Introduction: Grammars as Mental Representations of Language

Joan Bresnan
and
Ronald M. Kaplan

1. The Problem of the Psychological Reality of Grammars

A longstanding hope of research in theoretical linguistics has been that linguistic characterizations of formal grammar would shed light on the speaker's mental representation of language. One of the best-known expressions of this hope is Chomsky's *competence hypothesis:* ". . . a reasonable model of language use will incorporate, as a basic component, the generative grammar that expresses the speaker–hearer's knowledge of the language . . ." (Chomsky 1965:9). Despite many similar expressions of hope by linguists, and despite intensive efforts by psycholinguists, it remains true that generative–transformational grammars have not yet been successfully incorporated in psychologically realistic models of language use (Fodor, Bever, and Garrett 1974).

In discussing grammars as mental representations, it is essential to understand that the terms *grammar* and *theory* are both used on two entirely different levels of description. On the lower level of description, we speak of the grammar of a particular language such as Navajo. At this level, a grammar is a set of rules within a formal system. The grammar generates the language that it is a grammar of and, on analogy with the technical usage in other formal systems, the grammar is sometimes called a "theory" of the language that it generates. On the higher level of description, we speak of a theory of grammars. This is a set of primitives, axioms, and rules of inference (often unformalized) that characterizes the class of possible grammars of particular languages. A theory of grammar is sometimes referred to as a *Universal Grammar*.

These dual concepts of *grammar* and *theory* play very different roles in discussions of the mental representation of language. To learn to speak Navajo, one must acquire specific, if tacit, knowledge of the sentence patterns and pronunciation of Navajo. A grammar of Navajo, in that it provides specific rules for the construction of Navajo sentences, represents the kind of knowledge of that language that one must have to speak it. It is such grammars, grammars on the lower level, that we assume will represent the stored knowledge in competence-based models of linguistic performance. Grammar on the higher level, the Universal Grammar that is a theory of grammars, is not necessarily represented in such models in the same way. For example, principles of Universal Grammar might characterize aspects of the structure of the language-using device.

In response to the fact that generative–transformational grammars (on the lower level) have not been successfully incorporated in realistic models of language acquisition, comprehension, or production, many psycholinguists have come to the view that it is a mistake to adopt Chomsky's competence hypothesis; there need not be *any* transparent mapping between linguistically motivated formal grammars and psychological models of language use (for example, the language user may employ agrammatical heuristic strategies as knowledge representations (Fodor, Bever, and Garrett 1974)). If this is so, the knowledge representations actually used in language acquisition, production, and comprehension will not satisfy the postulates of Universal Grammar, which makes quite specific claims about the form, organization, and interpretation of rules of (lower-level) grammars. But then what mental structures is Universal Grammar a theory of?

In response, Chomsky has taken the view that it is a mistake to regard "psychological reality" as anything other than whatever linguistic theory is about: "Challenged to show that the constructions postulated in that theory have 'psychological reality,' we can do no more than repeat the evidence and the proposed explanations that involve these constructions" (Chomsky 1980b:191). Comparing the linguist to an astronomer studying thermonuclear reactions within the sun, Chomsky argues, "[I]n essence . . . the question of psychological reality is no more and no less sensible in principle than the question of the physical reality of the physicist's theoretical constructions" (Chomsky 1980b:192). However, neither Chomsky's response nor the response of those who have abandoned the competence hypothesis is satisfactory.

Consider first the view that rejects the competence hypothesis. In rejecting the hypothesis, proponents of this view do not also reject the notion that some form of stored linguistic knowledge is employed in all forms of language behavior. For example, it is generally acknowledged that models of linguistic comprehension must include a set of previously learned phonological and syntactic patterns for the language (English, Chinese, or Navajo) being comprehended. Similarly, models of linguistic production require stored information about the syntactic and phonological structure of the language being spoken, in order to convert the input (the speaker's message) into the form of speech. Likewise, models of language acquisition assume that the learner constructs and stores linguistic knowledge structures, forming an "internal grammar" on the basis of the primary data in the linguistic environment and a set of built-in constraints on the induction process. Each state of the learning process can be represented by a distinct internal grammar; but the sequence of these internal grammars must converge, in the final state of the learning process, on a "target" grammar which represents mature knowledge of linguistic structure. Finally, the formal grammars of linguistics are themselves abstract models of mature knowledge of language, as reflected in linguistic judgments and the other adult verbal behavior studied by linguists.

If it is uncontroversial that stored knowledge structures underlie all forms of verbal behavior,[1] the question arises of how these different components of linguistic knowledge are related. To reject the competence hypothesis is to adopt the theoretical alternative that a different body of knowledge of one's language is required for every type of verbal behavior. While this state of affairs is certainly a logical possibility, it is the weakest hypothesis that one could entertain, since it postulates multiple stores of linguistic knowledge that have no necessary connection. In contrast, the competence hypothesis postulates an isomorphic relationship between the different knowledge components and is thus the strongest and simplest hypothesis that one could adopt. On methodological grounds, it should be given priority over weaker alternatives: it enables us to unify our theories of the mental representation of language, to construct our processing models on the basis of a theoretical understanding of the structure of the knowledge domain, and to bring mutually constraining sources of evidence to bear on studies of process and of structure.

Granting that the competence hypothesis is desirable in principle, though, is it tenable in fact? In particular, if we do maintain the competence hypothesis, how can we then explain the conflict between psycholinguistic studies and linguistic theories of the mental representation of language? This is the scientific challenge posed by work on the "psychological reality" of grammars, as presented by Fodor, Bever, and Garrett 1974, Levelt 1974, and others. In response to this challenge, Bresnan 1978 pointed out that these psycholinguistic studies presupposed a transformational characterization of linguistic knowledge which could simply be wrong, a possibility that had also been suggested in the ATN-based work of Wanner, Kaplan, and others (Wanner and Maratsos 1978, Kaplan 1972). Wanner's and Kaplan's studies showed that a psycholinguistic model of syntactic processing based on a competence grammar can be computationally implemented and used to generate detailed and experimentally testable predictions. Bresnan argued that a more radical decomposition of competence grammars into an expanded lexical and contracted syntactic component promises to have far greater explanatory power than the current versions of transformational grammar, permitting a unification of linguistic and psycholinguistic research. Subsequent collaborative research by Bresnan, Kaplan, Ford, Grimshaw, Halvorsen, Pinker, and others has begun to bear this out, as the studies in this volume demonstrate. This work shows that the competence hypothesis is indeed tenable.

Consider next Chomsky's view, that psychological reality is whatever linguistic theory is about. Recalling the quotation cited earlier, we see that Chomsky construes the problem of psychological reality as an *ontological* problem: "[I]n essence . . . the question of psychological reality is no more and no less sensible in principle than the question of the physical reality of the physicist's theoretical constructions." In other words, he takes the question to be whether the rules and other constructs of linguistic theory "have reality," whether they describe real mental entities and processes. While Chomsky's answer to this question is surely a reasonable one, this is the wrong *question*. The cognitive psychologists, computer scientists, and linguists who have questioned the psychological reality of grammars have not doubted that a speaker's knowledge of language is mentally represented in the form of stored knowledge structures of some kind. All theories of the mental representation of language presuppose this. What has been doubted is that these internal knowledge structures are adequately characterized by transformational grammars—or indeed, by any grammars that are motivated solely by intuitions about the well-formedness of sentences. The challenge to Chomsky's theory is not the philosophical question that he addresses (whether theoretical constructs correspond to real mental entities and processes), but the scientific question (whether these theoretical constructs can unify the results of linguistic and psycholinguistic research on mental representation and processing). To the latter question, Chomsky's response is plainly inadequate: "Challenged to show that the constructions postulated in that theory have 'psychological reality,' we can do no more than repeat the evidence and the proposed explanations that involve these constructions" (Chomsky 1980b:191).

