Introduction

The first recognizable work of generative grammar was Noam Chomsky's *Morphophonemics of Modern Hebrew*, his MA thesis written at the University of Pennsylvania in 1951, in which he wrote:

> It is assumed that the sole purpose of the grammar is to generate a closed body of sentences, these having already been determined. Hence the grammar must be designed in such a way as to be the most efficient, economical, and elegant device generating these sentences. (Chomsky 1979 [1951], 3)

Chomsky's 1951 design for grammar included a syntactic component that was to yield "permitted arrangements of morphemes in sentences," and also an interpretive component consisting of "a series of morphological and morphophonemic statements transforming any grammatical sequence of morphemes into a sequence of phonemes." To measure the efficiency, economy, and elegance of the grammar, Chomsky proposed a simplicity metric:

> Given the fixed notation, the criteria of simplicity … are as follows: that the shorter grammar is the simpler, and that among equally short grammars, the simplest is that in which the average length of derivation of sentences is least. (Chomsky 1979 [1951], 6)

The basic goal of seeking economy in linguistic description in terms of short descriptions of the set of grammatical sentences using short calculations leading from one level of representation to others has remained constant for most researchers in generative grammar, at least in word, if not deed, as reflected in the following quote from a recent book explaining the Minimalist Program:

> We should consider theories that have a least effort flavor, e.g., requiring that derivations be short, or movements be local or operations simple or that there be no vacuous projections or operations, etc. (Hornstein *et al.* 2005, 9–10)

The generative program has yielded amazing results. Within ten years of the founding of the graduate program in linguistics at MIT, more was learned about the workings of human languages than had been learned in all previous history. But while the basic strategy of modern linguistics has changed little since 1951, its details have changed greatly. The style of our modern descriptions and the set of background assumptions we make today concerning both form and substance in language and in grammar are radically different from what they were twenty, thirty, or forty years ago. Consider, for example, the history of how rules of different kinds have been assumed to relate to each other.

0.1 Rule interaction

In *The Logical Structure of Linguistic Theory* (Chomsky 1975 [1955]), *Syntactic Structures* (Chomsky 1957), and *The Grammar of English Nominalizations* (Lees 1960a), phrase structure rules abbreviated a large but finite set of kernel sentences, which were then operated on by two sorts of transformational rules: generalized transformations that took two trees and put them together, and singulary transformations that modified individual trees. In the first phase of generative grammar the sequence of application of transformations of both types was controlled by a complicated set of "traffic rules" (Lees 1960a). In 1963 Charles Fillmore presented a scheme that effectively handled most known interactions between generalized and singulary transformations. Fillmore proposed an architecture of syntax in which all of the singulary transformations applied first to a tree, after which a generalized transformation attached the resulting tree to the bottom of a new kernel sentence. The resulting tree would then be recycled through the singulary transformations, and so on until the root of the tree had been completely processed.

A short time after Fillmore published his proposal, Chomsky's (1965) *Aspects of the Theory of Syntax* appeared in which a different scheme with the same effect as Fillmore's was put forward. Now the base component was charged both with structuring simple clauses *and* with the embedding of clauses within clauses, obviating the need for generalized transformations altogether. An output of the base component now produced a structure containing everything necessary for semantic interpretation, and this structure was then subjected to a sequence of singulary transformations that deformed the complex tree that the base component provided. Instead of Fillmore's scheme for recycling structures back through the singulary transformations as soon as a new clause

had been grafted on, *Aspects* imposed the auxiliary assumption of a cyclic principle of application for transformational rules: they were to apply to the most deeply embedded clauses first, then to the next most deeply embedded clauses, and so on until the root of the tree had been transformed. The results were the same as in Fillmore's scheme.

In the earliest phase of transformational grammar, the need for independent rules of semantic interpretation was already recognized, but when such rules were discussed at all explicitly, it was assumed that they could take all strata of the syntactic derivation into consideration. In the *Aspects* model, where deep structures contained all the meaningful pieces of the eventual surface structure, Chomsky proposed that rules of semantic interpretation applied to deep structure, an idea that was developed under the banner of generative semantics in the work of Lakoff, McCawley, Postal, and Ross. These researchers went a bit further and identified deep structure with semantic representation, thereby making rules of semantic interpretation unnecessary.

The generative semantic idea was quickly rejected by Chomsky and others, and in the heated debate that ensued, the generative semantic program became very much a minority position. Separate rules of semantic interpretation were reinstated. In the framework of Chomsky and Lasnik (1977), deep-structural configurations were preserved throughout the derivation, and all interpretation was shifted to S-structure.

