\}}$ will be victorious

The phenomenon of split antecedence is similarly accommodated, as displayed in (209a–b).

(209) a. $John_{\{i\}}$ told $Mary_{\{j\}}$ that $they_{\{i,j\}}$ should leave
b. $John_{\{i\}}$ told $Mary_{\{j\}}$ that $they_{\{i,j,k\}}$ should leave

Several other possibilities might also be considered. Thus, in place of the resort to set indices, we might enrich the interpretation provided for simple indices of the sort considered earlier. Consider the following interpretive procedure:

(210) a. Suppose NP and α are coindexed. Then
i. if α is an anaphor, it is coreferential with NP;
ii. if α is a pronoun, it overlaps in reference with NP.
b. Suppose NP and α are contraindexed. Then they are disjoint.

The standard cases of coreference, distinct reference, and disjoint reference now fall into place. In (195)–(198) contraindexing is required by Condition B, and the pronouns are interpreted as disjoint. In (200)–(204) coindexing is permitted, and (210aii) yields the intended interpretation of overlap in reference. It remains, however, to deal with the phenomenon of split antecedence, and further questions arise in the case of more complex constructions that we have not considered.

Another possibility would be to unify the indexing and interpretive procedures along with the binding conditions themselves, dispensing

with indexing and simplifying (210) to (211), where D is the relevant local domain.

(211) a. If α is an anaphor, interpret it as coreferential with a c-commanding phrase in D.
b. If α is a pronoun, interpret it as disjoint from every c-commanding phrase in D.

Following Lasnik (1976), we restate the former indexing requirement for r-expressions along the same lines.

(212) If α is an r-expression, interpret it as disjoint from every c-commanding phrase.

Nothing is said about interpretation in other cases. The standard examples are interpreted straightforwardly. Split antecedence is now understood to be a special case of free reference. Thus, in (209) any interpretation is permitted, including those indicated in (209), and also others, for example, an interpretation in which *they* is taken to refer to John and some third party, but not Mary.

What about more complex cases such as (213) (Wasow 1972)?

(213) the woman who loved him$_i$ told him$_j$ that John$_i$ was intelligent

Here, we have to exclude the interpretation in which the two pronouns and *John* all corefer. The problem is that the binding conditions permit both *John*$_i$ and *him*$_j$ to corefer with *him*$_i$. It then follows, incorrectly, that *John*$_i$ and *him*$_j$ can be coreferential. In the theory outlined earlier, this was excluded by the fact that coindexing is an equivalence relation, so that coindexing of both *John*$_i$ and *him*$_j$ with *him*$_i$ entails that *John* is coindexed with *him*$_j$, which is barred by Condition C. But we now have no coindexing, hence no equivalence relation.

However, the same result is achieved simply as a consequence of the interpretation itself (Lasnik 1976). By (212), *John*$_i$ is disjoint from *him*$_j$. Free interpretation allows the two pronouns to corefer and allows *John*$_i$ to corefer with *him*$_i$. If we adopt these options, *him*$_j$ and *John*$_i$ corefer, and we have an inconsistent interpretation, with *John*$_i$ both coreferential with and disjoint from *him*$_j$. Nothing further need be said. Many other complex cases follow in the same way.

The theory outlined earlier, which is the standard one, involved an indexing procedure that satisfies the binding conditions and (explicitly or implicitly) an interpretive procedure. The approach just sketched unifies

all three into an interpretive procedure. Whichever approach is followed, it now remains to consider the "local domain" in which anaphors must be bound and pronouns free.

Thus far the local domain has been the minimal clause containing the anaphor or pronoun. But this characterization is inadequate for a wider range of phenomena. In (214) the anaphor is free in its minimal clause, yet the example is well formed.

(214) John$_i$ believes [himself$_i$ to be clever]

Similarly, (215) is deviant even though the pronoun is free in the complement clause.

(215) *John$_i$ believes [him$_i$ to be clever]

We take the relevant difference between these examples and the embedded clause cases considered earlier to be in terms of government. In (214) and (215) the main verb governs the subject of the infinitival complement, as is evident from the accusative Case that shows up on that subject. In (216), on the other hand, there is clearly no such government relation, and the grammaticality judgments are the reverse of those in (214), (215).

(216) a. *John$_i$ believes [himself$_i$ is clever]
b. John$_i$ believes [he$_i$ is clever]

The local domain, or *governing category* as it is frequently called, involves reference to government, roughly as in (217), as a first approximation.

(217) The governing category (GC) of α is the minimal clause containing α and a governor of α.

In (214) and (215) the GC for the anaphor or pronoun is the entire sentence, since the governor, *believe*, is in the higher clause. Since both the anaphor and pronoun are bound in that domain, the former example is good, in obedience to Condition A, and the latter bad, in violation of Condition B. In (216) the GC is the lower clause, since the subject is assigned nominative Case by a governor internal to that clause, finite I (assuming that government is defined in terms of m-command). Since within the lower clause, there is no binder for the subject of that clause, (216a) is in violation of Condition A, and (216b) is in conformity with Condition B. Note that (217) correctly predicts that the difference between finite and infinitival complements is limited to subject position. With respect to object position, finite and nonfinite clauses are parallel.

(218) a. *John$_i$ believes [Mary likes himself$_i$]
b. him$_i$

(219) a. *John$_i$ believes [Mary to like himself$_i$]
b. him$_i$

In all four examples the GC for the anaphor or pronoun is the embedded clause, since the verb of the embedded clause is a governor of its object.

The local domain for Conditions A and B can be NP as well as IP, as seen in (220).

(220) *John$_i$ likes [$_{NP}$ Bill's stories about himself$_i$]

This suggests that (217) should be extended in the obvious way to include NP. The large NP would then be the GC for *himself* since *about* governs that anaphor. However, matters are slightly more complicated than that: unexpectedly, (221) is grammatical.

(221) John$_i$ likes [stories about himself$_i$]

Under the suggested extension, (221) should also be bad.

Note that in (220), in contrast with (221), the large NP contains not just the anaphor, but also a "potential" binder, that is, another NP that c-commands the anaphor. Our final modification incorporates this observation, and also generalizes from NP and IP to *complete functional complex* (CFC), where a CFC is a projection containing all grammatical functions compatible with its head.

(222) The GC for α is the minimal CFC that contains α and a governor of α and in which α's binding condition could, in principle, be satisfied.

This correctly distinguishes (220) from (221). As noted above, there is a potential binder, *Bill's*, for the anaphor in the large NP in (220), but none in (221). In the latter example the GC for the anaphor is thus the entire sentence, and Condition A is satisfied. Under the hypothesis alluded to in section 1.3.2 that subjects are base-generated internal to VP, the VP will be the GC, with the trace of the subject (which has itself moved to the [Spec, IP]) serving as the binder.