On Chomsky's view, then, a grammar is psychologically real if it contributes to the explanation of linguistic judgments and the other verbal behavior studied by linguists, and nothing more need be said. This, however, is a much weaker conception of psychological reality than we would like. (The following examples are taken from *The American Heritage Dictionary* and particularly the appendix "Indo-European Roots.") For example, the English words *baritone* and *grieve* both derive from an Indo-European root *gwer-* 'heavy'. Historically, the word *baritone* came into English from Italian (ultimately from Greek), while *grieve* came into Middle English from Old French (ultimately from Latin). Latin and Greek, of course, emerged historically from distinct ancient dialects of the common Indo-European mother tongue. Now it is possible to construct a formal system of morphophonemic rules and abstract representations that deductively account for these historical relations. By such rules one can formally derive English words from their Indo-European roots. Thus, the labiovelar *gw* is the source of both the initial labial *b* in *baritone* and the initial velar *g* in *grieve*. The same relationship appears in *bar* and *gravid,* as well as many other examples. Would such a formal rule system be psychologically real? With Chomsky's conception of psychological reality, we could answer affirmatively. The rule system might well contribute in some ways to an explanation of English speakers' linguistic judgments of what are well-formed sentences of their language. It could even be argued that the rules and representations of the system do characterize the *competence* of an idealized speaker–hearer, abstracting away from "performance" limitations of memory or perhaps education.

Given the stronger conception of psychological reality that we would like to maintain, this conclusion is absurd. It is most implausible that the remote historical derivations of the English lexicon are part of contemporary English speakers' mental representations of their language. Although this illustration slightly exaggerates the practice of some linguists, it serves to make two important points. First, linguistically motivated descriptions of a language need not bear any resemblance to the speaker's internal description of the language. Therefore, one cannot justifiably claim "psychological reality" for a grammar (in any interesting sense) merely because the grammar has some linguistic motivation. Second, the concept of *competence* has often been abused; in the above argument, for example, it now appears to mean that a linguistic rule system need not play *any* role in *any* model of performance. But the true import of the competence hypothesis is exactly the opposite: it requires that we take responsibility not only for characterizing the abstract structure of the linguistic knowledge domain, but also for explaining how the formal properties of our proposed linguistic representations are related to the nature of the cognitive processes that derive and interpret them in actual language use and acquisition. Chomsky's current conception of psychological reality represents a retreat from this more ambitious, and scientifically far more interesting, goal.

One might think that abuses of the above kind could be ruled out by appealing to theoretical simplicity: perhaps the rules deriving English words from Indo-European roots would not fit in elegantly with the simplest grammar for English. But simplicity is itself a theory-bound notion; as Chomsky 1970 has argued, the choice of a simplicity metric is made on the same empirical grounds as the choice of a theory. Moreover, it is easy to imagine even highly elegant and deductively satisfying rule systems that lack psychological reality in the sense we would like. There is evidence, for example, that the standard mathematical axiomatization of arithmetic differs from the system of conceptual competence that children display in counting (Greeno, Riley, and Gelman to appear). Although the two rule systems may be extensionally equivalent, it appears that they are built up from different sets of basic concepts and procedures. If this is so, we would *not* want to say that the standard mathematical axiomatic system is "psychologically real"; for while it does describe the conceptual structure of the knowledge domain, it appears to differ in essential ways from our internalized characterization of that knowledge.

Another response to our example would be to deny that there is anything absurd about it. The rules and representations of a linguistic system characterize the competence of an *ideal* speaker–hearer, who interprets his or her knowledge fully and faithfully, is immune to distraction, has no memory limitations, does not make errors, etc. The on-line behavior of real native speakers, to which a formal linguistic system may have no discernible relation, is jointly determined by that system *and* these other performance factors, and the latter are the source of any processing discrepancy. For this reason, processing considerations can have no direct bearing on grammatical theory.

As Chomsky (1980 and elsewhere) has pointed out, idealization has been a crucial ingredient for progress in the so-called mature sciences. Science progresses by abstracting away from the flux of experience to reveal underlying and invariant relationships. Likewise, linguistics and cognitive science will not make scientific progress unless we identify and focus on crucial idealizations, such as the ideal native speaker. There is nevertheless something suspicious about this appeal to idealization. It is used mainly to restrict the kind of evidence that may be brought to bear on representational issues. Thus, it appears to be a way of insulating linguistic theory from the other cognitive sciences.

This is not the role that idealization should play in science. The lawful relations that are constructed around such concepts as ideal gases, frictionless planes, infinitesimal masses, or infinite distances become scientifically interesting only when it can be shown that the behavior of real gases, planes, and particles can be made to come arbitrarily close to the postulated ideal as extraneous sources of variation are identified and controlled. In other words, there is a scientific responsibility to show that the real *does* asymptotically approach the ideal under appropriate circumstances. If this responsibility is taken seriously, there are more, not fewer, ways in which processing results might bear on representational issues, contrary to what Chomsky's appeal to idealization suggests. In particular, we must discover ways of showing that the actual behavior of real native speakers converges on the ideal behavior predicted by our grammatical theory, as interfering performance factors are reduced. Developing methods of reducing such interferences is of course a very hard scientific problem, but analogous problems had to be solved in the other sciences as a precondition for their major advances. Only as we make progress toward this goal will we be justified in speaking of a grammar as a model of the ideal native speaker's knowledge.

In attributing psychological reality to a grammar, then, we require more than that it provide us with a description of the abstract structure of the linguistic knowledge domain; we require evidence that the grammar corresponds to the speaker's *internal* description of that domain. Since we cannot directly observe this "internal grammar," we must infer its properties indirectly from the evidence available to us (such as linguistic judgments, performance of verbal tasks in controlled experimental conditions, observation of the linguistic development of children, and the like). The data of linguistics are no more or less privileged for this inquiry than any other data. The formal representations of linguistic theory, however, when joined with the information-processing approach of computer science and with the experimental methods of psycholinguistics, provide us with powerful tools for investigating the nature of this internal grammar and the processes that construct and interpret it. The methods and results of these different approaches can mutually constrain the form of a competence-based model of linguistic performance. In what follows we suggest that constraints on knowledge representation derived from considerations in theoretical linguistics may serve to limit the class of compatible information-processing models and, likewise, that computational considerations may have direct implications for the form in which linguistic representations must be cast.

2. Theoretical Constraints on Knowledge Representation

Theoretical constraints on representations of knowledge affect the classes of computations that can construct or interpret those representations. Thus, the choice of grammars can limit the choice of process models. A grammar may prove to be unrealizable in explanatory models of language processing because of the basic representational assumptions that it embodies. To amplify this point, let us compare the representational assumptions of two very different theories of grammar: the transformational theory and the lexical–functional theory. We will focus on the representation of grammatical relations; these are the associations between the semantic predicate argument structures (or thematic role structures) of sentences and their surface constituent structures. The term *grammatical relations* is thus used here in a theory-neutral sense, to be distinguished from *grammatical functions* such as SUBJ(ect) and OBJ(ect).

Let us consider first a simple, concrete example illustrating the different representational assumptions of the two theories. The basic motivation for syntactic transformations during the early days of generative grammar was the systematic relationship exhibited by pairs of sentences like *Mary kissed John* and *John was kissed by Mary*. In the first (active) sentence, the noun phrase preceding the verb has the agent or "kisser" role; in the second (passive) sentence, it has the theme or "kissed" role. If there were no transformational rules that moved the noun phrase following the verb to the position in front of the verb (so it was argued), we would have to postulate two different lexical entries for *kiss*—one with an agent subject which appears in the active form, and one with a theme subject in the passive form. The postulation of syntactic transformations removes this redundancy and expresses the regularity that systematically relates active sentences to passive sentences.

Given the original assumptions of transformational grammar (which lacked any mechanisms for representing grammatical relations other than transformations of phrase structures), the above argument was certainly persuasive. At a time when the lexicon was considered to be simply a collection of idiosyncratic properties of morphemes, it seemed obvious that a single lexical entry is better than two. When a richer theory of lexical rules and representations becomes available, though, the argument loses its persuasiveness. Lexical entries can represent semantic predicate argument structures independently of phrase structure forms, and lexical rules can capture the redundant relationship between the two lexical entries for *kiss*. For the two sentences given, we need to know only that in one, the subject is the agent and the object is the patient, and in the other, the subject is the patient and the *by*-object is the agent. In the lexical–functional theory, we represent this information in the lexical entries for the active and passive verbs, as in (1) and (2). (In (2) we have designated the English *by*-object by the more general function name OBL(ique)$_{\text{AG(ent)}}$.)

