JANUA LINGUARUM

STUDIA MEMORIAE
NICOLAI VAN WIJK DEDICATA

edenda curat

C. H. VAN SCHOONEVELD

Indiana University

Series Minor, 192/2

FORMAL GRAMMARS IN LINGUISTICS AND PSYCHOLINGUISTICS

VOLUME II

Applications in Linguistic Theory

by

W. J. M. LEVELT

1974

MOUTON

THE HAGUE · PARIS

PREFACE

Since the publication of Chomsky's *Syntactic Structures* (1957), linguistic theory has been strongly under the influence of the theory of formal languages, particularly as far as syntax is concerned. Investigations have dealt not only with the extent to which "pure" regular or phrase structure grammars can be used as models for a linguistic theory, but also with "mixed models", i.e., grammars to which a transformational component is added.

The most influential mixed model was that of Chomsky's *Aspects of the Theory of Syntax* (1965), but a number of other transformational grammars have been developed, such as dependency grammars and mixed adjunct grammars, with very different formal structures. Each of these grammars has its own specific advantages and disadvantages to offer. This volume presents a survey of the most important pure and mixed models, their formal structure, their mutual relations, and their linguistic peculiarities.

The formal structure of many transformational grammars has not been worked out in detail. This holds in particular for the syntax of *Aspects*. This fact may be considered as a simple esthetic fault, but on closer examination many deeper problems appear to be connected with it. The formalization of the grammar in *Aspects* has proven that the grammar in its standard form, as well as in later developments of it, cannot handle essential linguistic questions, such as that of the learnability of natural languages and the existence of a universal base grammar. A separate chapter deals with these problems.

Finally, attention will be given to the application of probabilistic grammars in linguistics.

This volume is concerned exclusively with linguistic questions. Psychological matters closely connected with them, such as the distinction between *competence* and *performance*, and the structure of linguistic intuitions, will be treated in Volume III.

Volume II presupposes acquaintance with the essentials of the material on formal grammar theory contained in Volume I. Cross-references to this Volume are made throughout the text. In its turn, the present Volume is preparatory to Volume III.

Nijmegen, June 1973

TABLE OF CONTENTS

1

LINGUISTICS: THEORY AND INTERPRETATION

In this volume we shall discuss the ways in which formal languages are used as models of natural languages, and formal grammars as models of linguistic theories. It was pointed out in Chapter 1 of the first volume that several concepts which have been incorporated into the theory of formal languages have lost something of the meaning they had in linguistics. As in the present volume our attention will be turned to natural languages, it will be necessary to re-examine such essential concepts as "sentence", "language" and "grammar", but in order to do so, we must first make a careful distinction between linguistic theory on the one hand, and its empirical domain on the other.

1.1. THE EMPIRICAL DOMAIN OF A LINGUISTIC THEORY

With respect to linguistic THEORY the problem of definitions mentioned above is trivial, and hardly anything need be changed in the given formal definitions. The formulation of linguistic theory must also aim at EXPLICITNESS and CONSISTENCY. The propositions of a linguistic theory (for example, "sentence x belongs to language L", "string y is a nominalization in language L", etc.) must be *explicit*, that is, it must be possible for anyone verifiably to deduce them from the principles of the theory, without making use of knowledge of the language in question obtained outside the theory. Consequently concepts such as "sentence", "language", etc. may not have the intuitive vagueness in the theory which they have in ordinary speech. As in formal language theory, a LANGUAGE in a theory of natural languages is a set of sentences, a SENTENCE is a string of elements (to be defined more fully), which satisfies the condition that it be generated by a GRAMMAR, which is a system of production rules, defined over a terminal and a nonterminal VOCABULARY. Such a theory is *consistent* if it does not lead to contradictions, that is, if it is impossible to deduce both a proposition and its negation within the theory. Obviously it is impossible to decide whether a theory is consistent or not if it is not explicit.

On one point, however, due to historical circumstances, a certain terminological ambiguity has come into being. As we have mentioned, a formal grammar is complete in that it is an exhaustive description of the sentences of a language and of their structure. In linguistics, the notion of grammar originally was used primarily for "syntax and morphology", the study of syntactic structure and the structure of words. Seen in this way, a linguistic grammar is not as complete as a formal grammar. A linguistic theory is not complete without phonological and semantic descriptions, concerning respectively the aspects of sound and meaning in the natural language. At first, applications of formal grammars to linguistic theory dealt exclusively with grammatical aspects in the original linguistic sense of the word: semantics was excluded and phonology was considered a more or less independent component and was studied separately. The impression was often given that the formalization thus obtained enjoyed the same degree of completeness in the description of natural languages as formal grammar in the description of formal languages. The notion of "grammar" became synonymous with that of "linguistic theory". It is still often used in this general sense, even now that the essential interest of semantics to linguistic theory is again emphasized in all quarters. If semantics is considered to be a subdivision of grammar, or if it is seen as indistinguishable from syntax, it remains something essential to grammar, and not something aside or apart from it. Some linguists maintain the original terminology, and use the word "grammar" only for syntax-and-morphology. There is no point in rejecting either terminology as "incorrect"; it is simply a question of scientific tradition and pragmatic considerations. On the basis of such pragmatic considerations, we shall use the word "grammar" only in its more limited sense, for it is quite evident that the clearest and most influential applications of formal language theory to linguistics have been related to syntax and morphology. Phonology has indeed been greatly formalized, but this has seldom been the result of *direct* applications of formal language theory. There have also been applications to semantics, but these have by far been neither as deep nor as extensive as those to syntax. As the subject of this volume is in fact applications to syntax, we shall, unless otherwise mentioned, use the notion of "grammar" as limited to syntax and morphology. Phonology and semantics will only rarely enter the discussion (semantics primarily in Volume III). Therefore we can, without risk of confusion, use the term "grammar" for "grammar in the limited sense", thus maintaining the connection with Volume I as far as possible.

Even within these limitations, however, it still holds that grammatical concepts to be used in linguistic theory should not have the vagueness (and wealth) of connotation which they might have in ordinary usage. Concepts must be defined entirely within the theory, and this holds as well for the concepts just mentioned as for other linguistic concepts such as "verb" and "noun phrase" which have not yet been discussed. They should all be fully defined within the formal description, and the relationships among them are established by the definitions and rules of the grammar. There is never reason for rejecting such concepts separately, but at most for rejecting the grammar as a whole.

But a linguistic theory is also an EMPIRICAL theory: it is designed to explain certain observable phenomena in verbal communication among human beings. As a whole, the observable phenomena with which a theory is concerned is called its EMPIRICAL DOMAIN. The size of the domain is not determined beforehand. Some verbal phenomena which seldom or never occur spontaneously might be elicited by various means; one can, for example, pose directed questions to the native speaker. These observable phenomena correspond with that which is called LANGUAGE in the theory. The theory is an abstract description of the kind of system a natural language is. This description must maintain a direct and understandable relationship with certain aspects of the observable linguistic phenomena. Thus the concept of "sentence" must in some way be related to that which is observable as an "utterance", the concept of "grammatical" (i.e. "generated by the grammar") ought to have something to do with the native speaker's judgment of which utterances are or are not "acceptable" or "good" English, Dutch, etc. The theoretical concept of "paraphrase" is perhaps related to the judgment of a hearer that a speaker means the same thing with two different utterances, and so forth. The network of theoretical concepts must be composed in such a way that the theory reflects the linguistic reality. In order to determine whether or not a theory satisfies this condition, as complete a description as possible must be given of the relations between the theoretical concepts on the one hand, and the empirical domain on the other.

In linguistics there are many cases in which the relations between theory and observable phenomena are simple and acceptable. We already know that various sentences (the relationship with "utterances" will be discussed later) must enjoy the status of "grammatical" in the theory. Any native English speaker will confirm that *the boy walks on the street* is good English, and that *on walks the street boy the* is not. Therefore a theory of the English language should be constructed in such a way that the grammar generates the first string and not the second. Such incontrovertible data are numerous enough to allow the construction of a linguistic theory and to test certain aspects of it, and we might hope that for less clear cases the theory itself might decide (for example, is *if he comes then she will go then he will come* a grammatical sentence?). This means that we can simply notice whether or not the sentence can be generated by the grammar which is composed on the basis of clear cases.