Note that the presence or absence of a potential binder (as opposed to an actual one) should play no role for Condition B, since there is no requirement that a pronoun have a binder at all. Hence, the minimal CFC containing α and a governor of α (where α is a pronoun) should always be the minimal such CFC in which α's binding condition could, in principle, be satisfied. This predicts that (223) and (224) should *both* be good, if in fact the NP object of *likes* in (224) qualifies as a CFC.

(223) John$_i$ likes [Bill's stories about him$_i$]

(224) John$_i$ likes [stories about him$_i$]

As expected, (223) is perfect. (224), while perhaps slightly worse, is still reasonably acceptable. This latter example thus provides one context where the usual distinctness in distribution between anaphors and pronouns seems to break down. (221), with *himself* in place of *him*, was also, of course, grammatical. Note that, as predicted, distinct distribution is maintained if there is an actual binder within the large NP, as in (225).

(225) a. I like [John's$_i$ stories about himself$_i$]
b. * him$_i$

The NP *John's stories about* ____ is the smallest potential CFC in which Condition A or B could be satisfied. While in (225a) Condition A is satisfied in that domain, in (225b) Condition B is not.

There is some evidence that the apparent overlap in distribution seen in (221), (224) is only illusory. In (224), where *him* is construed as *John*, the stories are not taken as John's. This becomes even clearer in (226), since in that example the meaning of the verb virtually forces the stories to be John's.

(226) ?*John$_i$ told [stories about him$_i$]

This suggests that (224) actually can have a structure similar to (223), but with the subject of the NP phonetically null. In that case the NP object of *likes* would clearly constitute a CFC. In (226), on the other hand, even if the NP object of *told* has a null subject, *him*$_i$ will still be illicitly bound in the minimal CFC, since that subject is understood as *John*.

However, there is one other situation where the usual disjoint distribution definitely breaks down. English has, to a limited extent, configurations permitting "long-distance" anaphors. (227) is a representative example.

(227) Mary$_i$ thinks [[pictures of herself$_i$] are on display]

Though *herself* is free within both an NP and a finite clause here, it is bound in its GC, the entire clause. There is no potential binder for the

anaphor anywhere in the lower clause, so Condition A could not be satisfied, even in principle, within the lower clause. Thus, *herself* is permitted to seek its binder in the upper clause, where, in fact, it finds it. Now note that a pronoun is possible in place of the anaphor.

(228) Mary_i thinks [[pictures of her_i] are on display]

The NP *pictures of her* (if it has a phonetically null subject), or the embedded clause (otherwise), is the smallest CFC that contains *her* and a governor of *her* (*of* or *pictures*, depending on certain assumptions about assignment of genitive Case; see section 1.4.3) and in which *her* could, in principle, be free. And *her* is, in fact, free in that domain. The limited overlap in distribution that exists is thus correctly accounted for by the "relativized" notion of GC in (222).

There is one remaining problem to consider before we leave this topic. Recall example (216a), repeated here as (229).

(229) *John_i believes [himself_i is clever]

Under the earlier absolute notion of GC, this was correctly excluded by Condition A. But under the characterization in (222), it is not. Though *himself* has a governor (finite Infl) in the lower S, there is no potential binder. The GC should therefore be the entire sentence, and *John* should be available as a legal binder. Assuming the basic correctness of the formulation of binding theory we have been developing, something other than Condition A must be responsible for the ill-formedness of (229). We suggest that the relevant condition is one discussed in section 1.4.1, which excludes traces from configurations in which they are not properly governed. On the face of it, this condition might seem irrelevant, because there is no trace evident in (229). However, it is plausible to regard the relation between a reflexive and its antecedent as involving agreement. Since agreement is generally a strictly local phenomenon, the reflexive must move to a position sufficiently near its antecedent. This might happen in the syntax, as in the cliticization processes of the Romance languages. If not, then it must happen in the LF component. In (229) this movement will leave a trace that is not properly governed. This approach directly accounts for the familiar observation that binding relations and movement processes fall under abstractly very similar constraints. Further, if it is, indeed, the requirement of agreement that is forcing the (LF) movement of the reflexive, (230), which otherwise could have been problematic, is ruled out.

(230) *himself left

Notice that there is no potential binder for the reflexive, so Condition A does not exclude the example, given the formulation of GC in (222). However, in the absence of an antecedent, the agreement requirement cannot be satisfied. These speculations suggest that for reflexives without agreement, there will be no locality requirement (Yang 1983, Pica 1987).

Given that the Condition A requirement on reflexives is thus partially subsumed under the proper government requirement on traces, the question arises of whether these two constraints fall together even more generally. Heim, Lasnik, and May (1991), expanding upon a proposal of Lebeaux (1983), suggest that the locality requirement between reciprocal expressions and their antecedents is attributable to conditions on movement. To the S-Structure of sentence (231), an LF operation of *each*-movement, adjoining the distributor *each* to its "antecedent," will be applicable, giving (232).

(231) The men saw each other

(232) $[_{IP}[_{NP}[_{NP}$ the men$]_i$ $\text{each}_j]$ $[_{VP}$ saw $[_{NP}$ t_j other$]]]$

In (233) this LF movement can be long distance. One reading (the noncontradictory one) of this sentence is representable as (234).

(233) they said that they are taller than each other

(234) $[_{IP}[_{NP}[_{NP}$ they$]_i$ $\text{each}_j]$ $[_{VP}$ said $[_{CP}$ that they_j are taller than $[t_j$ other$]]]]$

When the verb of the main clause is a "nonbridge" verb, however, movement is characteristically blocked. Compare (235) with (236).

(235) who did they say that they are taller than *t*

(236) ?*who did they mutter that they are taller than *t*

Correspondingly, the "wide scope" reading for *each* is unavailable with the nonbridge verb, leaving only the contradictory reading.

(237) they muttered that they are taller than each other

Thus, both major classes of lexical anaphors, reflexives and reciprocals, display constraints suggestive of movement.

We turn finally to the question of the level(s) of representation relevant to the binding conditions. (238), whose derivation involves raising

of the antecedent to the appropriate position to bind the reflexive, suggests that D-Structure need not meet Condition A.

(238) John$_i$ seems to himself$_j$ [t_i to be clever]

The issue is not entirely clear-cut, given the considerations of the preceding discussion, but we will tentatively assume that this is correct. Now observe that (239), from a D-Structure like that of (240), indicates that Condition C likewise need not be satisfied at D-Structure.