(1)
kiss: kiss⟨ (SUBJ) (OBJ) ⟩
AGENT PATIENT

(2)
kissed: kiss⟨(OBL$_{\text{AG}}$) (SUBJ) ⟩
AGENT PATIENT

Note that both lexical entries (1) and (2) have the same lexical predicate argument structure: kiss⟨AGENT PATIENT⟩. They differ in the grammatical functions that express the agent and patient arguments. The grammatical functions SUBJ, OBJ, and OBL$_{\text{AG}}$ are universal, but their phrase structure realizations vary from language to language, just as the surface forms of particular languages vary. Thus, in English the SUBJ is realized as an NP preceding the verb; the OBJ is realized as an NP following the verb; and the OBL$_{\text{AG}}$ is realized as a prepositional phrase marked with *by*. Finally, in order to capture the systematic relationship between the two lexical entries, we may propose a lexical rule that changes SUBJ to an optional OBL_{AG} and OBJ to SUBJ:

(3)
(SUBJ) → (OBL_{AG}) / ϕ
(OBJ) → (SUBJ)

By such a rule, the passive lexical form (2) can be derived from the active form (1). Chapter 1 of this volume gives a detailed analysis of passive constructions in these terms. In general, systematic relations between lexical forms can be expressed by means of lexical rules or principles for associating grammatical functions with predicate arguments.

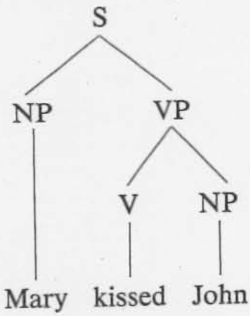
With this example of the active–passive relation in mind, let us now consider the underlying representational assumptions that result in these particular analyses. Despite continually changing theoretical assertions, the basic representational assumptions of transformational grammar have remained surprisingly constant since at least the "standard theory" (Chomsky 1965). The first assumption is that there is a one-to-one correspondence between the semantic predicate argument structure of a sentence and a set of ("deep") grammatical functions, such as "(logical) subject," "(logical) object," etc. The second assumption is that these grammatical functions can be reduced to a set of canonical phrase structure configurations: for example, the "logical subject" becomes the NP immediately dominated by S in the deep phrase structure representation of the sentence, and the "logical object" becomes the NP immediately dominated by VP. These two assumptions permit semantic predicate argument structures to be represented by canonical, or deep, syntactic phrase structures. Because the surface forms of sentences are also represented as syntactic phrase structures, the mapping between semantic arguments and surface form (which constitutes the grammatical relations of a sentence) must be expressed through operations on phrase structures. This naturally gives rise to the third representational assumption, that there is a set of structure-dependent operations, or syntactic transformations, which map between the deep phrase structure (representing semantic predicate argument structure) and the surface phrase structure. Analogues of these assumptions underlie current versions of transformational theory. For example, in Chomsky's 1981 government and binding theory, the one-to-one correspondence between predicate argument structures (thematic structures) and deep phrase structures (D-structures) appears as the *Theta Criterion*. The mapping between D-structures and surface phrase structures is accomplished in two steps: first, syntactic transformations (restricted to the form "Move α") map D-structures onto S-structures (abstract surface structures which contain indexed empty categories); second, a component of "stylistic" transformations, deletion rules, and morphophonological rules maps S-structures onto surface forms.

Under these assumptions, the grammatical relations of the sentence *John was kissed by Mary* must be represented in a particular way. First, by the assumption that there is a one-to-one correspondence between semantic predicate argument structure and deep grammatical functions, the surface subject of the passive sentence must have the same deep function as the surface object of the active sentence *Mary kissed John,* for these two NPs have the same thematic role of patient. Second, by the assumption that such deep functions reduce to a canonical set of underlying phrase structure configurations, *John* must be represented in the same deep structure position in both the active and the passive sentence. Third, by the assumption that the mapping between thematic role structure and surface form is effected by syntactic transformations, *John* must be moved from its deep structure position to its surface structure position. Thus, these basic assumptions about the representation of grammatical relations require that one of the constructions in question be formed by the movement of an NP from an underlying phrase structure position (as the logical object, for instance) to a surface phrase structure position (as the surface subject).

In sum, the guiding representational idea of transformational theories of syntax is that semantic predicate argument structure must be characterized in the vocabulary of constituent structure representation and must, in fact, be directly reflected in the forms of phrase structures. This is the meaning of Chomsky's Theta Criterion. As we have seen, this immediately leads to a theory in which the grammatical relations of sentences are encoded by phrase-structural computations (such as the transformational derivation).

In contrast, the theory of lexical–functional grammar (LFG) maintains a very different set of representational assumptions. First, on this theory, there is *no* one-to-one correspondence between semantic predicate argument structure and grammatical functions; the "theta criterion" of transformational theories is rejected. Instead, a single predicate argument structure may have several alternative lexical assignments of grammatical functions, governed by universal principles of function–argument association (see chapters 2, 3, and 5, M. Rappaport 1980, M. Baker 1982). This is illustrated by the active and passive lexical forms of the verb *kiss* in (1) and (2). Second, on the lexical–functional theory, grammatical functions are *not* reducible to canonical phrase structure configurations; on the contrary, the phrase structure categories themselves appear reducible to functional primitives (Jackendoff 1977 and chapter 5 of this volume), and the relation between structural configurations and grammatical functions is clearly many-to-many, varying across language types and even within languages (chapters 5 and 8, Simpson and Bresnan 1982). Since this theory requires no "normalized" phrase structure representation to express predicate argument relations, the structural component of the grammar can be vastly simplified, the entire transformational derivation being replaced by a single level of phrase structure representing the surface form of a language, the *constituent structure* (c-structure) (see chapters 4 and 5 of this volume). Third, in the lexical–functional theory, the mapping between thematic role structure and surface form is *not* effected by syntactic transformations (or equivalent structural computations). Rather, it is effected by correlating the grammatical functions that are assigned to lexical predicate argument structures with the grammatical functions that are syntactically associated with c-structure forms; *functional structures* (f-structures) formally represent these correlations. (See figure 1 for a simple illustration.) F-structures represent grammatical relations in an invariant universal format which is independent of language-particular differences in surface form. The f-structures are semantically interpreted (Halvorsen 1982), while the c-structures are phonologically interpreted.

Thus, in the lexical–functional theory, unlike the transformational theory, phrase structure computations play only a very restricted and superficial role in the mapping from predicate argument structure to surface form; the greater part of this mapping is lexically encoded independently of phrase structure form by the assignment of universal grammatical functions to lexical predicate argument structures. The guiding idea of the LFG theory is that *only* lexical rules can alter these function–argument associations. Syntactic rules must therefore preserve function–argument correspondences. This is called *the principle of direct syntactic encoding* (chapters 1 and 4 of this volume). One consequence of this principle is that active and passive verbs, because *c-structure*



f-structure

SUBJ	[PRED	'MARY']
TENSE	PAST	
PRED	'KISS⟨(SUBJ)(OBJ)⟩'	
OBJ	[PRED	'JOHN']

Figure 1

they induce different grammatical relations, must have different lexical entries. Another consequence is that there can be no NP-movement transformations like the Passive transformation.

Because the various function–argument correspondences are already encoded in lexical entries, no phrase structure manipulations are needed to express the grammatical relations of sentences. Instead, active and passive lexical items are lexically inserted directly into surface constituent structures (c-structures). The syntactic instantiations of the grammatical functions can be read off from functional annotations to these surface structures, as described in chapter 4 of this volume. A phrase structure satisfying the active verb *kiss* must have both an object and a subject; a phrase structure satisfying the passivized form of the verb must have a subject, but no object. How the phrase structure subject is interpreted—whether as "the kisser" or "the one kissed"—depends on which lexical form of the verb is satisfied by the phrase structure tree.

This theory of grammatical representation explains why passivization in English appears to involve NP movement (and why other grammatical-relation-changing rules of English do so as well). The lexical rule of Passivization shown in (3) refers only to the grammatical functions SUBJ, OBJ, and OBL_{AG}, and not to the phrase structures in which these functions are syntactically expressed. In English, the SUBJ and OBJ functions are structurally expressed by different positions in the phrase structure: the subject NP appears before the verb and is immediately dominated by the S node, while the object NP appears after the verb and is immediately dominated by the VP node. It follows that when an active verb like *kiss* is lexically inserted into a structure, its patient argument, which has been lexically assigned the OBJ function (as in (1)), will be structurally expressed by a postverbal NP; and when a passivized verb like (*be*) *kissed* is lexically inserted into a structure, this same patient argument, which has now been lexically assigned the SUBJ function (as in (2)), will be structurally expressed by a preverbal NP. The lexical change in these function–argument correspondences will therefore induce a change in the phrase structure positions that express these arguments. Thus, the syntactic effects of Passivization in English will appear to involve the "movement" of an NP from one phrase structure position to another.