This method has the advantage that a maximum of theoretical construction can be realized with a minimum of troubling about procedures of data gathering and processing. The history of transformational grammars has shown that this kind of capitalization on immediate intuitive evidence is indeed an extraordinarily fruitful approach. The understanding of the structure of natural languages has probably never grown as rapidly as since the publication of Chomsky's *Syntactic Structures* (1957) in which this very programme was presented.

The reader should notice that this method is based upon a more or less explicit conception of the empirical domain of linguistics. This domain is in fact much less broad than that which we understood by "observable linguistic phenomena". This type of theory is descriptive of only some aspects of verbal phenomena, and the best generic name for those aspects is *linguistic intuitions.* Intuitions are observable in the form of *metalinguistic judgments,* that is, judgments the objects of which are verbal entities. Thus, from the point of view of theory, the objects of judgments on paraphrase relations and grammaticality are SENTENCES, the objects of judgments on cohesion within sentences are PHRASES, and so forth. It may be said that such linguistic theories concern metalinguistic data. It is the reflection of the native speaker on his own speech, which is formalized in the theory. This can also be stated in another way. One of the more noticeable characteristics of a natural language is that it is its own metalanguage; this means that by using a language one can speak about that language itself. The attention of theoretical linguistics (in the sense we have just mentioned) is principally directed toward those verbal activities whose object is the language itself. This restriction is scarcely necessary, and we shall repeatedly return to its attractive and unattractive implications in Volume III, Chapters 1 and 2. Yet until now the application of formal grammars to linguistic theory has in general presupposed such a limitation of the empirical domain.

In the following chapters we shall discuss the adequacy of various formal grammars as models, on the basis of linguistic intuitions as described above. This means that we suppose for the sake of discussion that the relationship between theory and observation is directly visible and acceptable, or in other words, that every reliable data processing procedure will support the intuitive insight.

1.2. THE INTERPRETATION PROBLEM

This, of course, does not close the discussion. Not only is the restriction of the empirical domain of a linguistic theory to linguistic intuitions far from always clear or attractive, but it is also the case that we do not always dispose of such direct evidence, and even when we do, some very essential questions remain to be answered. We speak of the INTERPRETATION PROBLEM when the relationship between theory and observation is itself the object of investigation. The question then becomes how the theory should be interpreted in terms of observable phenomena. We shall at this point mention three cases in which important linguistic interpretation problems occur.

The first case, in which the problem of interpretation makes itself increasingly felt, is that of the *use of linguistic intuitions* which we have just mentioned. It has slowly but surely become clear that it is not possible, on the basis of incontrovertible, directly evident data, to construct a theory so extensive that all less obvious cases can be decided upon by the grammar itself. It is becoming more and more apparent that decisions on very important areas of theory are dependent on very unreliable observations. In Volume III, Chapter 2 we shall mention an experiment (Levelt, 1972) which shows the frightful unreliability in judgments on grammaticality which occurs in modern transformational linguistic investigations. There is a tendency toward preoccupation with extremely subtle distinctions, not the importance, but rather the direct observability of which can seriously be called into question. Better methods must be developed for testing linguistic intuitions, and this is certainly a realizable possibility (cf. Volume III, Chapter 2, paragraphs 3 and 4). Moreover, a methodological tradition has existed in linguistics for some time, in which more value is given to some intuitions than to others. This tradition in general is based on implicit conceptions of the problem of interpretation. This can clearly be seen, for example, in the treatment of very long sentences. As we have noticed, there is a tendency to establish a relationship between theoretical grammaticality and the judgment 'sentence x is good English'. But there are sentences which an informant might call 'bad English' or 'not English' on the basis of circumstances which we could not easily formalize in a LINGUISTIC theory. Such is the case for very long sentences. They are unacceptable because our limited memory capacity makes it impossible for us to understand them. It seems undesirable to include an upper limit of sentence length into a grammar; it would be wiser to handle the limitation of length in a psychological theory as a systematic property of human memory, than in a linguistic theory as a systematic property of natural language.

The fact that such an intuition is disregarded by the linguist clearly shows that psychological presuppositions are implicit in the theoretical interpretation of certain linguistic observations. The example, moreover, is by no means incidental. Motivational, socio-psychological and other psychological factors must also be sifted out by the linguist in the interpretation of linguistic intuitions.

This first interpretation problem is thus of a psychological nature. One might wish that every linguist would fully recognize the role of psychological assumptions in the formulation of his theories. Unfortunately this is not the case. On the contrary, many linguists maintain that an adequate psychological theory of verbal behavior is possible only to the extent that linguistic knowledge is available (cf. Chomsky, 1965). The only consequence of this attitude is that at present psychological theory is implicitly working its way into the formulation of linguistic theory, instead of explicitly being taken into account and thus held in control. There is no reason to suppose that the common sense psychology of linguists is in any way better than the common sense linguistics of psychologists.

The second case of the interpretation problem is related to the first; it occurs when an adequate grammar is given and the question is asked as to whence the linguistic structures described by the grammar proceed. Before developing this matter, we must first clarify the notion of "adequacy". A grammar is called OBSERVATIONALLY ADEQUATE if it generates the sentences of a given language and only the sentences of that language. Because judgment on the observational adequacy of a concrete grammar can be given only on the basis of a concrete and therefore finite corpus of sentences (and at best a finite set of non-sentences), Chomsky calls a grammar observationally adequate when "the grammar presents the observed primary data correctly" (Chomsky, 1964). A grammar is DESCRIPTIVELY ADEQUATE if it gives a correct formalization of the linguistic intuitions. A descriptively adequate grammar is obviously also observationally adequate, because the decision as to whether or not a sentence belongs to a language is also based on an intuitive judgment of the native speaker. A descriptively adequate grammar moreover gives a correct reflection of intuitions about the structure of sentences, the relations between words and word groups, the relations among similar sentences, etc. There are always several possibilities of writing an observationally adequate grammar for a language. A sufficient number of examples of this have been given in volume I to make the point (see, for example, Figure 2.3 and the accompanying text).

In the same way a linguist can probably also dispose of various options for writing a DESCRIPTIVELY adequate grammar. One way of choosing from among several formulations is to compare them with the grammars of other languages. In a GENERAL linguistic theory, the elements common to all natural language, the general systematic properties of natural language, also called UNIVERSALS, will be described. There is thus only a limited degree of freedom in the description of a specific language. Neither for general linguistics nor for the description of individual languages are there generally accepted criteria for the choice of adequate grammars. We refer the reader to Volume I, Chapter 8 in which a number of problems are discussed which can appear in the comparison of grammars, and to this volume, Chapter 5 in which it is shown that certain very common suppositions on the form of a general linguistic theory cannot be tested empirically.

In spite of these important and unsolved problems concerning observational and descriptive adequacy, still a third form of adequacy can theoretically be imagined. Given a descriptively adequate general theory of linguistics, one can wonder by which psychological, biological and cultural factors this systematic structure of natural languages is determined. A linguistic theory which also answers these questions is called EXPLANATORILY ADEQUATE. This is clearly an extremely hypothetical field. Finally, there is as yet very little in linguistics which might be called adequate even from an observational point of view. Nevertheless, reasoning back from this abstraction, a number of important questions can be posed concerning the problem of interpretation, questions which lend themselves to empirical investigation even without disposing of complete and adequate grammars.

The explanation of the existence of certain grammatical properties must in the first place be brought back to an explanation of the linguistic intuitions with which it is connected. At present practically nothing is known of the nature of linguistic intuitions. We do not know how they come into being, how they are related to conceptions of one's own linguistic behavior in concrete situations, how they change under the influence of situational circumstances, what interaction there may be between them and perceptual aspects of time and place, how they are related to the systematic physical and social structure of the environment. It is also unknown whether or not, and if so to what extent such intuitions are trainable and how they develop in the growth of the child.