(239) [who that John$_i$ knows] does he$_i$ admire

(240) he$_i$ admires [who that John$_i$ knows]

Compare sentence (241), a standard Condition C violation.

(241) *he$_i$ admires everyone that John$_i$ knows

Further, (242) indicates that LF satisfaction of Condition C would not suffice. The LF representation of (241), following QR, shown in (242), is structurally very similar to the S-Structure (and LF, presumably) of (239).

(242) [everyone that John$_i$ knows]$_j$ [$_{IP}$ he$_i$ admires t_j]

The relevant difference between (239) and (242) seems to show up neither at LF nor at D-Structure, but rather, only at S-Structure. Alternatively, as discussed in section 1.3.3, reconstruction could be at issue here. Under the null hypothesis that the binding conditions apply in a block, the level of representation at which they apply is S-Structure, or, assuming reconstruction, LF.

With respect to Condition A, we have considered the distribution and interpretation of reflexives. The empty category PRO, which was briefly discussed in section 1.3.1, is very similar in its interpretation and in some aspects of its distribution. Controlled PRO generally has just the interpretation that a reflexive would have. This, in fact, was the motivation for the *self*-deletion analysis of these constructions offered in Chomsky and Lasnik 1977. Further, the principles relevant to the control of PRO appear, on first inspection, to be similar to those involved in the assignment of antecedents to anaphors. For example, as already discussed, an anaphor as subject of an infinitival clause can successfully be bound by the next subject up, as in (243), just as a PRO can be bound in the parallel configuration in (244).

(243) John$_i$ believes [himself$_i$ to be clever]

(244) John$_i$ tries [PRO$_i$ to be clever]

And as subject of a finite clause, neither is permitted.

(245) *John$_i$ believes [himself$_i$ is clever]

(246) *John$_i$ promises [PRO$_i$ will attend class]
cf. John$_i$ promises [PRO$_i$ to attend class]

Further, while both are allowed as the subject of a nonfinite clause, in most circumstances the antecedent must be the next subject up for both.

(247) *John$_i$ expects [Mary to believe [himself$_i$ to be clever]]

(248) *John$_i$ expects [Mary to try [PRO$_i$ to be clever]]

However, alongside these similarities, there are striking differences in the distributions of PRO and standard anaphors. For example, the paradigmatic position for an anaphor—direct object—is unavailable to PRO.

(249) John injured himself

(250) *John injured PRO

Further, even in the kinds of structural positions allowing both PRO and anaphors, as in (243) and (244), the precise distribution is, in general, complementary rather than identical, as seen in the contrast between (243), (244), on one hand, and (251), (252), on the other.

(251) *John believes [PRO to be clever]

(252) *John tries [himself to be clever]

Thus, there are clear, and well-known, obstacles standing in the way of analyzing PRO simply as an anaphor, and thus determining its distribution and interpretation via Condition A. There have been a number of interesting attempts to overcome these obstacles, some of them involving appeals to the theory of Case, which we will explore in section 1.4.3. Suppose, for example, that *himself* requires Case, since it is lexical, while PRO does not tolerate Case, because it is not. Then (250) is immediately accounted for: PRO is Case-marked. (252) is straightforwardly explained on the standard assumption that *try* cannot "exceptionally" Case-mark; that is, it can Case-mark a complement NP but cannot Case-mark the subject of a complement clause. And (251) is ruled out since *believe* does

exceptionally Case-mark, as seen in (243). But there are aspects of the distribution of PRO that cannot be deduced in this way. Consider (253).

(253) *John believes sincerely [Mary to be clever]
cf. John believes sincerely that Mary is clever

In (253) *Mary* fails to receive Case, perhaps because of the adjacency requirement on Case assignment. But (254) is no better than (251).

(254) *John believes sincerely [PRO to be clever]

Thus, a filter proscribing Case for PRO is insufficient.

Further examples, of the sort that have been widely discussed, indicate additional deficiencies of a purely Case-theoretic account of the distribution of PRO. Since PRO in (255) is not in a configuration of Case assignment (a lexical NP is impossible here), that example might be expected to be grammatical, presumably with an "arbitrary" interpretation for PRO, as in (256).

(255) *it is likely [PRO to solve the problem]

(256) it is important [PRO to solve the problem]

And (257) might be expected to be grammatical with an arbitrary interpretation, or possibly with PRO controlled by *John*, given the general lack (or at least amelioration) of Condition A effects in clauses with expletive subjects, as illustrated in (258).

(257) *John believes [it to be likely [PRO to solve the problem]]

(258) John_i believes [it to be likely [that pictures of himself_i will be on display]]

(259)–(260), discussed in section 1.4.1 (see (139)), display one further configuration in which Case marking is inapplicable, yet PRO is nonetheless impossible.

(259) *my belief [Harry to be intelligent]
cf. my belief that Harry is intelligent

(260) *my belief [PRO to be intelligent]

In Chomsky 1981a it is argued that the crucial factor determining the distribution of PRO is government. In particular, (261) is offered as a descriptive generalization (see also section 1.4.1).

(261) PRO must be ungoverned.

Under the standard assumption that Case marking requires government, this will entail that PRO will not be Case-marked. But the requirement is now broader, since there is government without Case marking. This is what we find in (254), (255), (257), and (260). The distribution of PRO is thus correctly described.

(261) can itself be deduced from more general properties, namely, Conditions A and B. If we take PRO to be simultaneously both an anaphor and a pronominal, as suggested in section 1.3.1, it will then follow that it will never have a GC, since if it did, contradictory requirements would be in force, given that *free* entails not bound. (261) now follows, since a governed element will always have a GC. The relevance of this to the present discussion is that control must now be independent of Condition A, the condition determining antecedence for (pure) anaphors, since to exist at all, PRO must trivially satisfy Condition A, by virtue of having no GC.

This is widely viewed as an unfortunate, or even intolerable, consequence, and a substantial amount of research has focused on redefining PRO and/or providing alternative characterizations of "governing category." For particularly interesting discussions along these lines, see, for example, Bouchard 1984 and Manzini 1983. We suggest here that control is different enough from anaphor binding that a separate mechanism for antecedent assignment is, in fact, justified. Consider first the familiar observation that in addition to the instances of control by a subject illustrated above, a controller can regularly be an object, as in (262).

(262) John told Mary_i [PRO_i to leave]

Thus far there is no evidence for distinguishing control from binding, since binding too can be by an object.

(263) John told Mary_i about herself_i

But at least two differences emerge on closer inspection. First, control is generally by a specifically designated argument. (See Nishigauchi 1984.) (264), with control by a subject instead of an object, is ill formed.