The idea that semantic form is to be read off a late stage of derivation survived the transition to Government and Binding, where the level is called "S-Structure" (Chomsky 1981), through Principles and Parameters (Chomsky and Lasnik 1993), to Minimalism, where the level is named "Spell Out" (Chomsky 1993). It remains in place today. But there have been big changes elsewhere in mainstream assumptions concerning the organization of grammar. Generalized transformations (rebranded "Merge") are back and deep structure is gone again. Generative syntactic theory has returned to its roots:

> Within anything like the *LGB* framework, then, we are driven to a version of generalized transformations, as in the very earliest work in generative grammar. (Chomsky 1993: 188)

Since deep structure no longer exists, there can be no cycle of transformations applying to increasingly higher clauses. Instead, there are phases (Chomsky 2001), defined more or less by propositional nodes, resulting in a system that

is strikingly similar to Fillmore's (1963) architectural arrangement of almost fifty years ago. Recycling of some sort is not just a characteristic of transformational grammars, it is also apparently a characteristic of *theories* of transformational grammars.

If we consider also the position of morphological operations with respect to syntactic operations, we find again that nearly every possible answer has been proposed:

> Some morphology is created in the syntax and some after the syntax (Lees 1960b)
> Some morphology precedes syntax and some is in the syntax (Anderson 1982)
> All morphology precedes syntax (Lapointe 1980)
> All morphology follows syntax (Halle and Marantz 1993).

0.2 The automodular solution

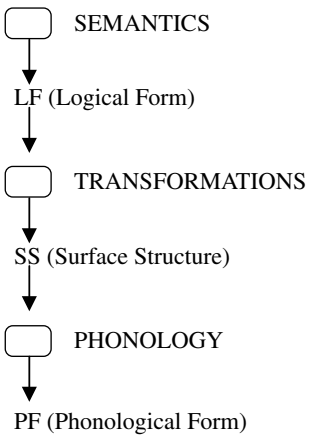
Despite decades of nearly continuous discussion, no arrangement of embedding rules, movement rules, rules of semantic interpretation, and morphological rules has won any long-term acceptance. The architectural arrangement of components that will be presented in this book settles questions by eliminating feeding relations between components of the grammar. Instead, it assumes that the following is true:

(1) The Modularity of Grammar Hypothesis
Grammatical rules of different informational types do not interact.

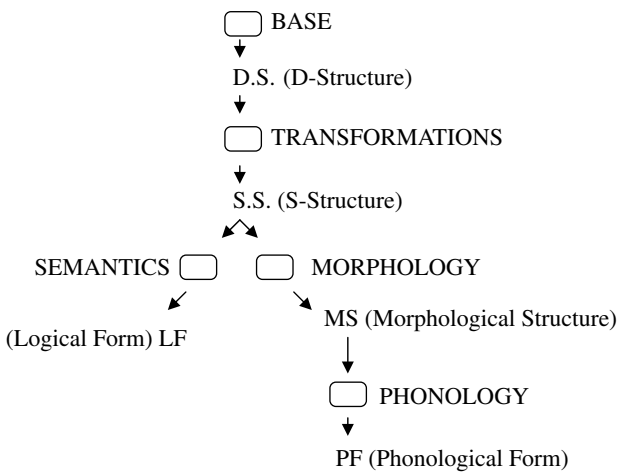
This extends Jerry Fodor's (1983) large-scale modularity of mind hypothesis to one of his modules: the language faculty. The idea is that syntactic rules, for example, operate in complete isolation from morphological and semantic rules, and furthermore, morphological rules are independent of syntax and semantics, and semantic rules are independent of morphology and syntax. Each of the modules of grammar is "informationally encapsulated," to use Fodor's felicitous term. Adopting (1) allows one to answer Anderson's (1982) famous question "Where's morphology?" this way: morphology is in the morphology. And I would add that syntax is in the syntax, and semantics is in the semantics.

Modularity in this sense does not characterize transformational generative grammars, all versions of which include at least one interpretive component whose input includes the output of another component. Diagram (2), for example, represents one form of the generative semanticists' architecture, and (3) the organization presented in Halle and Marantz (1993).

(2)



(3)



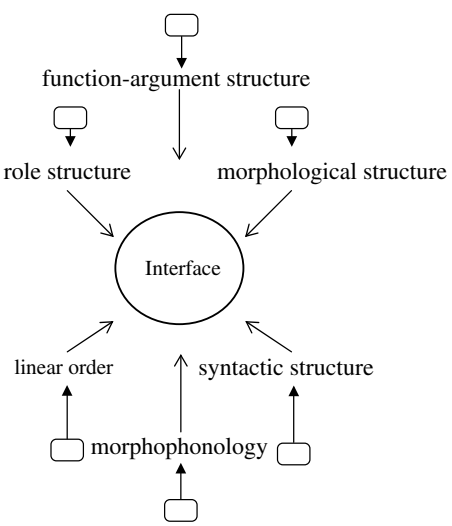
No theory with interpretive components of the kind we find in (2) and (3) is modular in the way that I, following Fodor, define modularity, since an interpretive component cannot operate independently of other modules. Without input from the base component, Halle and Marantz's transformational component, for instance, produces no output and therefore does not qualify as a module.