In the second place the explanation of a grammar must be brought back to the genetic question of how the language develops in the child. Research should be done on the means with which the child makes the language of his environment his own, to what extent these means are the same as those which the child uses in learning perceptual and conceptual skills in general, and whether the nature of these "cognitive strategies" is also determinant for some structural properties of natural language. Is it possible to give a general characterization of the relationship between language structure and the learnability of a language? A few mathematical aspects of this question have already been treated in Volume I, Chapter 8, and we shall return to them in Volume III, Chapter 4. It should suffice here to point out that the nature of linguistic intuitions and their development in the child is one of the most fundamental facets of the problem of interpretation. It is quite obvious that both of these aspects are very largely of a psychological nature.

A third case in which the problem of interpretation is of great importance in the formulation of linguistic theory occurs when one has to deal with the *analysis of a given corpus*, without the benefit of access to linguistic intuitions. Not only does the linguist have to cope with this situation in the study of dead languages, but also in applied linguistics he will find it to be less the exception than the rule. Translation and style analysis, for example, are most often performed on the basis of a corpus and without further access to the person who produced the text. From a formal point of view, however, the problem of corpus analysis has been most difficult in the analysis of children's language. One cannot base the development and testing of a grammar for the language of a three-year-old principally on linguistic intuitions. The number of metalinguistic utterances which a small child makes is quite small, and it is possible only on a very limited scale to elicit linguistic judgments from him. The small child is not comparable to the adult as an informant. If the domain of linguistic theory is limited to linguistic intuitions, the study of children's languages becomes a nearly impossible task. The data which can be obtained consists primarily of spontaneous utterances from the child and of observations relative to the circumstances under which they are produced. With ingenious experiments some information may be added to this, but the problem of determining what it is in verbal behavior which corresponds to the theoretical concept of "sentence" still remains, and usually demands a study in itself. It would not be advisable, for example, to consider every recorded utterance as a sentence in the child's language. Different utterances will often be taken for separate occurrences of a single sentence on the ground of acoustical criteria. If agreement can be reached at all on what these criteria are, it is still possible that not every class of equivalent utterances thus obtained will correspond to sentences in the theory. The grammar can often be simplified by excluding certain utterances or classes of utterances such as, for example, imitations which clearly have not been understood, and utterances of a very infrequent sort. Statistical and other data processing procedures will sometimes need to be adopted for the interpretation of the theory in terms of observable verbal behavior. Should the status of "sentence" be accorded to every utterance which the child understands? Decidedly not, for, as everyone knows, the child can understand much more than he can produce. But where can one draw the line? A sentence not produced by the child can very well have the status of "sentence" in the grammar. This is in fact the interpretation problem *par excellence*. Further interpretation problems occur in applied linguistics. In style analysis one might like to make use of data on distributions of sentence length and word frequencies. In the analysis of children's languages likewise, such data are often useful as parameters of growth or verbal skill. These matters can be considered as linguistic applications of inference theory (cf. Volume I, Chapter 8), where the interpretation problem is of central importance.

1.3. A FEW DESCRIPTIVE DEFINITIONS

In the preceding, careful distinctions were made between theory and observation, and between theoretical and empirical concepts. Theoretical concepts are determined entirely by their formal relations within the theory; in this connection we have already repeated the formal definitions of "sentence", "language" and "grammar", which may be found in greater detail in Volume I, Chapter 1. To these we can add definitions of "morpheme" and of "syntactic category". Unless otherwise stated, MORPHEMES are the terminal elements of the grammar; together they form the terminal vocabulary. For the sake of simplicity, we shall often refer to the terminal elements as WORDS, except where this might lead to confusion. We would point out that in transformational grammar another term, FORMATIVES, is also used for the terminal elements.[1] The nonterminal vocabulary (cf. Volume I, Chapter 1), by definition, is made up of SYNTACTIC CATEGORIES. The symbolic abbreviations for these are called CATEGORY SYMBOLS.

As a support to the intuitions which will often be called upon in the following chapters, we present a few descriptive definitions of the correspondents in the empirical domain of the concepts "sentence", "morpheme", "word", and "syntactic category". These are not formal definitions, and are meant only to make the ideas a bit clearer. They pretend to no completeness, but refer to each other as do the theoretical concepts.

SENTENCE. This theoretical concept has, as mentioned above, something to do with the empirical concept of "utterance". A hearer might consider two utterances to be identical, in spite of acoustical differences among them. Whether it is John or Mary who says *the weather is nice*, in the intuition of the native speaker, the two acoustical forms are simply occurrences of the same sentence. The intuition "the same statement (question, exclamation, etc.)" thus determines classes of equivalent utterances. Let us call each of those classes a *linguistic construction*. The relationship between this abstract empirical concept "linguistic construction" and the theoretical concept "sentence" remains complicated. On the one hand, there are linguistic constructions which we might prefer to represent theoretically as combinations of sentences; a story, for example, is a linguistic construction which we would ordinarily prefer to analyze as a sequence of sentences. On the other hand, there are linguistic constructions which we would rather consider to be parts of sentences than complete sentences; thus, for

[1] Some authors (Katz and Postal et al., 1964) make a further distinction between morphemes and formatives. Others (Chomsky, 1965) treat only the concept "formative".

example, the answer to the question *where is the hat?* is *on the table. On the table* is a linguistic construction (which might sound very different when spoken by different speakers or at different times), but we do not consider it a sentence. The principal reason for this is that *on the table* is dependent on the question which precedes it. Thus *on the table* cannot follow the question *what is your age?*. The principle used here is that of *distributional independence* (for a definition and discussion of the principle, see Lyons, 1968). Within a linguistic construction to which we should like to accord the status of sentence, various distributional dependences exist. Thus *the lazy nurse stood up* is good English, but *the lazy stone stood up* is not (it could be good English at best in a metaphorical context, but we shall not discuss such cases here). There are limitations of the nouns which can follow the adjective *lazy*. Names of inanimate objects are excluded in this connection (abstraction made of idioms such as "lazy day", etc.). This is a distributional limitation. We let the concept "sentence" correspond to linguistic constructions *within* which distributional limitations hold, but *among* which no distributional limitations hold. Consequently *on the table* is not a sentence, because it is dependent on the earlier question. However this does not yet solve the problem, as we see in the following question and answer situation: *where is your aunt?*, *she is coming*. The distributional dependence between these sentences is expressed in the intuition that at first sight *he is coming* is an unacceptable sequence to *where is your aunt?*. Yet we would like to represent *she is coming* in the grammar as a sentence. To allow for this, we can make an exception for pronouns in the rule of distributional independence. In other words, we represent these linguistic constructions as sentences in the grammar, adding that the pronoun stands for the noun mentioned in the other sentence. But this too falls short of solving all the problems. Additional criteria can always be given to provide the theoretical concept "sentence" with as careful an empirical basis as possible. Criteria concerning the intonation of the utterance would be an example of this. For the ends of the present volume, however, no further differentiation is needed.

MORPHEME (*formative*). Just as the theoretical concept "sentence" corresponds to the empirical concept "utterance", the theoretical concept "morpheme" corresponds to the empirical concept "morph". Roughly defined, morphs are the smallest meaning-carrying elements of a linguistic construction. Thus *the boys walked* can be segmented as *the-boy-s-walk-ed*; each segment is a morph with a functional or referential meaning. Some morphs can occur "independently" in a linguistic construction. This is the case here for *boy* and *walk*. Others occur only in combination, such as *s* (for the plural) and *ed* (for the past tense) in the present example. The status of *the* is less clear in this connection. Nevertheless we do not wish to limit the terminal elements of a grammar to such observable elements. The linguistic construction *the children ran*, segmented as *the-child-ren-ran*, makes the reason for this quite clear. The meaning of plural is carried by the morph *ren*, and we might thus consider the morph *s* of the preceding example and the morph *ren* as variants of the same grammatical element. The corresponding morpheme in the grammar can be written abstractly as *plural*, or simply as *pl*. The question becomes more abstract, however, when we compare *walk-ed* and *ran*. Our intuition tells us that *walk* is related to *walked* in the same way as *run* is related to *ran*, but in the latter case there is no separate morph which expresses the past tense. Change of tense occurs in the form of a vowel shift in the root. By analogy with *walked*, we can represent *ran* as *run* + *past tense*, or simply as *run* + *past t*, where *past t* represents the past tense morpheme. The consequence of this, however, is that without further additions such words as *ran*, *walked*, *boys*, *children* can no longer be generated by the grammar. The terminal strings will contain such pairs as *run* + *past t*, *boy* + *pl*, etc. Therefore rules must be added to the grammar to change these strings to the correct froms (*ran*, *boys*, etc.). The part of grammar called *morphology* is concerned with such rules. Morphological rules will not be explicitly discussed here, and we shall suppose that a morphology exists for changing the terminal strings of morphemes into the proper word forms, and will represent the terminal elements directly as words, unless this in a given case might lead to confusion. Morphemes such as *run, boy, walk* are called *lexical formatives,* and *pl* and *past t* are called *grammatical formatives.*