(264) *John_j told Mary [PRO_j to leave]

Binding, on the other hand, has no such constraint in English, as seen in the grammaticality of (265).

(265) John_j told Mary about himself_j

Thus, there is an optionality concerning choice of binder that does not regularly exist for choice of controller, a significant difference between the two phenomena.

Now, it is well known that there are languages unlike English with respect to this property of binding. In particular, there are languages where, apparently, only subjects can be binders. Polish is one such language, as illustrated in the following paradigm, from Willim 1982:

(266) Jan_i opowiadał Marii_j o swoim_i ojcu
John telling Mary about self's father
'John was telling Mary about his father'

(267) *Jan_i opowiadał Marii_j o swoim_j ojcu
John telling Mary about self's father
'John was telling Mary about her father'

These languages display a second difference between binding and control. For, while anaphor binding by a nonsubject is impossible, control by a nonsubject is possible (or even necessary), just as in English.

(268) Jan_i kazał Marii_j [$\text{PRO}_{j/*i}$ napisac' artykuł]
John told Mary write article
'John told Mary to write an article'

The precise nature of the parameter distinguishing English-type anaphor binding (any c-commander as the binder) from the Polish type (only subject as the binder) is far from clear. But what does seem clear is that this parametric difference does not carry over to control. For this and other reasons, there is considerable evidence for the existence of a distinct control module in the theory of grammar.

1.4.3 Case Theory

In some languages (Sanskrit, Latin, Russian, ...), Case is morphologically manifested, while in others, it has little (English, French, ...) or no (Chinese, ...) overt realization. In line with our general approach, we assume that Case is always present abstractly. In nominative/accusative languages, the subject of a finite clause is assigned *nominative* Case; the object of a transitive verb is assigned *accusative* Case (with some parametric and lexical variation, as discussed by Freidin and Babby (1984), Neidle (1988), among others); and the object of a pre- or postposition is assigned *oblique* Case (again with substantial variation). The basic ideas of Case theory grew out of the investigation of the distribution of overt

NPs, those with morphological content. Chomsky and Lasnik (1977) proposed a set of surface filters to capture this distribution, but Vergnaud (1982) observed that most of their effects could be unified if Case is assigned as indicated just above, and if Case is required for morphological realization, as stated in (269), the *Case Filter*.

(269) Every phonetically realized NP must be assigned (abstract) Case.

Chomsky and Lasnik's filters, and Vergnaud's replacement, were largely concerned with subject position of infinitival clauses. By and large, a lexical NP is prohibited in this position.

(270) *it seems [*Susan* to be here]

(271) *I am proud [*Bill* to be here]

Finite counterparts of these constructions are possible.

(272) it seems [that *Susan* is here]

(273) I am proud [that *Bill* is here]

This is as predicted, since in (272)–(273) the italicized NP is assigned nominative Case, while no Case is available for the corresponding NPs in (270)–(271).

Certain empty categories are permitted in place of the lexical NPs in (270)–(271). In (274) we have the trace of raised *Susan*, instead of the NP *Susan* itself, and in (275) we find PRO in place of *Bill*.

(274) Susan seems [*t* to be here]

(275) I am proud [PRO to be here]

Indeed, as discussed in section 1.3.1, it is the Case requirement that forces the movement producing (274) from an underlying structure like (270). (269) then need not be satisfied at D-Structure, but rather is a condition on a derived level of representation.

(276) displays another construction permitting PRO as subject of an infinitive while disallowing lexical NP.

(276) a. Bill tried [PRO to be here]
b. *Bill tried [Mary to be here]

In surprising contrast with the complement of *try*, we find just the reverse behavior with *believe*.

(277) a. *Bill believed [PRO to be here]
b. Bill believed [Mary to be here]

As seen in section 1.4.2, (276a) versus (277a) receives an account in terms of binding theory. The CP complement of *try* is a barrier to government of the subject of the complement, so PRO is allowed, having no GC in this configuration. Under the assumption that *believe*, as a lexical property, takes just an IP complement, PRO in (277a) is governed, hence has a GC. Either Condition A or Condition B is then necessarily violated. However, (277b) is not yet explained. In that example *Mary* is not the subject of a finite clause, the object of a transitive verb, or the object of a preposition, so (269) should be violated. The fact that the example is acceptable indicates that *Mary* does receive Case; (278) indicates that that Case is accusative (or oblique) rather than nominative.

(278) Bill believed [her (*she) to be here]

Further, there is evidence that the Case assigner is the matrix verb *believe(d)*. Perhaps because of the meager overt Case system in English, Case assignment generally conforms to an adjacency requirement, as illustrated in (279).

(279) a. Bill sincerely believed Sam
b. *Bill believed sincerely Sam

The same requirement exhibits itself with respect to the subject of the infinitival complement of *believe*.

(280) a. Bill sincerely believed [Mary to be here]
b. *Bill believed sincerely [Mary to be here]

Evidently, *believe* can assign accusative Case not only to its object (the core situation) as in (279a), but also to the subject of its infinitival complement, a phenomenon often referred to as *exceptional Case marking* (ECM). Recalling that (277a) shows that there is a government relation in this configuration, we conclude that Case is assigned under government (and, parametrically, adjacency), a slightly weaker requirement than the head-complement relation at its core. We tentatively take nominative Case also to fall under government, in this instance government of the subject by the inflectional head of IP (assuming an m-command definition of government).

In English the lexical heads V and P appear to be Case assigners, while N and A do not. This is why NPs can occur as direct complements of the former, [−N], categories, but not of the latter, [+N], categories, despite the fact that X-bar theory would lead us to expect the same range of

complements in both situations. Thus, while *proud* can take a clausal complement, as seen in (273) and (275), it cannot take a bare NP.

(281) *I am proud my students

Likewise, while the verb *criticize* takes an NP complement, its nominalization *criticism* does not.

(282) John criticized the theory

(283) *John's criticism the theory

In place of the NP complements in (281) and (283), we find an apparent prepositional phrase with a semantically null preposition *of*.

(284) I am proud of my students

(285) John's criticism of the theory

It seems that *of* is inserted to provide a Case assigner for a lexical NP that would otherwise be Caseless. Insertion of a pleonastic element to fulfill a morphosyntactic requirement is a rather common process. "*Do*-support" salvages an inflectional affix isolated from V by movement of I to C, as in (286).

(286) did John leave

But there is some reason to question such an account of (284)–(285). In particular, none of the other Case Filter violations enumerated above can be salvaged by the insertion of *of*.