According to the architecture of grammar that I have espoused for more than twenty-five years and which I will further defend here, each distinguished level of representation is independently generated, the set of structures at each level comprising a separate language. These include the traditional levels of

morphological structure, syntactic structure, and semantic structure (or logical form), and three others that will be suggested in what follows. One could speak of the modules that generate these sets of representations as the syntax of morphological structure, the syntax of syntactic form, the syntax of logical form, and so on, though I will not use that terminology here.

In a grammar governed by the modularity hypothesis, as in most others, the description of an expression in a language is a set of representations, one for each distinguished level of representation that is recognized in the grammar. What is different here is that none of those representations is calculated from other representations and instead, all of them are generated independently. There must obviously be some sort of mechanism for coordinating the members of a set of representations to ensure that they are representations of the same expression. The syntactic structure of the sentence *Mary had a little lamb* is well formed and so is the semantic structure of the sentence *I've got a lovely bunch of coconuts*, but that pair of structures does not characterize any sentence of English. In an automodular grammar, then, an interface component is needed to check the compatibility of a set of representations from the several modules in order to assure that they can count as representations of the *same* expression. The particular form of modular grammar that is assumed in this book has six independent modules as shown in (4).

(4)



Automodular grammar is a refinement of the earlier versions of modular theory that went by the less-than-transparent name of autolexical syntax, e.g., Sadock (1985a, 1991), and Schiller, Steinberg, and Need (1995). The scheme

of description found in Jackendoff (1997, 2002), and Culicover and Jackendoff (2005) is also fully modular, the only other such framework I am aware of.

I will try to convince readers of the virtues of the automodular arrangement by providing a wide-ranging and detailed analysis of grammatical phenomena, mainly drawing on the facts of English, about which research in transformational grammar has yielded a tremendous amount of knowledge. I have endeavored, with incomplete success, to avoid polemics and to let the benefits of assuming autonomous modules of grammar speak for themselves. I think it is important, however, to call the attention of the reader to those features of the model that I think are especially salubrious. Besides offering a straightforward way to avoid the vexing problem of ordering the components of a grammar, the architecture depicted in (4) has several other important advantages over grammatical theories that are less than fully modular. The remainder of this book should be seen as an extended attempt to demonstrate the superiority of the automodular approach in each of these areas. For now I will just outline the benefits I claim for it.

0.3 Formalizability

If syntactic, semantic, morphological, and other types of information are dealt with in separate modules, each will deal with a single kind of information and this results in a simplification of all components, making it possible to formalize them straightforwardly. In what follows I will assume that no individual component of grammar has formal power greater than that of a context-free phrase structure grammar. The basic rules of the several components that I will talk about will be presented in enough detail to allow for formal description of grammatical phenomena, particularly those of English.

Chomsky (1957) and its much lengthier underpinnings in Chomsky (1975 [1955]) were formal enough to allow for the explicit statement of every rule that was proposed. As time went on, however, grammatical descriptions became sketchier and sketchier. I hope that the adoption of an automodular framework will help to restore the rigor of the earliest versions of generative grammar.

0.4 Concreteness

The various levels that I will employ in describing English and the other languages that will be touched upon are all reasonably concrete. What I mean by this is that the facts they describe are either directly observable, as is the case with linear order, or allow some degree of confirmation from intuition.

At the quite superficial level of analysis that I call syntax, we are pretty sure about constituent structure from a variety of tests. We also know what sentences mean and can consult this knowledge to determine scope, reference, entailment, and the like. We have a good pretheoretical understanding of what words are and how they are related to other possible words. Similar facts hold for the other modules.

Not all linguistic scientists would agree with me, but I think this concreteness is a good thing, at least at the present state of our understanding. There is nearly a compulsion among grammarians to wax abstract, to propose sweeping principles, and to assign to them grand names. But when these principles are examined, they usually turn out to be defined in terms of yet more abstract principles, some of which are themselves decomposable into other abstract postulations. There's nothing wrong with this in principle; it's what chemists and physicists do. But in grammatical practice the abstract notions are so far removed from empirical foundations that there is much disagreement as to their content.

Take, for example, the Empty Category Principle (ECP). In its most elegant and frequently cited form it says:

(5) ECP: A non-pronominal empty category must be properly governed.