WORD. This concept will be used only to mean "terminal element", as mentioned above. Theoretically this concept would more properly belong to morphology, which, as we have stated, will be left largely untouched. A rather good definition of the concept "word" is 'a minimal free form'. There are various ways of interchanging morphemes in a sentence and of adding new morphemes without changing the character of the sentence. If, for example, *the boys are walking* is a sentence, then *are the boys walking* is also a sentence, and *the big boys are walking* likewise. In shifts of this kind, some morphs always remain coupled, like *walk* and *ing*, and *boy* and *s*. Such internally connected groups are called words. The smallest free forms in the example are *the, boys, are, walking*, and the form *big* which was added later. This definition is certainly not exhaustive, but should be sufficient to serve as a memory aid for the rest of the book.

SYNTACTIC CATEGORY. This concept is the most difficult to define. Two things may be borne in mind in connection with it. In the first place, one can relate the concept to that which is ordinarily called a "phrase", such as "a noun phrase" (e.g. *the big boy, John's carpentry, old folks*) or "a verb phrase" (e.g. *goes to school, does not give himself away, is a bit lazy, plays the piano*). In the second place, the concept relates to *classes of morphemes,* such as "number" for the class of morphemes consisting of "singular" and "plural", "tense" for the class of morphemes of time (*past t, pres t,* etc.), "verb" for the class of morphemes like *walk, run, sing*, etc. Formal models of natural languages tend to show considerable divergence in the choice of syntactic categories. Therefore in the following we shall give supplementary definitions when needed.

2

PURE MODELS: PHRASE-STRUCTURE GRAMMARS

In this chapter formal grammars of the pure types 3, 2, and 1 will be examined on their value as models for linguistic grammars. When these grammars are used in linguistics, they are denoted by the generic term PHRASE STRUCTURE GRAMMARS, or CONSTITUENT STRUCTURE GRAMMARS. These designations are related to the fact, discussed in Volume I, Chapter 2, that derivations in such grammars can be represented by means of tree-diagrams. The reader may remember that this held for type-1 grammars only when their production rules were of context-sensitive form (cf. Volume I, paragraph 2.4.1); in the following we shall continue to respect that condition. A tree-diagram clearly shows the phrases of which a sentence is composed. Phrases may also be called CONSTITUENTS, whence the second term for this family of grammars. In linguistics, tree-diagrams for sentences are often called PHRASE MARKERS or simply P-MARKERS.

2.1. GENERATIVE POWER AND STRUCTURAL DESCRIPTION

The order of subjects to be discussed in this chapter will be determined by the methodological principles which have consistently served as the basis of the investigation of formal models in linguistics. Thus the strongest possible model is chosen first to see if that model can be maintained for the description of natural languages. Only if the model can convincingly be rejected can one go a step higher in the hierarchy and repeat the procedure. In this way one can be sure that the grammar used will never be too broad for the language (cf. Volume I, Section 2.1). Some clarification of what is meant by a "model which can be maintained" or a "tenable model" will be useful. It is only in the more limited sense of "observational adequacy" that we can see precisely what is required, namely that a grammar generate all and only the sentences of a language. One can speak here of the WEAK GENERATIVE POWER of the grammar; this is the language which is generated by the grammar. The weak generative power of a class of grammars (for example, that of the class of regular grammars) is the set of languages generated by the grammars in that class. Thus the weak generative power of the class $\{G_1, G_2, \ldots\}$ is the set $\{L_1, L_2, \ldots\}$, where language L_i is generated by grammar G_i.

It is much less easy to decide whether or not a grammar is descriptively adequate, that is, whether or not it correctly reflects the intuitions of the native speaker. This requirement is often operationalized in the criterion of whether or not the grammar assigns the correct STRUCTURAL DESCRIPTION to the sentences generated. The structural description is the information about the sentences, given by the grammar. This information is contained completely in the GENERATION of the sentence (the LEFTMOST DERIVATION for context-free grammars; cf. Volume I, paragraph 2.3.4). It shows how the sentence is composed of terminal elements, the syntactic categories to which words and phrases in the sentence belong, which production rules were used in the derivation and in what order. On the basis of such structural data other intuitions can also be formalized, such as intuitions concerning the relations among various sentences. Structural descriptions for regular and context-free grammars are identical with the P-marker. Derivations in context-sensitive and type-0 grammars cannot unambiguously be shown in tree-diagrams, and consequently further definition of "structural description" will be necessary. For the present, however, we may decide that the structural description of a sentence will be denoted by the symbol Σ. The set of structural descriptions given or generated by a grammar G is called the ANALYZED LANGUAGE $A(G)$, generated by G. Thus

$A(G) = \{\Sigma_i \mid \Sigma_i$ generated by $G\}$. $A(G)$ is also called the STRONG GENERATIVE POWER of G. The strong generative power of a class of grammars $\{G_1, G_2, \ldots\}$ is $\{A_1, A_2, \ldots\}$, where $A_i = A(G_i)$.

It is possible that an observationally adequate grammar might assign structural descriptions to sentences, while those structural descriptions conflict with various intuitions. It does not seem likely, however, that we might ever be able to expect proof of the intenability of a grammar on grounds of descriptive inadequacy; such a grammar would rather be rejected on the basis of increasing inconvenience in working with it. We shall see later, moreover, that contrary to current opinions, observational inadequacy has never been strictly proven for any class of grammars whatsoever.

Having defined strong generative power in addition to weak generative power, we must give the same extension to the concept of EQUIVALENCE of grammars. In Volume I, paragraph 1.2, we stated that two grammars G_1 and G_2 are weakly equivalent if $L(G_1) = L(G_2)$. We add to this that two grammars G_1 and G_2 are STRONGLY EQUIVALENT if $A(G_1) = A(G_2)$. If G_1 and G_2 are strongly equivalent context-free grammars, they assign the same (set of) tree-diagrams to sentences. ("Set" is added between parentheses to cover cases where sentences are ambiguous and may therefore have several different tree-diagrams; cf. Volume I, Figures 2.4 and 2.5). The inverse, however, does not hold in all cases: for context-sensitive grammars the same tree-diagram can be obtained for two different derivations (cf. Volume I, paragraph 2.4.1). The concept of "strong equivalence" is of linguistic interest because of the problem of the descriptive adequacy of grammars. Thus if G_1 and G_2 are strongly equivalent and G_1 is descriptively adequate, then G_2 is also descriptively adequate. Yet the concept presented in its usual form is rather trivial; two strongly equivalent grammars are identical, with the possible exception of a few uninteresting details. They may only differ in unusable production rules, i.e. rules which, if used, do not lead to terminal strings, or in vocabulary elements which cannot be used. Linguistics is decidedly in need of formalization of "equivalence of structural description", but the strength of that concept should be attuned to the tolerance of our intuitions toward syntactic structures. The only effort in this direction known to us is that of Kuroda (1972).

2.2. REGULAR GRAMMARS FOR NATURAL LANGUAGES

How can a regular grammar be imagined as a model for a linguistic grammar? This can best be illustrated by an example.

EXAMPLE 2.1. Let $G = (V_N, V_T, P, S)$ be a regular grammar, where $V_N = \{S, A, B\}$, $V_T = \{the, bites, dog, cat, scratches, black\}$, and P contains the following productions:

1. $S \rightarrow$ *the* A	5. $B \rightarrow$ *bites*
2. $A \rightarrow$ *black* A	6. $B \rightarrow$ *scratches*
3. $A \rightarrow$ *cat* B	7. $B \rightarrow$ *bites* S
4. $A \rightarrow$ *dog* B	8. $B \rightarrow$ *scratches* S

This grammar can generate such sentences as *the dog bites*, *the black cat bites*, *the black cat scratches*. The derivation of the last sentence, for example, is $S \overset{1}{\Rightarrow}$ *the* $A \overset{2}{\Rightarrow}$ *the black* $A \overset{3}{\Rightarrow}$ *the black cat* $B \overset{6}{\Rightarrow}$ *the black cat scratches* (the numbers written above the arrows refer to the productions used in the derivation step). The corresponding tree-diagram is given in Figure 2.1.