(287) *it seems of *Susan* to be here (cf. (270))

(288) *I am proud of *Bill* to be here (cf. (271))

(289) *Bill tried of *Mary* to be here (cf. (276b))

(290) *Bill believed sincerely of *Sam* (cf. (279b))

(291) *Bill believed sincerely of *Mary* to be here (cf. (280b))

To the (271) versus (288) paradigm, with an adjectival head of the construction, could be added (292a) versus (292b), where the head is nominal.

(292) a. *my proof *John* to be here
b. *my proof of *John* to be here

That *proof* can take a clausal complement is evidenced by (293).

(293) my proof that John is here

Further, it would be expected to take an infinitival complement as an option since the verb to which it is related can.

(294) a. I proved that John is here
b. I proved John to be here

It is important to note that under other circumstances "*of*-insertion" is available with *proof*, as illustrated in (295).

(295) a. *my proof the theorem
b. my proof of the theorem

Two requirements emerge from the data examined so far. First, "*of*-insertion" takes place in the context of a [+N] head (N or A) and not otherwise. And second, *of* is available only for the *complement* of an appropriate head. It is not possible in "exceptional" circumstances. This suggests a different perspective on *of*-insertion. Instead of *of* being inserted, as a sort of last resort, before the complement of an A or N, suppose A and N are, in fact, (genitive) Case assigners, as is overtly visible in German (Van Riemsdijk 1981). *Of* can then be regarded as the realization of this genitive Case in this configuration in English. Following Chomsky (1986b), we then distinguish the *structural Cases* accusative and nominative, which are assigned solely in terms of S-Structure configuration, from *inherent Cases*, including genitive, which are associated with θ-marking. That is, inherent Case is assigned by α to NP only if α θ-marks NP. In (292), then, *John* cannot receive inherent Case from *proof* since it receives no θ-role from it. Structural Case has no such thematic requirement, but *proof*, being a noun, has no structural Case to assign. Thus, *John* receives no Case at all and violates the Case Filter. Note that under the inherent Case approach to *of*-insertion, the abstract Case needed for the satisfaction of the Case Filter can be either structural or inherent.

Passives are another construction in which Case is evidently not available, but where *of*-insertion, now viewed as inherent genitive assignment, still does not obtain. (296) illustrates this for "exceptional" Case.

(296) *it is believed (of) Mary to be here

Compare *Mary is believed to be here* and *it is believed that Mary is here*. These examples show that a passive verb, unlike a preposition or active verb, is not a structural Case assigner. The impossibility of *of* here is not surprising, given the thematic requirement we have seen. (297) is more problematic.

(297) *it is believed (of) Mary
cf. Mary is believed

Again, structural Case is unavailable, indicating that, as suggested in Chomsky and Lasnik 1977, passive verbs are not [−N]. But since *Mary* is the θ-marked complement of *believed*, inherent genitive Case might be expected. The fact that it is not possible indicates that a passive verb, while not a verb ([+V, −N]), is not an adjective ([+V, +N]) either. Rather, it is a neutralized [+V] category with no marking for the feature [N]. Alternatively, as in Baker, Johnson, and Roberts 1989, the passive morpheme is actually an argument receiving the subject θ-role of the verb and the accusative Case that the verb assigns. Accusative Case is then unavailable for the object of the verb, or for the subject of a clausal infinitival complement.

The Case Filter was originally proposed as a morphological requirement, and while such a requirement might well be at its core, there are relevant phenomena that do not seem amenable to an account in morphological terms. The trace of *wh*-movement generally must conform to the Case Filter; note that virtually all of the contexts examined thus far where a lexical NP is prohibited also disallow a *wh*-trace.

(298) *who does it seem [*t* to be here]

(299) *who are you proud [*t* to be here]

(300) *who did Bill try [*t* to be here]

(301) *who are you proud *t*

(302) *which theory did you understand the proof *t*

(303) *who is it believed *t*

Though traces have features, they have no morphological realization, so (298)–(303) are unexpected. It might be thought that it is actually the *wh*-phrase antecedent of the trace that must satisfy (269), with Case somehow being transmitted from the trace via the links of the movement chain in well-formed *wh*-questions such as (304).

(304) who did you see *t*

However, the paradigm is replicated in constructions where even the moved operator need not have overt morphological realization, as in the relative clauses in (305) or the complex adjectival constructions in (306).

(305) a. the man (who) I see
b. *the man (who) it seems to be here
c. *the man (who) you are proud to be here
d. *the man (who) Bill tried to be here
e. *the man (who) I am proud
f. *the theory (which) you understand the proof
g. *the man (who) it is believed

(306) a. Mary_i is too clever [Op_i [for us to catch t_i]]
b. *Mary_i is too reclusive [Op_i [for it to seem t_i to be here]]
c. *Bill_i is too unpopular [Op_i [for you to try t_i to be here]]

Evidently, both phonetically realized NPs and variables (traces of operator movement) must have abstract Case. Arguably, *pro*, the null pronominal subject in such languages as Italian and Spanish, must also since it typically occurs as the subject of a finite clause. In terms of phonetics and morphology, these three NP types constitute an unnatural class. It is for this reason that we instead might attribute Case Filter effects to θ-theory. As mentioned in section 1.3.1, we assume that an argument must be visible for θ-role assignment, and it is Case that renders it visible. This correctly distinguishes overt NPs, variables, and *pro*, on the one hand, from NP-trace on the other hand. Only the former are arguments.

We now assume, then, that the Case Filter is, in effect, part of the principle of θ-marking: a chain is visible for θ-marking only if it has a Case position. Economy conditions (Last Resort) block further movement if a Case position has been reached in chain formation. Given the interface condition on D-Structures, we derive the Chain Condition: in an argument chain $(\alpha_1, \ldots, \alpha_n)$, α_1 is a Case position and α_n a θ-position.

In discussing the Chain Condition in section 1.3.1, we noted two major problems: concerning expletives and PRO. The former were discussed in section 1.3.3; it remains to deal with the fact that argument PRO appears in non-Case positions, a fact that apparently compels us to adopt a disjunctive version of the Visibility Condition that falls short of a true generalization.

(307) A chain is visible for θ-marking if it contains a Case position (necessarily, its head) or is headed by PRO.

The problems concerning PRO are in fact more serious. Thus, PRO is like other arguments in that it is forced to move from a non-Case position, and cannot move from a Case-marked position, facts left unexplained even by the unsatisfactory disjunction (307).

The first problem is illustrated by such constructions as (308).