But this apparent NP movement is only an illusion, an effect arising from the way that these universal grammatical functions happen to be syntactically encoded in the surface forms of English. In other languages, the SUBJ, OBJ, and OBL_{AG} functions are syntactically expressed through morphological case markings and not through distinctive positions in phrase structure. In such languages, predictably, the syntactic effects of Passivization involve, not the apparent movement of an NP, but an apparent change in morphological case (chapter 8). The syntactic mechanisms of LFG successfully generalize across radically different language types. For detailed analyses of varying language types within LFG, see Fassi Fehri 1981, Neidle 1982, Simpson in preparation, Simpson and Bresnan 1982, Klavans 1982, Bresnan, Kaplan, Peters, and Zaenen 1982, and Levin, Rappaport, and Zaenen to appear, in addition to the chapters in this volume. In contrast, syntactic transformations do not generalize across these language types and therefore fail to provide a universal mechanism for representing grammatical relations. Recognizing this, transformational theories, including the government and binding theory (Chomsky 1981), simply make the unsatisfying assumption that language types may differ in their fundamental syntactic mechanisms.

Other explanatory advantages of the LFG theory are discussed in Bresnan 1978 and elsewhere in this volume. They include explanations for the boundedness and structure-preserving properties of relation-changing rules and the fact that rule interactions follow from the basic organization of the grammar rather than from stipulated conditions on rules or representations.

3. Implications for Process Models

Let us now consider the implications of these representational theories for process models. To establish a clear basis of comparison, we will make explicit several definitions and assumptions. First, for the reasons given in section 1, we assume that there is a *competence grammar* that represents native speakers' tacit knowledge of their language. Next, suppose that we are given an information-processing model of language use that includes a processor and a component of stored linguistic knowledge K. As a minimum, we assume that K prescribes certain operations that the processor is to perform on linguistic representations, such as manipulating phrases or assigning grammatical functions. K may also include other kinds of information as well. For example, it could contain indexing information that helps the processor quickly determine which representational operations it should perform in a given situation, frequency information to aid in making heuristic decisions, and so forth. We call the subpart of K that prescribes representational operations the *representational basis* of the processing model. (The representational basis is the "internal grammar" of the model.) Since not all components of the internal grammar are necessarily utilized in every linguistic behavior, we do not require all information in the representational basis to be interpreted by every processing model. However, we do require that every rule of the representational basis be interpreted in a model of *some* behavior; thus, the internal grammar cannot contain completely otiose rules. We can now say that a model satisfies the *strong competence hypothesis* if and only if its representational basis is isomorphic to the competence grammar.

We are now in a position to see how the choice of grammatical theories can affect the choice of process models. Natural languages frequently require highly intricate feeding relations among rules. This is as true of phonological and morphological rules as it is of the rules that determine syntactic relations. Examples are easy to construct. Consider (4).

(4)

a. Someone is giving too many gifts to politicians. (active)

b. Someone is giving politicians too many gifts. (active–Dative)

c. Politicians are being given too many gifts. (active–Dative–Passive)

d. Too many gifts are being given to politicians. (active–Passive)

e. %Too many gifts are being given politicians. (active–Passive–Dative)

f. There is someone giving too many gifts to politicians. (active–*There* Insertion)

g. There is someone giving politicians too many gifts. (active–Dative–*There* Insertion or active–*There* Insertion–Dative)

h. There are politicians being given too many gifts. (active–Dative–Passive–*There* Insertion)

i. There are too many gifts being given to politicians. (active–Passive–*There* Insertion)

j. %There are too many gifts being given politicians. (active–Passive–Dative–*There* Insertion or active–Passive–*There* Insertion–Dative)

In these examples, the mappings between the surface subjects and objects and their semantic arguments depend upon the feeding relations of the rules. If the Dative rule feeds the Passive rule, for example, the subject will correspond to a different argument of the verb *give* than if it does not or if the reverse is the case, and if any rule feeds *There* Insertion, the subject will correspond to no argument of *give*. Similar intricacies appear elsewhere in English and in other languages.

In transformational theories of syntactic representation, as we have seen, the predicate argument structure to surface form mapping is performed by phrase structure computations. In such theories, grammatical relations which are encoded by intricate feeding relations must be represented by an ordered sequence of structural computations which is the transformational derivation.[2]

In the lexical–functional theory, in contrast, the predicate argument structure to surface form mapping is performed by lexical operations and surface function annotations. The feeding relations are expressed by the composition of operations that derive lexical entries. Thus, given the LFG representations, no special phrase structure computations are required to decode the grammatical relations which arise from these intricate feeding relations. As pointed out in Bresnan 1978, even the lexical computations are not required in generating sentences, since such lexical rules, as long as they have a finite output, can always be interpreted as *redundancy rules,* and in fact there is some independent motivation for doing so (Jackendoff 1975, C. L. Baker 1979). As such, the rules could be applied to enter new lexical forms into the mental lexicon, and the derived lexical forms could subsequently simply be retrieved for lexical insertion rather than being rederived (Miller 1978). The search space for any lexical form is bounded, since only a finite number of these lexical rules can be defined in the theory (chapter 10).

Consider now how we might model the process of decoding the grammatical relations of natural language sentences. Let us suppose that the model derives a representation of the predicate argument structure of a sentence from a representation of its surface form together with a store of linguistic knowledge structures represented by a grammar—the "internal grammar" of the model. Let us further suppose that the model satisfies the strong competence hypothesis. If the linguistic knowledge that is required in this process is represented by a transformational grammar, grammatical relations which are encoded by intricate feeding relations must be decoded by an ordered sequence of phrase structure computations (corresponding to the sequence of transformations in a transformational derivation of a sentence). The complexity of the decoding process is then a direct function of the length of the transformational derivation, a hypothesis known to psychologists as the *derivational theory of complexity.*[3] Despite important early work in its support (Miller 1962, Miller and McKean 1964), psychologists now appear to be universally agreed that this theory is false (Fodor, Bever, and Garrett 1974).

If the linguistic knowledge that is required in decoding grammatical relations is represented by an LFG, however, a range of possible models with very different complexity metrics can be obtained. The intricate syntactic feeding relations that can change grammatical relations will now be represented by sequences of *lexical,* not syntactic, rules. If these lexical rules are interpreted as redundancy rules, then all of the possible function–argument correspondences will already be expressed in the lexicon by finite sets of lexical forms. Thus, the lexical entry of a verb like *give* will include passive and dative lexical forms, which were derived by the lexical operations when the active form of the verb *give* was first entered in the lexicon. Because the outputs of these lexical redundancy rules already exist in the stored knowledge component of the model, the processor need not perform the operations specified by these rules as the model decodes the grammatical relations of a sentence. If we further assume that all lexical forms are accessed in parallel, then in this model the complexity of syntactic computations will not reflect the complexity of the lexically encoded feeding relations, but only the complexity of the analysis of the surface phrase structure tree. In this model, then, the relative complexity of active, active–Passive, active–Dative–Passive, and other sentences of (4) will depend only on the relative complexity of their surface structure analyses. Let us refer to this as *Model I*.

We can derive an interesting variant of Model I if we further suppose that semantic interpretation is interleaved with the syntactic analysis—an assumption compatible with LFG grammars because of their order-free composition property (see section 4 and Halvorsen 1982). Assume that the arguments of lexical forms can be accessed both by the *functions* they select (SUBJ or OBJ, etc.) and by their *semantic properties*. Then the process of extracting the grammatical relations of a sentence could be facilitated by semantic information that differentiated the arguments of a lexical form. For example, if we are analyzing a sentence like *The kite was admired by the girl,* and we know that *the kite* denotes an inanimate object, then we can use that semantic information to match *the kite* to the correct (theme) argument of *admire* (since kites can be admired but cannot admire). But if we are analyzing a sentence like *The woman was admired by the girl,* this semantic accessing of the lexical predicate argument structure would not facilitate the analysis (since women can both admire and be admired). In this model, *Model II,* the complexity of syntactic analysis could be reduced by asymmetries in the semantic structure of the predicates. Thus, Model II would produce "nonreversability" effects like those reported by Slobin 1966.

A *Model III* could be designed to accord with the derivational theory of complexity. Such a model can be obtained by using the lexical rules on-line to generate forms for lexical insertion. In other words, the processor performs the operations specified by the lexical rules as the model decodes the grammatical relations of a sentence. Thus, in order to analyze an example like (4c), *Politicians are being given too many gifts,* the Dative and Passive lexical rules would be applied to the active lexical form *give* and the outputs matched with the syntactic analysis. Thus, Model III would make the same complexity predictions as a transformational grammar-based processing model satisfying the strong competence hypothesis.