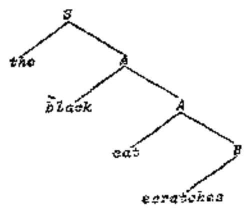


Fig. 2.1. *P*-marker for the sentence *the black cat scratches* (Example 2.1).

The grammar in fact generates an infinity of sentences. By virtue of production 2 which is recursive, the adjective *black* can be indefintely repeated, as in *the black black black dog bites*. The grammar can also generate compound sentences thanks to productions 7 and 8 which reintroduce the start symbol S. This produces such sentences as *the dog bites the black cat scratches*, etc.

The equivalence of regular grammars and finite automata shown in Volume I, Chapter 4 suggests that a finite automaton (*FA*) can be constructed which will be equivalent to this grammar. The following *FA* is equivalent to G:

$FA = (S, I, \delta, s_0, F)$, with $S' = \{S, A, B\}$, $I = \{$*the, bites, dog, cat, scratches, black*$\}$, $s_0 = S$, $F = \{S\}$, and the following transition rules:

$\delta(S, the) = A$ | $\delta(A, cat) = B$
$\delta(A, black) = A$ | $\delta(B, bites) = S$
$\delta(A, dog) = B$ | $\delta(B, scratches) = S$
$\delta(\text{-},\text{-}) = \varphi$ for all other cases

The transition diagram for this automaton is given in Figure 2.2. The diagram clearly shows which sentences the automaton accepts,

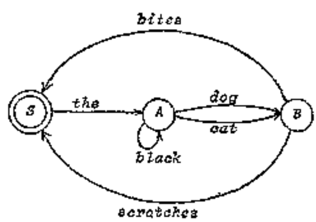


Fig. 2.2. Transition diagram for the finite automaton in Example 2.1.

and consequently the sentences which the grammar generates. Every path shown by the arrows from the initial state S to the final state S corresponds to a grammatical sentence. To complete the presentation, we give the transition table for this automaton in Table 2.1. The attentive reader will have noticed that this example is formally identical with Example 4.1 of Volume I.

It should be evident that many variations are possible here. It might be so arranged that the terminal vocabulary is made up of

TABLE 2.1. Transition Table for the Finite Automaton in Example 2.1.

		Input Symbols					
		the	*black*	*cat*	*dog*	*bites*	*scratches*
States	*S*	*A*					
	A		*A*	*B*	*B*		
	B					*S*	*S*

morphemes instead of words; it could even be made up of letters or phonemes. The grammar can be rendered more abstract by composing its terminal vocabulary of *classes* of words and morphemes. Thus, regularities may be formulated as, for example, "an article can be followed by a noun or by an adjective", "a noun can be followed by a verb", etc. In this way SENTENCE SCHEMAS are generated, such as "article-adjective-noun-verb". The grammar must then be completed with "lexical rules" showing which words are articles, which are adjectives, and so forth.

One may search in vain in linguistic literature for an explicit proposal to model linguistic theory on regular grammars, in spite of appearances to the contrary.[1] Some confusion exists on this point, since linguists have not seldom used the terms "finite grammar" and "finite state grammar" interchangeably, and readers may be led to think they are referring to a regular grammar when they only mean a finite system of rules. The most explicit use of the model may be found in the application of communication theory to natural languages. The origin of a verbal message is described as a so-called *Markov-source*, which in essence is a probabilistic finite automaton. We shall return to this in Chapter 6, paragraph 1. Suffice it here to point out that this has never been presented as a model of linguistic theory in the strict sense, but only as a model for summing up global statistical properties of a text (oral or written). It has never touched the structure of such messages in detail. Linguistics has indeed gone through a period of 'flirtation' with the

[1] Diligent searching revealed one exception to this, Reich (1969). The large number of essential errors in this article, however, gives rise to some doubt as to the carefulness of the editors of *Language*.

model under the influence of communication theory, but this never went beyond the implicit. Since Chomsky's explicit linguistic formulation and his rejection of the model in 1956, no linguist has seriously proposed it as a model for linguistic grammars.

Chomsky (1956, 1957) rejects the model on grounds of observational inadequacy. The enormous influence which this argument has had on the development of modern linguistics justifies a rather detailed discussion of it. It is also the case that the argumentation as given in *Syntactic Structures* is not completely balanced (the same is true, to a lesser degree, of Chomsky's treatment of the question in 1956). A consequence of this has been that the same sort of evidence is incorrectly used for the rejection of other types of grammars, and, as we shall see, simply erroneous conclusions have been drawn.

Before dealing with the form of the reasoning, we must first consider the fact that every argument is based on the supposition that a natural language contains an infinite number of sentences. Every finite set of sentences can, in effect, be generated by a regular grammar (cf. Theorem 2.3 of Volume I). What is the linguistic justification for this supposition? Three reasons are ordinarily advanced. The first of these has already been mentioned in Chapter 1, paragraph 2 of the present volume, namely, that from a linguistic point of view it is not advisable to set an upper limit to sentence length. The inacceptability of very long sentences can be justified better on the basis of a psychological theory than on the basis of a linguistic theory. A language is infinite if for every sentence another sentence can be found which is longer than the first, and this is clearly an intuitive fact. The second reason, closely related to the first, is the possibility of coordination of sentences. If sentences s_1 and s_2 are declarative (for example, *John is walking* and *it is raining*), then s_1 *and* s_2 also form a declarative sentence (*John is walking and it is raining*). Thus for every pair of declarative sentences, a new declarative sentence can be found which is longer than either. If a language contains one declarative sentence, it contains an infinity of them. The third and principal reason, however, is the following. Imagine a natural language of finite size. According to Theorem 2.3 of Volume I, a regular grammar can therefore be written for it. But this grammar will have no recursive production rules (i.e., production rules which make it possible to use a given nonterminal element repeatedly for an indefinite number of times in a derivation, like productions 2, 7, and 8 in Example 2.1). Excluding trivial cases, such rules lead to the generation of infinite languages. But if recursive production rules are excluded from the grammar, the number of production rules will be of the same order of magnitude as the number of sentences in the language. Such a grammar would scarcely be helpful in clarifying the structure of the language; a list of all the sentences would be quite as good. The assumption of infinitude is, in other words, a fundamental decision designed for finding a characterization of the language which is as general and as simple as possible.

The argument of inadequacy advanced in *Syntactic Structures* is of the following form: (a) a language with property x cannot be generated by a regular grammar, (b) natural language L has property x, therefore (c) L is not a regular language. The argumentation here is balanced, but the difficulty lies in demonstrating (b). Let us examine this more closely on the basis of the argument.

For property x we shall take self-embedding. From Theorem 2.8 in Volume I we know that self-embedding languages are not regular. This is step (a) in the argument. We must now show for (b) that English is a self-embedding language. This is done by referring to self-embedding subsets (called *subparts* in *Syntactic Structures*) in English. Thus, for example, if s_1 is grammatical, one can add a relative clause to it without loss of grammaticality, as in s_2:

s_1: *the rat ate the malt*
s_2: *the rat the cat killed, ate the malt*

One can now add a relative clause to the relative clause in s_2, as in s_3:

s_3: *the rat the cat the dog chased, killed, ate the malt*

The embedded structure of s_3 becomes obvious when we add parentheses:

(the rat (the cat (the dog chased) killed) ate the malt)

There is no fundamental limit to the number of possible self-embeddings of this kind. The sentences become complicated, but always remain completely unambiguous in meaning. When necessary, one can verify or falsify such a statement, as the following, completely unnatural sentence:

(William II (whom William III (whom William IV (whom William V succeeded) succeeded) succeeded) succeeded William I)

Another example of self-embedding in English is the following sequence:

s_1: *If John says it is raining, he is lying*
s_2: *If John says Joe says it is raining, he is lying*
s_3: *If John says Joe says Mike says it is raining, he is lying*
and so forth.