(308) we never expected [there to be found α]

If α is an indefinite NP, the counterpart to (308) is grammatical in many languages and marginally acceptable in English (more so, with "heavy NPs" such as *a hitherto unknown play by Shakespeare*); at LF, α raises to the position of the expletive, giving a chain that satisfies the Visibility Condition. But with α = PRO, the sentence is completely excluded, though all relevant conditions are satisfied: PRO occupies a θ-position as object of *find*, and choice of arbitrary PRO should satisfy the "definiteness condition," giving the meaning 'we never expected that some arbitrary person would be found'. Overt raising of PRO to the position of *there* is possible, as in (309), but with an entirely different meaning involving control by *we*.

(309) we never expected [PRO to be found *t*]

As a descriptive observation, yet to be explained, we conclude that PRO must move from a non-Case position at S-Structure, while other arguments must move from such a position either at S-Structure or at LF.

To bar (308), we might appeal to the requirement that PRO be ungoverned (see section 1.4.2). We must, however, now assume that this condition applies at S-Structure; if the condition follows from Conditions A and B of the binding theory, then these too apply at S-Structure. To account for (309), we might modify Last Resort to permit movement of PRO from a governed position.

Both the assumption that binding theory applies at S-Structure and the extension of Last Resort are open to question. Furthermore, they are empirically inadequate, because of the second problem: like other arguments, PRO is not permitted to move from a Case-marked position, even to escape government. The problem is illustrated in such forms as (310).

(310) a. α to talk about β
b. α to strike β [that the problems are insoluble]
c. α to seem to β [that the problems are insoluble]

Suppose that (310a) is a D-Structure in the context *it is unfair* —, with α = *e* and β = John. Last Resort bars raising of β to position α, yielding

(311a), because the chain (*John*) is already visible for θ-marking without movement. Suppose β = PRO. On the assumptions now under consideration, PRO must raise to the position α to satisfy the nongovernment requirement. But that movement is impermissible, even though α is a legitimate position for PRO in other constructions, as in (311c).

(311) a. *it is unfair [John to talk about *t*]
b. *it is unfair [PRO to talk about *t*]
c. it is unfair [PRO to talk about John]

One might argue in this case that there is a θ-theory violation, the subject being an obligatorily θ-marked position (a dubious move, as illustrated by nominalizations in which no external θ-role is assigned; see Chomsky 1981a). But that argument will not suffice for (310b–c) (Lasnik 1992). Here α is in a non-θ-position, so that the sentences are well formed with α = expletive *it* and β = *John* as in (312a–b).

(312) a. it is rare for it to strike John that the problems are insoluble
b. it is rare for it to seem to John that the problems are insoluble

Still, β = *John* cannot raise to the position α, leaving trace, as in (313).

(313) a. *We want John to strike *t* that the problems are insoluble
b. *We want John to seem to *t* that the problems are insoluble

In the case of β = *John*, Last Resort accounts for the phenomena, Case being assigned in the trace position and therefore barring further movement. But suppose that β = PRO in (310). The requirement of nongovernment forces movement, to yield (314).

(314) a. PRO to strike *t* [that the problems are insoluble]
b. PRO to seem to *t* [that the problems are insoluble]

PRO is now in an ungoverned position, heading a θ-marked chain. Hence, all conditions are satisfied. But the constructions are radically ungrammatical, whatever the context.

We conclude, then, that the proposal to impose the nongovernment requirement for PRO at S-Structure and to incorporate this condition in Last Resort did not solve the problem. Even with these questionable moves, the disjunctive formulation of the Visibility Condition remains empirically inadequate, as well as unsatisfactory. Some other principle requires that PRO behave like other arguments, moving from non-Case positions and barred from moving from Case positions.

Notice that these anomalies would be overcome if PRO, like other arguments, has Case, but a Case different from the familiar ones: nominative, accusative, and so on. From the point of view of interpretation, we might regard PRO as a "minimal" NP argument, lacking independent phonetic, referential, or other properties. Accordingly, let us say that it is the sole NP that can bear null Case (though it may have other Cases as well, in nonstandard conditions that we will not review here). It follows that Last Resort applies to PRO exactly as it does to any argument: PRO is permitted to move from a non-Case position to a position where its Case can be assigned or checked, and is not permitted to move from a Case position. The Visibility Condition can now be simplified to (315).

(315) A chain is visible for θ-marking if it contains a Case position.

—necessarily, its head, by Last Resort.

Observe further that in some languages, agreement plays the same role as Case in rendering chains visible (Baker 1988). Thus, abstract Case should include agreement along with standard Case phenomena. The realization of abstract Case will depend on parametric choices for functional categories. Case is a relation of XP to H, H an X^0 head that assigns or checks the Case of XP. Where the feature appears in both XP and H, we call the relation "agreement"; where it appears only on XP, we call it "Case."

In English, Spanish, and other languages with minimal overt Case marking, agreement is often manifest with PRO as well as overt NPs, as in (316), where the predicate necessarily agrees with the subject of the lower clause.

(316) a. I want [them to be officers]
b. *they want [me to be officers]
c. they want [PRO to be officers]
d. Juan cree [PRO estar enfermo]
Juan believes [(himself) to.be sick]

Thus, PRO includes φ-features for agreement, elements of abstract Case if we construe this category in the manner just indicated. It is a small further step, then, to suppose that like other NPs, PRO contains standard Case as well as agreement features.

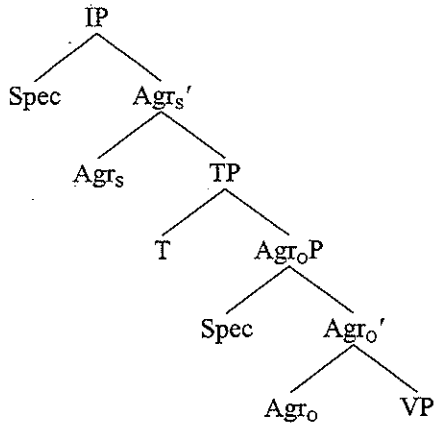
Where, then, is null Case assigned or checked (assume the latter, for concreteness)? Recall that nominative Case is standardly checked in

[Spec, IP], where I involves the features of tense and agreement (T, Agr). It is thus a realization of a Spec-head relation, with the head = I, the head of IP. It is natural, then, to take null Case to be a realization of the same relation where I lacks tense and agreement features: the minimal I checks null Case, and the minimal NP alone can bear it. More generally, we may assume that the infinitival element (with null agreement) and the head Ing of gerundive nominals check null Case, so that PRO will appear in such constructions as (317).