Could a transformational grammar-based model make the same complexity predictions as the LFG-based Models I and II? In LFG, the lexicon can store the finitely many lexical forms which are the outputs of the lexical redundancy rules. But in transformational grammars, there is no store of the infinitely many phrase structure trees which are the outputs of syntactic transformations. (This is clear if one notes that the passivized object NP itself can be of arbitrary depth, since NPs are recursive phrase structure categories.) Therefore, in order to decode the grammatical relations of sentences in a transformational grammar-based model that satisfies the strong competence hypothesis, the processor must perform the operations on phrase structure trees that the transformational rules specify. Because of the feeding relations among transformations, derivational complexity effects will result.

Is there any way to get around this problem? One way of doing so would be to use transformations, not to transform the infinitely many phrase structure trees that the phrase structure rules generate, but to transform the finitely many subcategorization frames of lexical entries. The Dative, Passive, and other structure-preserving transformations would be eliminated from the syntactic component, and corresponding lexical rules would be added to the lexical component. This approach substitutes for the transformational grammar knowledge structures an alternative representation of linguistic competence, one whose empirical divergences from the transformational theory were originally explored in Bresnan's 1976a, 1978 extended lexical grammars. Thus, to adopt this approach is to drop transformational grammars as the competence theory in favor of a precursor to the LFG theory.[4]

Another way of achieving the complexity predictions of an LFG-based processing model using a transformational grammar knowledge representation has been proposed by Berwick and Weinberg to appear. They argue that if one changes the underlying computational architecture of the transformational grammar-based model, the derivational theory of complexity no longer holds. What they actually show is that by using a finite amount of parallel processing, permitting parsing actions to be executed simultaneously, the analysis of a simple passive structure can take the same amount of time as the recognition of the corresponding active structure. In other words, the elementary operations of Passivization—the NP movement and the attachment of the passive verb to the analysis tree—can be accomplished simultaneously in unit time. To successfully mimic the results of the LFG-based model, however, they would have to demonstrate that the operations specified by *all* of the sequences of standard transformational operations—Dative, Dative–Passive, Dative–Passive–*There* Insertion, etc.—can be executed in unit time. But because these operations are in true feeding relationships, in which the necessary input of one operation is created by the output of another, it is simply not possible to execute them in parallel.

For every finite feeding sequence of transformations $t_1, \ldots, t_n$, it is possible to construct a new composite operation T_{1-n} that takes structures from the input form specified by t_1 directly to the output form specified by t_n. As long as there are only finitely many of the original transformations and as long as they are cyclic, bounded rules, this kind of recoding is possible. In this way the complexity effects of the LFG-based Models I and II may be simulated. But *the resulting model no longer satisfies the strong competence hypothesis;* that is, it is no longer true that the transformational grammar specifies the only operations that the processor can perform on linguistic representations. For example, the Dative transformation, which formerly played a role in the derivations of sentences like (4b,c,e,h,j), no longer exists in the model, and the same is true for the Passive transformation, along with all of the other cyclic transformations. Instead there is a new set of transformations, $T_1, T_2, \ldots, T_{1-2}, T_{2-1}, \ldots$, etc., which have been motivated solely on the grounds of reducing computational complexity in this model. The composition procedure loses information in the sense that, given the output of composite rules, it is in general not possible to identify what the original rules were.

It might be suggested that the grammar of composite operations, which does reduce comprehension complexity, is in fact the better grammar, the one that more accurately models the native speaker's competence. It is difficult to support this conclusion, however, since this grammar trades simplicity of comprehension for complexity of linguistic descriptions: sentences exhibiting complex grammatical relations can no longer be classified into simple syntactic types solely on the basis of their derivations. For example, the passive morphology of English could formerly be correlated with a characteristic component of the mapping that encodes grammatical relations (namely, movement from object position to subject position). In the new grammar, this correlation disappears, and passive morphology is associated with an arbitrary set of NP movements. For these reasons, the new grammar could actually increase the complexity of language acquisition.

We see, then, that a transformational grammar-based process model cannot achieve the results of the LFG-based Models I and II without relinquishing the strong competence hypothesis. This is because the basic representational assumptions of transformational grammar require that complex grammatical relations be encoded by complex transformational derivations. Grammatical relations cannot be prestored in any component of the competence grammar, since it is a finite knowledge representation and the syntactic transformation is an operation on infinite sets of syntactic phrase structure representations. Nor can grammatical relations be computed in parallel, because of the feeding relationships that grammatical-relation-mapping rules enter into. Thus, the derivational theory of complexity is implicit in the fundamental representational assumptions of transformational grammar. To the extent that there is evidence against the derivational theory of complexity, this can be taken as evidence against the psychological reality of transformational grammars.

These conclusions do not mean that we must weaken the strong competence hypothesis or give up the goal of a unified theory of the mental representation of language. There are systems based on very different representational assumptions, such as LFG, which are realizable in more explanatory models of language processing, satisfying both the strong competence hypothesis and the substantive constraints of a theory of Universal Grammar.

4. Theoretical Constraints on Knowledge Processing

Just as theoretical constraints on knowledge representations affect the classes of computations that can construct and interpret those representations, so also the converse is true: theoretically motivated constraints on knowledge processing can affect the choice of knowledge representations. In this way, assumptions underlying models of how linguistic knowledge is processed can determine the choice of grammars.

In this section we discuss an abstract computational theory of syntactic processing which provides a conceptual framework for constructing models of various linguistic behaviors. Since all such behaviors (beyond the word-level ones) involve a mapping between strings and grammatical relations (the so-called syntactic mapping), we will proceed by postulating various properties that such a mapping might have. These assumptions imply certain conditions that the representational basis of a processing model must satisfy, and thus, if the strong competence hypothesis is to be maintained, they limit the formalisms that are suitable for linguistic theory.

The *syntactic mapping problem* is the problem of computing, for any human language, the grammatical relations of any string of words of that language.[5] This is an extremely difficult problem—first, because of the complex, many-to-many relation between the sentences of any natural language and their grammatical relations, and second, because of the radical variations in surface form across languages. Yet we know that the human brain instantiates a general solution to this problem, for despite the exotic variety of the world's languages, any normal child is capable of mastering any language. To solve the problem in a general way, we will therefore introduce theoretical assumptions about the nature of the computation which will serve to constrain the set of admissible solutions. These mapping constraints are motivated by general properties of the computational problem that linguistic knowledge processes must solve, by properties that are intuitively true of the human mind (the one entity that we know is capable of executing those processes), or by properties that we believe a satisfying scientific theory of these phenomena must possess. As is true of the basic assumptions in any scientific theory, the validity of these postulates is not susceptible to direct empirical evaluation. Rather, they stand at the center of a rich deductive system which has testable consequences at its empirical frontier. These central theoretical constraints will be accepted to the extent that the remote empirical predictions of models that embody them are confirmed. The last chapters of this volume describe actual processing models that do embody these theoretical assumptions and whose specific predictions are the subject of ongoing experimental investigations.

As general conditions on the syntactic mapping problem, each grammatical string must be paired with a phonological interpretation and the representation of its grammatical relations must be paired with a semantic interpretation; a string is syntactically unambiguous if and only if the mapping assigns it a unique representation of grammatical relations; and a string is ungrammatical if and only if the mapping assigns it no well-formed representation of grammatical relations. The essential properties of all generative grammars reflect certain theoretical constraints on the set of possible processes that compute solutions to the syntactic mapping problem. The constraints are *creativity* (the domain and range of the mapping are theoretically infinite), *finite capacity* (there is only a finite capacity for the knowledge representations used in the mapping), and, though not all generative grammars have turned out to satisfy this constraint, *reliability* (the mapping provides an effectively computable characteristic function for each natural language). Let us briefly review these constraints in turn.

Creativity

Prior to generative grammatical theory, most American structuralist linguists considered the aim of linguistics to be to establish procedures for discovering the grammatical structure of a given corpus of linguistic utterances as presented in fieldwork transcriptions. This would be equivalent to taking the domain and range of the syntactic mapping to be finite. However, Chomsky 1957, 1964, emphasizing the creativity of language use, argued that the most revealing way of looking at the problem is to take the domain and range of the mapping to be infinite sets of data and grammatical structures. While it is true that the entire body of language that any language user produces or comprehends in a lifetime is finite, the finiteness of the linguistic corpus appears to be an arbitrary restriction from the point of view of linguistic structure. The problem to be explained is how the language user can construct mental representations of grammatical relations for endless numbers of *novel* sentences. In principle—idealizing away from all limitations on lifespan, memory, and performance—the language user can identify the infinite set of grammatical sentences of the language. This is the justification for the *creativity constraint,* which requires the domain of the syntactic mapping to include all strings over the lexical vocabulary of the language, the range to include infinitely many representations as well, and the mapping to characterize the infinite set of grammatical sentences of the language; that is, a string is a member of the language if and only if the syntactic mapping assigns it a well-formed representation of grammatical relations.