The notions of non-pronominals, empty categories, government, and proper government found here need to be made precise if the ECP is to have real content. For A to govern B, it seems to be agreed that A must command B, but there are two popular definitions of command and several less popular ones. It also is subject to the condition that A is a governor, an idea that has several alternatives, and that it is "in some sense" a minimal governor of B, or alternatively, that there is no barrier to government between A and B (Chomsky 1986), all notions that have various interpretations. Moreover, proper government usually has two (sometimes three) quite different cases, one of which is usually antecedent government and another of which is either lexical government or theta government, again a group of concepts whose exact formulations are up in the air.

"The ECP" apparently doesn't deserve its definite article. There is a huge range of possible ECPs, a surprising number of which have actually been suggested by one investigator or another. What enables this uncertainty is the tenuous relation between the component principles and replicable facts of language. It is often difficult to tease out all of the ramifications of a small change in the understanding of one of the very, very abstract ideas that abound in linguistic theory. The sorts of facts that choose between alternative formulations just seem to be absent in surprisingly many cases, and if facts are available

that could distinguish two versions, they often lie at the limits of our ability to judge. The simple components of an automodular model, in contrast, generally make use of notions that are much more directly connected to fact and about which there can therefore be much less disagreement.

Consider next the idea of shortest moves:

> [W]e have made no use of the notion "minimal domain." But this too has a natural interpretation, when we turn to Empty Category Principle (ECP) phenomena. I will have to put aside a careful development here, but it is intuitively clear how certain basic elements will enter … Looking at these phenomena in terms of economy consideration, it is clear that in all the "bad" cases, some element has failed to make "the shortest move." (Chomsky 1995, 181)

The idea of short movement turns out not to be all that intuitively clear. Some cases exist where two positions that are intuitively at different distances from a third in fact need to be considered equidistant. And chain link length needs to be defined as well as the domain of a chain. Sometimes two different chains are formed by movement. But then again, there is a possibility that instead of movement for checking, what we have is attraction. Alternative definitions of short movement or the terms included in its definition continue to be put forward.

These uncertainties arose in just the first two years of the theory that replaced the ECP with the Shortest Movement Condition (SMC) and, so far as I can see, an agreed-upon "careful development" has yet to be achieved. I believe that here too the distance of the theoretical constructs from facts that can determine their validity is the problem.

One of these too-abstract concepts is movement, a concept we are so familiar with from more than half a century of theorizing that we think of movement as real. But it is not. We detect a word-order difference (among others) between *It seems that the note has disappeared* and *The note seems to have disappeared*, and we fantasize movement. The automodular model attributes structural differences to different informational levels. The two sentences above have different syntactic structures – a fact that we can directly discern – but the same logical structure – a fact readily confirmed by our intuitions. In multi-modular grammar, nothing moves.

0.5 Empirical content

Owing to the ease of formulation of rules and the concreteness of the ideas that the rules express, it is much harder to wriggle out of the way of threatening facts in the automodular program than it is in a theory where vague

and extremely abstract postulations are typical and where brand new and ever more abstract ideas are constantly introduced. When empirical or conceptual problems with a particular formulation of the ECP or the SMC have come to light, it does not seem to have embarrassed scholars, who were quick to alter the definition of command, or minimality, or proper government, or distance. No grammatical framework that I know of actually prevents its adherents from changing assumptions to fit new facts, and that is true in the case of the system of grammar I advocate, as well. But there is a great deal of difference from theory to theory in the degree of shame that disconfirmatory facts produce.

0.6 Coverage

I have been convinced for some time that linguistics is not yet a full-fledged science but remains in a largely descriptive phase. The study of electricity and magnetism remained a largely descriptive and classificatory enterprise well into the nineteenth century, but that did not inhibit the rapid and accelerating accumulation of knowledge of electrical and magnetic phenomena from the beginning of the eighteenth century onward. The results of electrical investigations before Maxwell were never simply lists of experiments and their results, but were constantly the subject of hypothetical suppositions whose assumptions were never very far from contemporaneous established fact. This allowed for investigations to be guided by hypotheses while not preventing the description of new phenomena of all kinds.

Theoretical linguistics made rapid progress after Chomsky's philosophical revolution in the 1950s. As one who has been in the field for a long time, I have to report my feeling that the rate of discovery has slowed considerably. Constructions that have been treated many times before are still discussed from novel theoretical perspectives rather than turning out to be automatically accounted for in the new paradigm.

Take the case of *tough* movement, a phenomenon that was touted in the early days of generative grammar as supporting the new paradigm over the old, the claim being that the old paradigm could not account for its properties whereas the new paradigm provided a straightforward account. Yet within the tradition of Government and Binding that began some twenty-five years later, *tough* movement became, according to Holmberg (2000) "in principle unexplainable," and there it remained for another dozen years despite numerous attempts to explain it, including one in the theory's foundational document, Chomsky (1981). But the Minimalist Program, launched in Chomsky's 1995 paper, offered a solution and fourteen years later Glyn Hicks (2009) observed that "Despite the advances that the field has seen in nearly 50 years,