(317) a. PRO to VP (to be sick)
b. PRO Ing VP (being sick)

One striking anomaly still remains in Case theory. We are taking abstract Case to be an expression of an (XP, head) relation. But we still have two distinct relations of head to XP, leaving us still with an unsatisfactory disjunctive formulation: while nominative (and now null) Case is the realization of a Spec-head relation, accusative Case is assigned by V to an NP that it governs. In discussing the matter earlier, we extended government to m-command to incorporate nominative Case assignment; but apart from the Case relation, c-command appears to be the appropriate basis for government. It would be more natural to suppose that structural Case in general is the realization of a Spec-head relation, while inherent Case, which, as we have seen, is associated with θ-marking, is assigned by lexical heads. We have already touched upon this possibility in discussing the inflectional system in section 1.3.2, where we took it to have the form (318) (=(77)).

(318)



As before, the notations Agr_S and Agr_O are mnemonics; there is only one element Agr, a collection of φ-features. We continue to omit a possible [Spec, T] and negation, and to assume that at D-Structure the subject occupies the [Spec, VP] position.

Recall further that the V head of VP amalgamates with the heads Agr_O, T, and Agr_S; and at least by LF, V with its affixes has raised to eliminate all traces not c-commanded by their antecedents. Verbs may or may not have the ability to assign Case, which we may assume to be indicated by a two-valued feature [Case] for accusative and unaccusative verbs (Perlmutter 1978, Burzio 1986). If V has [+Case], then the amalgam [Agr_O, V] will also have this feature and will check accusative Case in the position [Spec, Agr_O]; if V has [−Case], an NP in [Spec, Agr_O] will not have its Case checked and must therefore move to [Spec, Agr_S]. The [Agr_S, T] amalgam checks either nominative or null Case in the position [Spec, Agr_S], depending on whether T has the value [+tense] or [−tense]. Structural Case in general is simply a manifestation of the [Spec, Agr] relation, with realizations as Case or agreement, depending on language-particular morphology.

As we have seen, one standard kind of parametric variation among languages has to do with the position of S-Structure in the derivation of LF from D-Structure. Thus, certain operations that are necessary for satisfying LF conditions may apply before or after the branch point to the PF component. The same is true of the operations that raise NP to the [Spec, Agr] positions for Case checking. Suppose that all the NP-raising operations are at LF and the language is left-headed, with V raising overtly to the inflectional position. Then as noted earlier (section 1.3.2), we have a VSO configuration at S-Structure, V and the inflectional elements having amalgamated and trace of V heading VP in (318), with subject and object remaining in their VP-internal positions. Subject will raise to [Spec, Agr_S] and object to [Spec, Agr_O] at LF. Suppose that subject raising is overt and object raising covert in the LF component. We then have an SVO configuration at S-Structure, with the VP headed by V or its trace depending on whether the language lowers inflections to V (like English) or raises V to inflection (like French, and English auxiliaries; see section 1.3.1). Suppose that the language is right-headed with overt object raising and covert subject raising; we then have OSV order at S-Structure (scrambling). If both subject and object raise overtly in a right-headed language, we will still have SOV order, but with traces in the original positions in VP. Other options are also possible.

The parameters involved are much like those that differentiate English-type languages that require overt raising of a question phrase from Chinese-type languages that leave all such phrases in situ. As discussed in section 1.3.3, we take the economy principles to prefer covert operations, which do not feed the PF component, to overt operations that do. Hence, unless a language requires that movement be overt, it will apply at LF, as in Chinese-type interrogatives or multiple *wh*-phrases in English-type languages. We might assume that what is involved is a condition on S-Structure Spec-head agreement, where the head is the C to which the *wh*-phrase raises—that is, a condition on Case, in the broad sense now under consideration. The conditions on Agr_S and Agr_O are similar. Only if S-Structure Spec-head agreement (Case, in the broad sense) is required is overt raising permissible: in English, for Agr_S but not Agr_O. For a formulation eliminating the S-Structure condition, see chapter 3.

This approach, which reduces Case agreement to a reflection of the Spec-head relation, requires that we modify the formulation of a number of the basic principles discussed earlier, while leaving their content essentially intact, for example, the Last Resort condition for movement and the associated Chain Condition. Consider the D-Structures (319).

(319) a. we believe [*e* to have [$_{VP}$ John won the election]]
b. we believe [*e* to have [$_{VP}$ been elected John]]

Assuming the VP-internal subject hypothesis, *John* is within VP in (319a) and must raise to the subject position *e*, as also in (319b), yielding the S-Structure forms (320).

(320) a. we believe [John to have [*t* won the election]]
b. we believe [John to have [been elected *t*]]

The standard account, reviewed earlier, explains this in terms of the Chain Condition, assuming an S-Structure requirement on Case assignment. Movement is a legitimate last resort operation.

We now cannot appeal to this argument for S-Structure movement. The problem is that the S-Structure forms (320) still do not satisfy the Chain Condition, because Case is checked only at the LF representations (321).

(321) a. we [John believe [*t'* to have *t* won the election]]
b. we [John believe [*t'* to have been elected *t*]]

This is one of a class of problems relating to the subject position [Spec, IP], a non-θ-position that can be occupied either by an argument

(raised from a θ-position) or an expletive, which may in turn be overt (*there, it*) or vacuous, that is, nothing but a target for movement. The expletive can be *pro*, if the language permits null subjects. In such a case, analogues to (319) would be acceptable in principle at S-Structure with *e* being *pro*, assuming the satisfaction of other conditions (the indefiniteness condition, etc.). Then LF movement would replace *pro* by its associate in the normal fashion.

Note that these problems arose in a different way in the standard account. In part the problems were conceptual: the standard account was based on the dubious assumption that Case must be checked at S-Structure, though on conceptual grounds we would expect the Visibility Condition, hence the Chain Condition, to apply only at the LF interface. In part the problems were similar to the ones just raised. Thus, in the construction (322), for example, the phrase *an error* is raised at S-Structure even though the target position is not assigned structural Case; this is checked (or assigned) only at LF, after expletive replacement.

(322) there was an error made *t* in the calculation

The problem is similar to the one we now face in the case of (319)–(321).

The EPP (see section 1.3.2) requires, for English, that the [Spec, IP] position be present through the course of a derivation, hence occupied by an expletive at D-Structure. Other optional positions (e.g., [Spec, Agr_O]) may be assumed to be inserted in the course of the derivation as part of the movement operation itself, inserting a target for movement in a manner conforming to X-bar theory. Where the expletive is inserted to satisfy the EPP, it must be either *pro* or a vacuous target for movement. English lacks the first option and must therefore accept the second: the vacuous expletive, which is only a target for movement.