Finite Capacity

The second constraint, *finite capacity,* was also proposed by Chomsky 1965. While the possible data and representations in the domain and range of the mapping are infinite, each language has only finite sets of elementary words and relations and there is only a finite capacity for representing and storing knowledge mentally. Given this constraint, the mapping must consist of the recursive composition of finitely many operations that can project a finite store of knowledge of a particular language onto infinite sets of data. Any mapping defined to be the recursive composition of finite elementary operations can be represented as a pair consisting of the specification in some formal language of the composition rules of those elementaries and a procedure which interprets those composition rules. For the syntactic processing case, the rules of composition are called *the grammar*, and we will use the notation m_G to name the procedure which applies to the rules. Thus, we can assume that the syntactic mapping decomposes into a grammar G of the particular language (which includes, for example, a set of syntactic patterns, or rules, of the language) and a procedure m_G which recursively matches the patterns, or applies the rules, of G to construct infinite sets of representations that relate surface strings of words to semantic predicate argument structures.

Reliability

The creativity constraint implies that our computation must characterize the infinitely many sentences of any natural language. The finite capacity constraint implies that this computation must decompose into a finite grammar G and a recursive procedure m_G for projecting G onto the infinitely many sentences of a language.These constraints are based upon the abilities of an ideal speaker of the language, abstracting away from the actual performance limitations of real language users. The same is true of the third constraint, reliability. Under the same idealization, a speaker of a language is regarded as a reliable arbiter of the sentences of his or her language. It is plausible to suppose that the ideal speaker can decide grammaticality by evaluating whether a candidate string is assigned (well-formed) grammatical relations or not. The syntactic mapping can thus be thought of as reliably computing whether or not any string is a well-formed sentence of a natural language. This motivates the *reliability constraint* that the syntactic mapping must provide an effectively computable characteristic function for each natural language.

The reliability constraint implies that the subset of data in the domain of the mapping for which there are well-formed grammatical relations is a recursive set (for the mapping must effectively compute whether an arbitrary string is grammatically well formed or not). Arguments for the recursiveness of natural language were first given by Putnam 1961:41, who concluded:

> [T]he self-containedness of language [i.e., the classifiability of sentences out of context], the usability of nonsense sentences, and the relative universality of grammar intuitions within a dialect group, taken together support the model of the [human] classifier as a Turing machine who is processing each new sentence with which he is provided according to some mechanical program.

In other words, independently of knowledge of specific context, even independently of meaningfulness, speakers can reliably classify sentences as grammatical or ungrammatical, and they do so with a high degree of consistency across individuals within a dialect group. This convergent behavior suggests that classification of strings as grammatical or ungrammatical is based on an automatic procedure. Even though actual speakers may hesitate or conflict in their judgments of grammaticality, it is hypothesized that if extraneous and confusing sources of variation are controlled and if the speakers are given more time and memory, their judgments will approach the behavior of the ideal classifier in the limit.

Putnam also pointed out the consequent need to constrain transformational grammars if they were to be taken as characterizations of the mentally represented linguistic knowledge that is used to classify sentences. Although Chomsky 1965:208, note 37, assumed that constraints on transformational grammars such as the recoverability of deletions restricted the generative capacity of transformational grammars to recursive languages, Peters and Ritchie 1973 disproved this conjecture. Chomsky has increasingly downgraded the importance of constraints on generative capacity, relying on the idea of an evaluation metric to limit the class of available grammars in a learning model (Chomsky 1977b) or on an analogy between language acquisition and biological maturation (Chomsky 1980b). Despite this deemphasizing of grammatical processing in favor of grammatical acquisition, it nevertheless appears to be a precondition for any satisfying computational theory of syntactic processing that it account for the capability of the ideal speaker to classify natural language sentences. Peters and Ritchie 1973 pointed out that the actual generative transformational grammars that had been written for natural languages were in fact recursive, and subsequently both Peters 1973 and Wasow 1978 discovered properties that would constrain standard transformational grammars to generate only recursive sets.

Psychologically-based arguments for the recursiveness of natural language—including Putnam's—have been criticized on the grounds of English speakers' reactions to well-known "garden path" sentences like (5) and (6).

(5)
The canoe floated down the river sank.

(6)
The editor the authors the newspaper hired liked laughed.

For example, Matthews 1979:212 claims that the classification of examples (5) and (6) as grammatical or ungrammatical can be predictably affected by their position in a list with sentences such as (7) and (8).

(7)
The man (that was) thrown down the stairs died.

(8)
The editor (whom) the authors the newspaper hired liked laughed.

Sentences (5) and (6), he asserts, will typically be classified as ungrammatical if they precede sentences similar to (7) and (8), but grammatical if they follow them. He claims that this fact is inexplicable in terms of memory and other performance limitations on an idealized effective procedure for classifying sentences. Matthews 1979:217 argues that the same evidence also counts against the view that real native speakers instantiate an effective procedure when pairing sentences with their underlying structural descriptions during comprehension. Following Fodor, Bever, and Garrett 1974, he assumes that these phenomena are to be explained by postulating a collection of agrammatical heuristic strategies that directly pair certain simple types of sentences with their structural descriptions; but then, he concludes, there is no need to assume that the language user instantiates an effective procedure in comprehending or classifying the grammatical sentences of the language.

In fact, however, we can draw precisely the opposite conclusion from such examples. Recent work on syntactic perception (reported in chapter 11 of this volume) has argued that the heuristic strategies approach fails to explain the ability of speakers to *recover* from garden paths. In general, if speakers are given sufficient time and resources (such as pencil and paper), they are able to recover all of the grammatical analyses provided by the competence grammar. The competence-based model of syntactic processing given in chapter 11 can explain both the ideal behavior (the recovery of all of the grammatical analyses) and the actual behavior (the experience of conscious garden paths) in terms of resource limitations on a recursive procedure for constructing grammatical relations.

In particular, note that examples (5) and (6) are locally ambiguous in their initial segments; with respect to the competence grammar of English, (5) permits a local initial analysis either as a simple sentence (*The canoe floated down the river*) or as a reduced relative (*The canoe* [*which was*] *floated* . . .), and (6) permits a local initial analysis either as a sequence of conjoined NPs (*The editor, the authors, the newspaper,* [*and*] . . .) or as a center-embedded relative clause like (8). General principles of local ambiguity resolution which are incorporated into an effective procedure for analyzing all and only the grammatical sentences of natural language can easily explain the initial false analyses assigned to these sentences (see chapter 11). The observed performance difficulties that many speakers have in recovering the true analyses without contextual clues can be explained by limitations on working memory and other computational resources during the grammatical analysis procedure.[6]

We see, then, that speakers' reactions to examples like (5)–(8) would actually follow from specific performance limitations on an idealized effective procedure for analyzing or classifying sentences. Although Matthews 1979 repeats the well-known arguments due to Fodor, Bever, and Garrett 1974 that a psycholinguistic theory of comprehension cannot plausibly incorporate a transformational grammar as the mental representation of linguistic knowledge, we have already observed that these arguments do not impugn the competence hypothesis, and hence they do not undermine the motivation for imposing the reliability constraint on the syntactic mapping.

A very different source of arguments against the recursiveness of natural language can be found in considerations of semantic interpretation. The most ingenious such argument is due to Hintikka 1974, 1979. Hintikka found that the acceptability of an infinite set of natural language sentences involving the quantifiers *any* and *ever* depends upon their semantic equivalence to other sentences. He argued that there is no effective procedure for determining the acceptability of these sentences, because (assuming that the semantic equivalence of natural language sentences is adequately modeled by logical equivalence) there is no effective procedure for testing the logical equivalence of even first order quantification sentences. This conclusion implies that the set of acceptable sentences is not only nonrecursive, but not even recursively enumerable, and hence well beyond the power of any generative grammar or Turing machine to specify.[7]

Such arguments for the nonrecursiveness of natural language enforce an important point. The automatic mental construction of grammatical relations must *not* require an evaluation of truth in the world, or logical truth conditions. The grammatical relations of the unacceptable sentences cited by Hintikka are in fact as easily perceived as those of other sentences. For example, his unacceptable *Louise ever kisses me* can be assigned perfectly well-formed grammatical relations, which specify that *Louise* and *me* are the subject and object arguments of a dyadic tensed active verb *kiss,* modified by a temporal quantifying adverb *ever*. This fact is unsurprising, since well-formed grammatical relations can be assigned even to nonsensical sentences like *Louise pilsely grisps a blawn neddle*. If such examples are judged "unacceptable" or "incorrect," this must then be attributed to some property of their nonsyntactic interpretation—whether it be the non-effectively-computable semantic property formulated by Hintikka or another, as yet undiscovered, property. If Hintikka's argument is correct, then semantics must diverge from syntax in a fundamental way, as he observes. The method of recursively characterizing the structure of natural language by means of a generative grammar may in fact be incapable of characterizing the semantics of quantification.