It would not be difficult to think of other examples. The conclusion is that on the basis of the self-embedding character of English (c) follows, i.e. English is not a regular language.

The self-embedding property (b) of English is however not yet demonstrated, in spite of appearances to the contrary. The only thing which has been proven is that English has self-embedding subsets. But it by no means follows from this that English is a self-embedding language. This can easily be seen in the following. Let language L consist of all strings over the vocabulary $V = \{a, b\}$, so that $L = V^+$. Language L is regular, because it is generated by a regular grammar with production rules $S \to aS$, $S \to bS$, $S \to a$, $S \to b$. Let us now consider the set $X = \{ww^R\}$, the set of symmetrical "mirror-image" sentences, where $w \in V^+$. It is clear that X is a subset of L. Moreover, X cannot be generated by any regular grammar, given its self-embedding property. Nevertheless L is a regular language. The reason why the argument errs is that sentences which are excluded by a grammar for X are nevertheless sentences of L. The omission in the argument for inadequacy is that nothing is said of the grammatical status of relative sentences (or sentences of other types discussed) which are not self-embedding.

Chomsky's original argumentation (1956), to which he refers in *Syntactic Structures*, is considerably more precise. In it he shows that it is necessary for the proof to demonstrate that a certain change in the sentences of a self-embedding subset must always be accompanied by a certain other change, on pain of ungrammaticality. But in the demonstration of that theorem he does not test whether or not this is in fact the case for English. Chomsky chooses the following intuition concerning English: if s_1 and s_2 are English sentences, then *if* s_1 *then* s_2 is also an English sentence. Repeated embedding shows *if* (*if* s_1 *then* s_2) *then* s_2 also to be an English sentence, and in general, $if^n\ s\ (then\ s_2)^n$, $n \geqslant 1$. Let us suppose that this holds for English (although this is itself an open question); it must then also be shown that $if^n\ s_1\ (then\ s_2)^m$ is ungrammatical if $n \neq m$.[1] Chomsky, however, does not do this, and, moreover, it does not hold. Grammatical counter-examples are *if John sleeps he snores* and *John drank coffee, then he left*. Similar objections may be made to the other examples in Chomsky (1956) and (1957).

Fewer problems occur when the "proof" is stated as follows (this is due to Dr. H. Brandt Corstius, personal communication). We follow a procedure of indirect demonstration. Assume that English is regular. We now construct the following regular set T. $T = \{William\ (whom\ William)^n\ succeeded^m\ succeeded\ William \mid n, m \geqslant 1\}$.[2] It has been proven by Bar-Hillel (see Hopcroft & Ullman, 1969) that the intersection of two regular sets is a regular set. Therefore, the intersection of English and T should be regular.

[1] It must at least be shown that $n \geqslant m$ for all sentences, or $n \leqslant m$ for all sentences, because not only is $\{a^n b^n\}$ non-regular, but $\{a^n b^m \mid n \leqslant m\}$ and $\{a^n b^m \mid n \geqslant m\}$ are likewise non-regular.

[2] A right linear grammar for T is: $S \rightarrow$ *William A*, $A \rightarrow$ *whom William B*, $B \rightarrow$ *succeeded C*, $B \rightarrow$ *whom William B*, $C \rightarrow$ *succeeded C*, $C \rightarrow$ *succeeded William*. A language with a right linear grammar is regular (cf. Theorem 2.1 in Volume I).

Let us therefore have a closer look at English $\cap$ T. Intuitively, the only grammatical sentences in T are those for which $n = m$, though some people have the intuition that one may delete occurrences of *succeeded* so that the grammatical sentences in T are those for which $n \geqslant m$. In both cases ($n = m$, $n \geqslant m$), however, the intersection is self-embedding; there is no regular grammar which can generate sets like $\{a^n b^n\}$, or $\{a^n b^m \mid n \geqslant m\}$. The intersection is, therefore, not regular. This contradicts the fact that the intersection should be regular, and hence our only assumption must be wrong, namely that English is a regular language.

Although this form of proof avoids the formal difficulties, the "proof" remains as weak as the empirical observation on which it is based. However, it is upon reaching this level of empirical evidence that one can decide in theoretical linguistics to formulate the state of affairs as an axiom: *natural languages are non-regular*. Given the independent character of a theory (see the preceding chapter), this is a more correct method of work than simply acting as though one were dealing with a *theorem* which could be proven, as linguists often do. The latter method is an incorrect mixture of theory and observation.

2.3. CONTEXT-FREE GRAMMARS FOR NATURAL LANGUAGES

Example 2.2 gives a context-free grammar for (part of) a natural language.

EXAMPLE 2.2. Let $G = (V_N, V_T, P, S)$ be a context-free grammar with $V_N = \{S, NP, VP, D, N, V, A\}$, $V_T = \{$*nice, the, and, congratulate, big, boys, children, little, malicious, girls, tease, defend*$\}$, and the following productions in P:

1. $S \rightarrow NP + VP$	6. $N \rightarrow \{$*boys, girls, children*$\}$
2. $NP \rightarrow NP + and + NP$	7. $V \rightarrow \{$*defend, tease, congratulate*$\}$
3. $NP \rightarrow (D) + (A) + NP$	8. $A \rightarrow \{$*malicious, nice, big, little*$\}$
4. $VP \rightarrow V + NP$	
5. $NP \rightarrow N$	9. $D \rightarrow$ *the*

Explanation of the notation: The category symbols stand for usual linguistic quantities, *S* for "sentence", *NP* for "noun phrase", *VP* for "verb phrase", *V* for "verb", *N* for "noun", *D* for "determiner", and *A* for "adjective". The sign + only indicates the concatenation of elements. It is used to avoid typographical indistinctness which could come about when elements are printed directly next to each other. Productions with elements surrounded by parentheses are in fact sets of productions; elements placed between parentheses may be used optionally in derivations. Thus production 3 stands for four productions: $NP \rightarrow D + A + NP$, $NP \rightarrow A + NP$, $NP \rightarrow D + NP$, $NP \rightarrow NP$. Braces indicate that only one of the several elements they surround may be used in a rewrite. Thus, in applying production 6 one may replace the *N* with either *boys* or with *girls*, or with *children*. The rule thus stands for three productions.

Sentences which may be generated by *G* are *boys defend girls*, *the little girls congratulate the big children*, *malicious big children tease nice little girls*, and so forth.[1] A leftmost derivation of *malicious boys and girls tease little children* is as follows (the numbers above the arrows indicate the productions applied):

$S \overset{1}{\Rightarrow} NP + VP \overset{2}{\Rightarrow} NP + and + NP + VP \overset{3,8}{\Longrightarrow} malicious + NP + and + NP + VP \overset{5,6}{\Longrightarrow} malicious + boys + and + NP + VP \overset{5,6}{\Longrightarrow} malicious + boys + and + girls + VP \overset{4}{\Rightarrow} malicious + boys + and + girls + V + NP \overset{7}{\Rightarrow} malicious + boys + and + girls + tease + NP \overset{3,9}{\Longrightarrow} malicious + boys + and + girls + tease + the + A + NP \overset{8,3,5}{\Longrightarrow} malicious + boys + and + girls + tease + the + little + children$. The *P*-marker for this sentence is given in Figure 2.3.

At this point we can proceed to the discussion of a few attractive qualities of context-free grammars for linguistics, problems of weak generative power, and problems of strong generative power. Much of what will be said here will hold also for context-sensitive grammars.

[1] The example has no pretensions; the grammar can also generate "sentences" like *nice the big boys congratulate girls*.

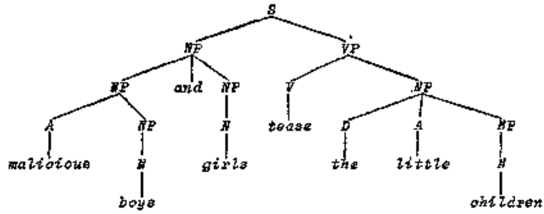


Fig. 2.3. Phrase marker for the sentence *malicious boys and girls tease the little children* (Example 2.2).