A vacuous expletive, being only a target for movement, must be eliminated "as soon as possible." Either it is eliminated by the very movement operation that inserted it as a target, or, if it was inserted at D-Structure to satisfy the EPP, it is eliminated at the first opportunity in the course of derivation, hence surely by S-Structure, in the course of cyclic application of rules from the most deeply embedded structure to the highest category. Indirectly, then, (320) is left as the only option for English. It is necessary to extend this reasoning to other constructions that exhibit a similar range of properties, a matter that requires a closer analysis of the notions of economy and the status of expletives. For discussion within a considerably simplified framework, see chapter 3.

Turning now to the new version of Case theory, we can account for the fact that raising takes place at S-Structure in such constructions as (319). And since English does not require S-Structure checking of accusative Case, overt operations cannot form (321). It remains to provide a new interpretation of the Chain Condition and Last Resort, to conform to the new assumptions.

These revisions are straightforward. The Visibility Condition took Case (now including agreement) to be a condition for θ-marking. We assumed before that this was a condition on chains (the Chain Condition). We now take it to be a condition on *linked chains*, where a linked chain is formed by linking two chains C_1 and C_2 of (323), where $\alpha_n = \beta_1$.

(323) a. $C_1 = (\alpha_1, \ldots, \alpha_n)$
b. $C_2 = (\beta_1, \ldots, \beta_m)$

The new linked chain C_3, headed by α_1 and terminating in β_m, is the LF object that must satisfy the Chain Condition. In the examples (319)–(321) we have the linked chain (*John*, *t'*, *t*) at LF, in each case. The account can be simplified further in ways that we will not explore here.

Turning now to Last Resort, its intuitive content was that operations should be permissible only if they form legitimate LF objects. We now relax that requirement, taking an operation to be permissible if it is a *prerequisite* to the formation of a legitimate LF object; had the operation not taken place, the derivation would not have been able to form such an object. S-Structure raising is now a permissible last resort operation because, were it not to apply, the derivation would not yield legitimate LF objects in the case of (320), (322); the latter case indicates that this interpretation of Last Resort was already necessary in the standard account.

In presenting the standard account, we noted that the Case Filter is not satisfied at D-Structure, but rather is a condition on a derived level of representation. Apart from expletive constructions, that level was S-Structure, for English. We have now moved to the conceptually preferable assumption that the Case Filter is satisfied only at the interface level. S-Structure movement, where required, follows from the economy conditions, the EPP, and properties of expletives (including the null subject parameter).

It remains to settle many other questions (see chapter 3). But the basic structure of the system is reasonably clear, and it offers some prospects for unifying the properties of Case theory and integrating it into the general framework in a natural way.

1.5 Further Topics

The review above is sketchy and incomplete, and leaves many important topics virtually or completely unmentioned. A number of examples have been noted, among them the status of morphology, a question with broad implications, however the problems are settled. The discussion of the computational system is also crucially too narrow in that it excludes the PF component. This restriction of scope not only omits major topics (see Chomsky and Halle 1968, Goldsmith 1976, McCarthy 1979, Clements 1985, Dell and Elmedlaoui 1985, Halle and Vergnaud 1988, among many others), but also begs certain questions; as briefly noted earlier, there are open questions as to whether certain operations and properties we have assigned to the LF component do not in fact belong to the PF component (section 1.3.3).

Similar questions arise about the actual "division of labor" between the PF component and the overt syntax. Consider, for example, the "parallelism requirement"—call it PR—that holds of such expressions as (324).

(324) John said that he was looking for a cat, and so did Bill [say that he was looking for a cat]

The first conjunct is several-ways ambiguous. Suppose we resolve the ambiguities in one of the possible ways, say, by taking the pronoun to refer to Tom, and interpreting *a cat* nonspecifically, so that John said that Tom's quest would be satisfied by any cat. The constraint PR requires that the second conjunct be interpreted in the same way as the first—in this case, with *he* referring to Tom and *a cat* understood nonspecifically. The same is true of the elliptical construction (325).

(325) John said that he was looking for a cat, and so did Bill

Here too, the interpretation satisfies PR (Lasnik 1972, Sag 1976, Ristad 1993).

On our assumptions so far, PR applies to the LF representation. If (325) is generated at S-Structure, we must assume that some LF process "regenerates" something like (324), which is then subject to PR. A simple alternative would be to deny that (325) is generated at S-Structure, taking it to be formed by a rule of the PF component that deletes the bracketed material in (324) to form (325), as in earlier versions of generative grammar. That alternative is strengthened by observation of a

distinctive phonetic property of (324): the bracketed phrase has a distinguished low-flat intonation. That property, we assume, is determined within the PF component. The deletion rule, then, could say simply that material with this intonational property may optionally delete. Since such expressions as (324) have their particular status in the language, they must be generated quite independently of their elliptical counterparts. We are left, then, with a very simple treatment of ellipsis: it reduces to deletion of phonetically marked material by a general principle. The problems of parallelism, and so on, must still be dealt with for such examples as (324), but that is true independently of how we handle ellipsis.

If this approach is correct, then a wide class of elliptical constructions will be formed within the phonological component, not by operations of the overt syntax. Numerous problems remain, for example, the status of such expressions as (326), derived from the presumed underlying forms (327), which are, however, ill formed in this case.

(326) a. John said that he was looking for a cat, and Bill did too
b. John likes poetry, but not Bill

(327) a. John said that he was looking for a cat, and Bill did [say he was looking for a cat] too
b. John likes poetry, but not Bill [likes poetry]

The solution to the problem might well involve significant changes in how inflectional processes and negation are treated in the overt syntax. We leave the question here, merely noting that an approach to ellipsis that has considerable initial plausibility involves PF component properties in ways that may have large-scale effects when pursued. In this respect too, omission of the PF component leaves important questions unanswered.

The discussion of modules of language is also seriously incomplete. We have, for example, said virtually nothing about θ-theory and argument structure (see, among many others, Gruber 1965, Jackendoff 1972, 1983, 1987, 1990b, Williams 1981, Bresnan 1982, Higginbotham 1985, 1988, Hale and Keyser 1986, 1991, Wilkins 1988, Grimshaw 1990, Pustejovsky 1992) and have barely mentioned the theory of control, topics that interact crucially with other aspects of syntax. Further inquiry into these topics raises the question whether the system of modules is, in fact, a real property of the architecture of language, or a descriptive convenience.

It is unnecessary to add that this sketch also omits many other major topics that have been the focus of highly productive inquiry and provides only a scattered sample of relevant sources on the topics that have been addressed. As explained at the outset, we have attempted no more than to indicate the kinds of work being pursued within the general P&P framework and to outline some of the thinking that underlies and guides it.