We can conclude that the reliability constraint on the syntactic mapping problem introduces no incoherence: "grammaticality" can be a recursive concept so long as it is not a function of truth in the world.

The above constraints—creativity, finite capacity, and reliability—are familiar from early work in generative grammar. Note that none of them is logically necessary for purposes of grammatical description. Their motivation comes rather from an abstract computational theory of syntactic processing. Exactly the same is true of two further constraints that we will now add: *order-free composition,* requiring that the grammatical relations that the mapping derives from an arbitrary segment of a sentence be directly included in the grammatical relations that the mapping derives from the entire sentence, independently of operations on prior or subsequent segments, and *universality,* requiring that the mapping incorporate a universal procedure for constructing representations of grammatical relations.

Order-free Composition

The constraint of *order-free composition* is motivated by the fact that complete representations of local grammatical relations are effortlessly, fluently, and reliably constructed for arbitrary segments of sentences. For example, the sentence fragments in (9) are immediately associated with grammatical representations which provide the same kinds of information as those of complete sentences.

(9)
a. There seemed to . . .
b. . . . not told that . . .
c. . . . too difficult to attempt to . . .
d. . . . struck him as crazy . . .
e. What did he . . .

In (9a) *there* is a grammatical but not a "logical" subject of *seemed* and is grammatically related to an infinitival complement of *seemed*. In (9b) the unexpressed subject of *told* can be the passivized indirect object of the verb, and *that* can be the complementizer of an unexpressed sentential complement corresponding to the "logical" object; other grammatical interpretations arise as well. The fragment in (9c) is syntactically ambiguous in the same way as the full sentence *It was too difficult to attempt to understand,* where the subject of *difficult* can be identified either with the clause *to attempt to understand (something)* or with an object within the clause (such as the object of *understand*). In (9d) *him* is the grammatical and the "logical" object of *struck,* and *crazy* is the predicative complement of the unexpressed subject of *struck*. The fragment in (9e) is interpreted as part of a question in which *what* is grammatically related to an unexpressed clause of which *he* is the grammatical subject. In each case, the grammatical relations for the fragment are complete in the sense that they express information that can be directly incorporated without change into the grammatical relations of some full sentence.

It is a fact, then, that given a sentence fragment one can always pick, entirely out of context, a *possible* reading from the finite set of grammatical alternatives, while *impossible* readings are excluded. It is no more possible to interpret the fragment . . . *easy to justify* . . . so that the subject of *easy* is also the subject of *justify* than it is possible to interpret *John's actions* as the subject of both *easy* and *justify* in the ungrammatical sentence **John's actions are easy to justify yours*. In contrast, the fragment . . . *unlikely to justify* . . . allows just this possibility, exactly as in the grammatical sentence *John's actions are unlikely to justify yours*. This fact suggests a very strong natural constraint on the syntactic mapping problem: the composition of operations as performed by m_G in the syntactic mapping must be order-free, in the sense that from an arbitrary sentence fragment, the mapping must derive—independently of operations on prior or subsequent segments—a set of possible grammatical relations, any of which can be directly included in the grammatical relations that the mapping derives from some sentence containing the segment. In other words, we have the following *order-free composition constraint* (referred to as *bounded context parsability* in Bresnan 1979a):

(10)
Order-free Composition Constraint
Under any valid interpretation of a string *xsy* consisting of adjacent segments *x*,*s*,*y*, the grammatical relations in $m_G(s)$ must be included in $m_G(xsy)$.

Note that this constraint allows for the fact that, in general, segments of a sentence—viewed as strings of words—have more possible grammatical relations in isolation than within the sentence. Example (11) illustrates this point:

(11)
a. . . . her candy killed Mother See.
b. A man who hated her candy killed Mother See.

The string of words in (11a) can be interpreted so that *her candy* is the subject of *killed* and *Mother See* is the object of *killed;* but the subject–verb relation between *her candy* and *killed* is not included in the grammatical relations of the whole sentence (11b). However, (11a) can also be interpreted as a sequence of two unrelated subsegments, *her candy* and *killed Mother See,* both of whose local grammatical relations are then directly contained in the grammatical relations of the entire sentence (11b). Note that in the limiting case, all of the words in a certain segment may be unrelated, as in (12a):

(12)
a. . . . to by for . . .
b. The one that he should be spoken to by for God's sake is his supervisor.

In this limiting case, the representation of grammatical relations of the segment can simply be identified with the lexical representations of the unrelated words. In sum, the order-free composition constraint asserts that sentential context may determine the *choice* of one of a set of locally computed grammatical relations for a segment, but the computation of grammatical relations for a segment may not involve the computation of the grammatical relations of the context. In other words, this postulate severely constrains the role of context-sensitive operations in the syntactic mapping.

The order-free composition property is equivalent to requiring that the mapping m_G be *monotonic*.[8] In other words, the mapping m_G from segments to sets of grammatical relations must preserve inclusion: under any valid interpretation, if one string of words (s_1) is contained in another (s_2), then the (compatible) grammatical relations of the one ($m_G(s_1)$) must be included in those of the other ($m_G(s_2)$). Intuitively, monotonic functions are incremental. Thus, the monotonicity constraint requires the operations of m_G to add increments of local information from the string to the global representation of information and not subtract from the global representation.

Universality

The fifth assumption that we shall adopt to constrain the syntactic mapping problem is the *universality* of the mapping procedure. It is plausible to assume that the procedure for grammatical interpretation, m_G, is the same for all natural language grammars *G;* that is, there is a universal m_U such that for any *G*, $m_G = m_U$. This constraint is motivated by the universality of the system for mentally representing natural language. We assume that the grammar *G* representing mature knowledge of a particular language is induced by a universal learning function. As in chapter 10 of this volume, we assume that data for the learning function are sets of pairs (s, r), where *s* is the perceived surface string of words and *r* is the mental representation of meaningful grammatical relations of *s;* the learning function maps these data onto hypothesized grammars. To test a hypothesized grammar G^*, there must be some universal effective procedure for constructing the mental representations *r* of lexical strings *s* given G^*; call it m_U. While it is conceivable that this universal procedure is different from the one that the language learner normally uses in comprehending language (m_G), the simplest, strongest, and most plausible assumption is that the procedures are the same. If so, the acquisition process must depend just as much upon the ability to comprehend language as the growing comprehension of language depends upon the acquisition process. The interaction of these two information-processing systems, linguistic acquisition and linguistic comprehension, motivates the universality constraint that $m_G = m_U$ for all G.

5. Implications for Syntactic Knowledge Representation

These processing constraints on the syntactic mapping problem—*creativity, finite capacity, reliability, order-free composition,* and *universality*—impose important limitations on the possible forms of syntactic knowledge representation, ruling out many possible systems of grammar—even apparently descriptively adequate ones—as systems of the mental representation of language. The creativity and finite capacity constraints require knowledge representations consisting of finite systems of explicit rules that are capable of characterizing infinite sets of natural language sentences. The reliabilty constraint further requires that these rule systems specify languages within the class of recursive sets, thereby ruling out unrestricted rewriting systems (Chomsky 1963), arbitrary augmented transition networks (Woods 1970), and arbitrary standard transformational grammars (Peters and Ritchie 1973) as representations of linguistic competence. The remaining two constraints impose still stronger conditions on possible grammatical theories.