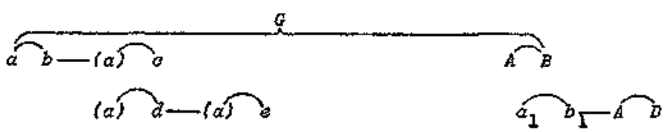
2.3.1. *Linguistically Attractive Qualities of Context-free Grammars*

Referring to the discussion in paragraph 2.2, we would first point out that context-free grammars have no difficulties with self-embedding. If a natural language is not regular, it is self-embedding, according to Theorem 2.8 of Volume I. A linguistic rendering of self-embeddingness calls at least for a context-free grammar. *NP* in Example 2.2 is a self-embedding element, as may be seen in Figure 2.3.

But context-free grammars were used as linguistic models in more or less implicit form, long before linguists became aware of the self-embedding property. An important reason for this was the undeniable need of *sentence parsing* in linguistics. Linguists have always been analyzing sentences into phrases. Sentence parts were labelled according to type, and their hierarchical articulation determined the levels of linguistic description. Thus at the lowest level minimal syntactic elements are distinguished which were called morphemes or otherwise. On a somewhat higher level one finds words. An element of higher level is composed of elements of a lower level: a word is composed of morphemes, a phrase is composed of words. A still higher level is found in the traditional distinction between subject and predicate, and so on. Context-free (and context-sensitive) grammars are very well suited to parsing in the form of *levels of labelled syntactic elements*, and we find these ideas in the most diverse linguistic traditions. For a survey of such models in modern English linguistics, we refer the reader to Postal (1964a); the article, although a bit one-sided, shows the "phrase structure" character of Hockett's linguistics, Lamb's *stratificational syntax*, Pike's *tagmemics*, and a few other theories, including that of the English linguist Halliday. But hierarchical parsing of sentences is a much older tradition, especially in Europe. Take, for example, Jespersen's "analytic syntax", in which parts of sentences are labelled according to function (subject, object, indirect object, etc.), or the important work of Wundt (1900) which is especially interesting for psycholinguistics. We can give an example from the latter work, Wundt's analysis of the following sentence from Goethe's *Wahlverwandtschaften*.

Als er sich aber den Vorwurf sehr zu Herzen zu nehmen schien ($a \frown b$) und immer aufs neue beteuerte (c), dasz er gewisz gern mitteile (d), gern für Freunde tätig sei (e), so empfand sie ($A \frown B$), dasz sie sein zartes Gemüt verletzt habe ($a_1 \frown b_1$), und sie fühlte sich als seine Schuldnerin ($A \frown D$).

Wundt gives the following phrase marker for this:



The *G* stands for *Gesamtvorstellung* or "general image", the psychological equivalent of "sentence". The brace and curves combine lower level elements "apperceptively" into higher level elements. "Apperceptively" means that there is a part-whole relationship between the lower level element and the higher level element. Straight lines indicate that the relationship is "associative", that is, there is no intrinsic relationship of part-to-whole, but only an accidental connection of elements. Notice also that Wundt sometimes puts elements between parentheses. Such elements repeatedly play a grammatical role in the sentence, but are not repeatedly pronounced. We shall return to this phenomenon of deletion, which got a first formalization in Wundt's diagrams.

In this tradition of parsing, the linguistic method of distributional analysis could thrive. Particular attention was paid to finding a distributional definition of syntactic elements which can play a certain part in sentence structure. This in turn led to distinguishing elementary sentence schemas. The hierarchical relations of inclusion among the labelled syntactic elements in Figure 2.3 give a very satisfying representation of our intuitions concerning the sentence they compose. Finally, we point out that such relations of inclusion make it possible to give justification for some structural ambiguities. The sentence given in Figure 2.3, *malicious boys and girls tease the little children,* is an example of an ambiguous sentence. It is an intuitive datum that *malicious* can refer to *boys and girls* (1), or only to *boys* (2). Even before the formalization of context-free grammars linguists of the "immediate constituent analysis" tradition knew that such ambiguities could be justified by way of inclusion relations. From a formal point of view a sentence is ambiguous, relative to a context-free grammar, when two leftmost derivations of it are possible in that grammar (cf. Volume I, section 2.3.4), and it consequently has two tree-diagrams. In Figure 2.3 we read meaning (1), for *malicious boys* is one noun phrase, and *girls* is another. It is easy to see that the grammar in Example 2.1 also generates the other structure, as given in Figure 2.4. It is quite clear that the correct treatment of structural ambiguities is one of the most important touchstones for the descriptive adequacy of a grammar.

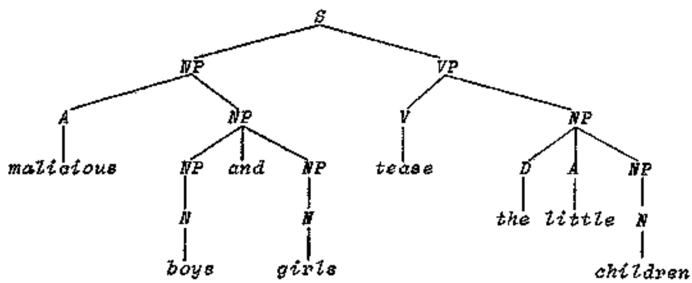


Fig. 2.4. Phrase marker for the sentence *malicious boys and girls tease the little children* (Example 2.2, alternative analysis).

2.3.2. *Weak Generative Power of Context-free Grammars*

Can context-free grammars be observationally adequate, that is, can they generate all and only the sentences of a natural language? Despite contentions of a different tenor in linguistic literature (Postal (1964b)), this question still remains unanswered. Postal "proves" the theorem (his term) that the North American Indian language Mohawk is not context-free by following the argumentation schema of *Syntactic Structures*: (a) a language with the property of "string repetition", as in the language $\{ww\}$ in which w is a string of elements from the terminal vocabulary such that every sentence consists of a string and its repetition, is not context-free. (b) Mohawk has the property of string repetition: there are sentences of the form $s = a_1a_2...a_nb_1b_2...b_n$, where a_1 "corresponds" to b_1, a_2 to b_2, and in general a_i corresponds to b_i. Therefore (c) Mohawk is not a context-free language.

This reasoning is as defective as the one, which we criticized, on the proposition that natural languages are not regular. It is erroneous to conclude the non-context-freeness of a language from the existence of non-context-free subsets.

To our knowledge, the literature does not yet contain a correct demonstration of the observational inadequacy of context-free grammars. However, Brandt Corstius (personal communication) recently proposed a proof along the following lines.

The proof is by indirect demonstration. Assume English is context-free. Consider the following regular set: $T = \{$*The academics, accountants, actors, admirals, ... in respectively Belgium, Bulgaria, Burundi, Brasil, ... are respectively calm, candid, canny, careless, ...*$\}$, or abbreviated: $T = \{$*The* a^k, *in respectively* b^m are respectively $c^n \mid k, m, n \geqslant 0\}$. It is not difficult to write a right-linear grammar for T. It has been proven by Bar-Hillel (see Hopcroft & Ullman, 1969) that the intersection of a regular language and a context-free language is context-free. Since we assumed English to be context-free it should be the case that $T \cap$ English is context-free. Let us therefore consider which sentences in T are grammatical English sentences. Intuitively, these are the strings for which $k = m = n \geqslant 1$. However, it is known (see Hopcroft & Ullman, 1969) that there is no possible context-free grammar for the language $\{a^n b^n c^n \mid n \geqslant 1\}$, i.e. the intersection of T and English is not context-free. This contradicts our earlier conclusion, namely that the intersection is in fact context-free. Hence, the only assumption that was made, i.e. that English is context-free must be wrong.

Again this "proof" is as strong as the intuitions about the grammatical subset of T. The *respectively*-construction is rather unnatural. One could probably use Postal's observations for proving non-context-freeness of Mohawk. But Postal is quite cryptic about the grammatical status of strings that do not exactly adhere to string repetition.

Much more convincing, at any rate, are other arguments against the context-free character of natural languages. But they will have to be advanced entirely in terms of strong generative power.

2.3.3. *The Descriptive Inadequacy of Context-free Grammars*

The impossibility for context-free grammars, and for phrase structure grammars in general, to describe a natural language in an intuitively satisfying way has been discussed in great detail in several places (see, for example, Chomsky (1957), and Postal(1964a) and their references). We shall give a short account of a few of the most important arguments here.