Because of these constraints, the syntactic mapping decomposes into a finite grammar G which generates the language and a finite set of operations m_G which recursively apply the rules of G to extract the meaningful grammatical relations of the language. In general, it is the syntactic (as opposed to the lexical) rules of generative grammars which specify infinite sets of structures; hence, it is these rules that m_G must recursively apply in extracting the grammatical relations of a string. Recall that the order-free composition constraint states that under any valid interpretation of a string xsy consisting of adjacent segments $x,s,y,$ the grammatical relations in $m_G(s)$ must be included in $m_G(xsy)$. Clearly, grammars whose syntactic rule components consist only of context-free phrase structure rules satisfy this requirement; in such grammars, a noun phrase is a noun phrase regardless of its external context. For these grammars, the order in which the segments are analyzed in the mapping from the data to the representation will therefore be irrelevant. However, non-context-free syntactic rules do not in general have this property; such rules may alter the analysis of the grammatical relations of a phrase according to the analysis of its context, and the resulting representation may therefore depend upon the order of application of these "context-sensitive" rules to different segments. Consequently, if m_G has the order-free composition property, the grammar G must be such that information about the surface-to-predicate argument structure associations given as the output of m_G is independent of the order in which any "context-sensitive" nonlexical grammatical rules are applied in the derivation of the representation from the data.[9] For reasons we have already seen (section 3), transformational grammars do not have this property. Transformational grammars explicitly operate by changing the grammatical relations for each region of the string. It is of course possible to modify transformational grammars so that local grammatical relations in the transformational derivation are always preserved in surface forms, and this formal effect is partially achieved by the Projection Principle (note 2). However, for reasons already discussed in note 2, transformational grammars with this provision still do not possess the order-free composition property: the order dependencies merely migrate to other components of the mapping onto surface forms. Augmented transition networks also lack the order-free composition property, for they enable register reassignments to modify previously determined grammatical relations.

The universality constraint implies that grammatical relations must be encoded in a form which allows a universal decoding process for all natural languages. One might think that one of the various well-known algorithms for context-free phrase structure parsing might provide just such a universal decoding process for natural languages, but recent work in theoretical linguistics forces us to reject this possibility. Bresnan, Kaplan, Peters, and Zaenen 1982 have shown that there is no context-free phrase structure grammar that can correctly characterize the parse trees of Dutch. The problem lies in Dutch cross-serial constructions, in which the verbs are discontinuous from the verb phrases that contain their arguments. The phenomenon of "discontinuous constituents"—that is, noncontiguous constituents defining single functional units—is pervasive in natural language, occurring in its most extreme forms in Australian aboriginal languages (Hale 1979, Pullum 1982). The results of Bresnan, Kaplan, Peters, and Zaenen 1982 show that context-free grammars cannot provide a *universal* means of representing these phenomena.

The processing constraints on the representation of syntactic knowledge may now seem too strong. On the one hand, *G* must have the order-free composition property of context-free phrase structure grammars, but on the other hand, *G* cannot be a context-free phrase structure grammar because context-free grammars cannot universally characterize the correct surface constituent structures of natural languages. Nevertheless, a solution to the syntactic mapping problem does exist. For any language *L*, where *S* is the set of strings over the lexical vocabulary of *L*, let us take *G* to be a *lexical–functional grammar* for *L*, and *R* (the representations of grammatical relations) to be the set of *functional structures* of the language as defined by *G*. Then a map $m{:}(S,G) \rightarrow R$ exists which satisfies all of the given constraints. This result is based on the mathematical characterization of lexical–functional grammars in chapter 4 of this volume. It provides the foundations of a computational theory for investigating the mental processes that construct representations of grammatical relations.

6. Conclusion: An Introduction

The work in this volume represents a new approach to the study of the mental representation of grammatical relations, based on the cognitive theory outlined above. The universality of the LFG theory of grammatical representation is supported by the chapters on grammatical relations in English, French, Russian, Icelandic, and Malayalam (a Dravidian language). The order-free composition property is illustrated by the f-structure solution algorithm given in chapter 4. Finally, the suitability of the LFG theory for modeling syntactic processes is shown in the last three chapters, which are psychological studies of competence-based models of language acquisition, comprehension, and production.

Notes

This material is based on work supported in part by the National Science Foundation under Grant No. BNS 80–14730 and in part by the Cognitive and Instructional Sciences Group, Xerox Corporation Palo Alto Research Center.

1. Some philosophers consider this controversial. Matthews 1982, for example, argues against this view as a form of "computational reductionism." However, he overlooks the fact that the attribution of an internal representation to a computational device may itself involve intensional descriptions. To take an example due to Brian Smith, the description of a machine as running LISP

cannot be reduced to purely extensional terms, because what counts as an instance of the LISP language depends on the functioning of the program and not on any particular electronic or mechanical configuration that realizes it.

2. This holds for true feeding relations—relations in which the output provided by each operation creates the necessary input for the next operation in a "cascade." If all syntactic rules could be applied simultaneously, there would be no true syntactic feeding relations. Chomsky 1981 has attempted to maintain a new representational principle, the *Projection Principle,* which—by making the transformational derivation "transparent" in S-structure—would in effect permit simultaneous application of syntactic transformations. However, there are grammatical phenomena which seem to force a structuralist theory into cascades, at one level of representation if not another. For even if the Projection Principle could be maintained for the mapping between D-structure and S-structure, the mapping from S-structure to surface form will itself fail to satisfy the principle of simultaneous applicability of rules; this problem is particularly obvious in the case of radically nonconfigurational languages (Klavans 1982, Simpson and Bresnan 1982), where the mapping from configurational to nonconfigurational form would involve operations of "scrambling," constituent breakup, and deletion, which produce different surface forms depending on their order of application. Thus, the Projection Principle does not eliminate feeding relations among transformational operations in the derivation of surface forms.

It is interesting to note that combining the Projection Principle with other assumptions in current transformational theory leads to empirically wrong results. First, as Chomsky notes, it is inconsistent with many analyses within his own framework, which remain to be reconciled with it. Second, that framework lacks any well-motivated analysis of a number of constructions that appear inconsistent with the Projection Principle, such as dative-alternation constructions and noncompositional idiom constructions (on the latter, see chapter 1 of this volume and Rothstein 1981). Third, recent work on "raising" constructions (chapter 5), derived nominals (M. Rappaport 1980), subcategorization (Grimshaw 1981), and small clauses (Williams 1982, Bresnan 1982, Neidle 1982) has brought forth evidence that optimal grammars for these phenomena will violate the Projection Principle.

3. There is a slight subtlety here. The statement that such a model will have computational complexity in proportion to the length of the derivation is a fact about that kind of model. The derivational theory of complexity involves another small step, the claim that complexity of the decoding process in the model will correspond to actual experienced cognitive load under some straightforward assumptions about psychological costs.

4. There are certain differences between a lexical–transformational theory and the LFG theory. For example, in the former the transformations will operate on structural categories—NP, V, PP, etc.—in the strict subcategorization frames, rather than on the functions SUBJ, OBJ, etc., of the LFG theory. This difference is in fact a disadvantage for the former theory; in recent work, Grimshaw 1981 has given linguistic evidence that lexical items subcategorize for *grammatical functions,* not for phrase structure categories.

5. Recall that the term *grammatical relations* is used here in a theory-neutral way to refer to the associations between the surface constituents and the semantic predicate argument structure of a sentence. Thus, the grammatical relations of a sentence can be represented by a (semantically interpreted) constituent structure tree in phrase structure grammars, by a pair of deep and surface phrase structure trees in transformational grammar, by a pair of initial and final relational strata in relational grammar, and by an f-structure in LFG.

6. Specifically, according to the perceptual theory of chapter 11, the true analysis of (5) would require morphosyntactic reanalysis of *floated* from a past tense verb to a passive participle after the grammatical analysis of the initial hypothesized sentence that contains it has been completed; the difficulty of recovering the correct analysis can therefore be explained by limitations on memory for morphosyntactic categorizations during the syntactic analysis of sentences. Similarly, the true analysis of (6) requires recognition of a center-embedded syntactic structure, which could impose an excessive burden on working memory during the grammatical analysis of a sentence, as Church 1980 argues.

7. See Hintikka 1979, 1980 for counterarguments to various objections to his argument that have been raised in the literature, such as Chomsky's 1980b proposal that a constraint on "Logical Form" accounts for Hintikka's generalization in an effectively checkable way.

8. Recall that a function $f:X \rightarrow Y$ is called *monotonic* if X and Y are ordered sets and $X_1 < X_2$ implies that $f(X_1) \leq f(X_2)$.

9. This "order-free" restriction does not refer to the absence of an "extrinsic" ordering on these rules (as in Chomsky and Lasnik 1977), but rather to the stronger requirement that the order of application of these rules in the derivation of the representation of grammatical relations from the data cannot alter that representation.