(1) A correct representation of the structure of a sentence often, if not always, calls for more than one phrase marker. The identification of structural description and (a single) phrase marker, as is the case for context-free grammars, leaves various intuitive syntactic insights undescribed. A few cases in which there is need of more than one phrase marker are *discontinuities*, *deletions*, and *phenomena of correspondence*.

Discontinuous constituents may be seen in sentences like *John put his coat on*. Intuitively, *put on* belongs together, just as in the nearly synonymous sentence, *John put on his coat*. A context-free grammar gives two different phrase markers, and in the case of the first sentence *put* and *on* fall into different phrases. The correct word order is thus described, but that is a question of observational adequacy rather than of the intuition that the words belong together. Therefore a *pair* of phrase markers is needed, one of which would group *put* and *on* together (in the same way for both sentences), while the other would give justification for the word order as it is met in fact (different in each sentence). This problem is felt more acutely when one is dealing with languages with freer word orders, such as Latin. An important generalization is lost if for every permutation of words in a sentence a new phrase marker must be made, although the meaning of the sentence does not change essentially because of the permutation.

In the case of deletions we have to do with words or phrases which do function in the sentence, but need not be repeated explicitly. As we have seen, Wundt put such elements between parentheses. This is just another way of showing that more than one phrase marker is involved in the description of the sentence in question, namely, the phrase marker which does contain the elements, and that which does not. The phenomenon of deletion occurs very frequently in coordinative constructions. If we wish adequately to describe the paraphrase relationship between the sentences *John came and auntie came as well* and *John came and auntie as well*, we will have to find some way of making the relationship between *auntie* and *came* explicit, and at the same time we will have to show that *came* does not appear a second time because of the influence of the use of *John came*. Phenomena of coordination will be mentioned separately under (2).

A third general case in which more than one phrase marker seems necessary for the description of a sentence occurs in the representation of correspondence. Compare the sentences *the painters mix the paint* and *the painter mixes the paint*; we see the correspondence here between the number (singular or plural) of the subject and that of the verb. Transgression of the rules of such relations of correspondence leads to ungrammaticality, as may be seen in **the painters mixes the paint* or **the painter mix the paint*.[1]

[1] It is customary to mark non-grammatical sentences with the sign*.

We are obviously dealing here with a very general property of English which should be expressed in the grammar. For this it will be necessary that *painters* be generated as *painter+pl*, *painter* as *painter+sg*, *mix* as *mix+pl*, and *mixes* as *mix+sg*. It must also be shown in some way that the morpheme *pl* must be added to *mix* only when the subject *(painter)* appears with *pl*, and the morpheme *sg* must be added to *mix* only when the subject appears with *sg*. In other words, the "underlying" form *mix* is changed to *mix+pl* or to *mix+sg* under certain conditions elsewhere in the sentence. But this means that *mix* and *mixes* must be described in two ways: on the one hand it must be shown that *mix* is *mix+pl* and that *mixes* is *mix+sg*, and on the other hand that *pl* or *sg* are not intrinsic to the verb, but rather dependent on a *pl* or *sg* earlier in the sentence.

(2) The description of coordination is a touchstone for every grammar, and therefore also for phrase-structure grammars (for a thorough study of this phenomenon, see Dik, 1968). In example 2.2 we find a context-free description of *NP*-coordination. By production 2 of the grammar, *NP* can be replaced by *NP+and+NP*. But what will happen when we want to coordinate several *NP*'s? We can apply the production repeatedly, but the hierarchical structure thus obtained would be rather uninforming. The noun phrase *boys and girls and children* will be set out either as *(boys and girls) and children* or as *boys and (girls and children)*. An ambiguity is thus introduced for which there is no intuitive basis: in this and other examples of coordination we prefer to see the elements as ordered really *coordinatively*. We might do this, for example, by making rules like $NP \rightarrow NP+and+NP+and+NP$, but then we would need a new production rule for every new string length. If there is no upper limit to the length of such coordinations, there will be an infinity of such productions. Another solution to the problem is the so-called *rule schema*: $NP \rightarrow NP^n+and+NP$, $n > 0$, by which strings like *boys, girls and children* of indefinite length can be generated. But whatever such rule schemas may be (there is noticeably little known of their mathematical structure relative to formal languages), they are not context-free production

rules. Thus context-free grammars give too much structure in the description of coordination phenomena.

But they also give too little. The phenomenon of deletion which often accompanies coordination is not satisfyingly accounted for by context-free grammars, as we have mentioned under (1). Especially for compound sentences like *Peter plays the guitar daily and John weekly*, a context-free grammar will either generate the deleted element, in which case no account is given for the deletion, or it will not, in which case no account is given for intuitively essential syntactic relations.

(3) Context-free and context-sensitive grammars treat ambiguities correctly only in some cases. Such a case was construed in the grammar of Example 2.2 which was capable of rendering the ambiguity *malicious boys and girls* correctly. There are, however, many cases in which phrase structure grammars fail concerning ambiguities. A few typical examples should make this point clearer. In *Italians like opera as much as Germans, Germans* either like or are liked; in *John watched the eating of shrimp, shrimp* either eat or are eaten; in *John is the one to help today*, John either helps or is helped. In all of these examples it is impossible to represent the ambiguities in an intuitively satisfying way by regrouping the syntactic elements, that is, by assigning alternative phrase markers to the sentences. In such cases a context-free grammar shows too little structure, as we have already seen in the case of deletions.

(4) A context-free grammar will often fall short of an intuitively satisfying representation of the relations *between* sentences. The passive sentence *the house was built by the contractor* is very directly related to the active sentence *the contractor built the house*. It is not clear how a context-free grammar might show that these sentences in important ways are paraphrases of each other. As soon as a similar structure is outlined for both sentences, as would be justified by intuition, account must be given for the fact that the sentences are very different in their elements and word order. To represent such relations, then, it will again be necessary to have a structural description which consists of more than one phrase

marker per sentence. Moreover we cannot write this off as an incidental case, given the generality of the active/passive relationship in English. Other general relations also yield such problems. An English yes/no question which contains an auxiliary verb stands in a simple relationship with the affirmative sentence; compare, for example, *has Peter been joking?* with *Peter has been joking.* In general this concerns a permutation of subject and auxiliary verb. But permutations yield discontinuous constituents, and the related problems for context-free grammars which we have already mentioned. It becomes much more difficult still to imagine a context-free grammar which correctly represents the relationship between the following sentences: *father gave mother roses* and *mother received roses from father.*

These and other kinds of inconveniences have slowly but surely led to the conviction that context-free grammars are descriptively inadequate, whatever their weak generative power may prove to be.

2.4. CONTEXT-SENSITIVE GRAMMARS FOR NATURAL LANGUAGES

For context-sensitive grammars the concept of "structural description" cannot be identified with the "phrase marker", as was the case for context-free grammars. In the first place, it is possible to construct phrase markers only when the grammar exclusively contains context-sensitive production rules (cf. Volume I, paragraph 2.4.1). In the second place, even in the latter case the phrase marker will not represent the derivation unambiguously. The contexts in which the various rewrites took place is especially not expressed. Likewise the sequence of strings obtained in the derivation of the sentence does not show what the contexts were in each step of rewriting. A context-sensitive structural description must, therefore, not only show the sequence of strings, but also the sequence of contexts. Context-sensitive phrase-structure grammars, that is, context-sensitive grammars with context-sensitive production rules, give structural descriptions which can best be defined as phrase markers, every nonterminal node of which is provided with the context in which it has been generated. This definition of structural description for context-sensitive phrase-structure grammars is used in Example 2.3.

Context-sensitive grammars can resolve some of the problems mentioned in the preceding paragraph, but quite as many new problems appear. Example 2.3 shows how, by the use of a context-sensitive grammar, one can treat the discontinuity which arises when an interrogatory sentence is generated. The example gives a very reduced grammar, developed especially for this problem, and without further pretensions.

EXAMPLE 2.3. Let $G = (V_N, V_T, P, S)$ be a context-sensitive grammar with $V_N = \{S, NP, NP', VP, N, V, V'\}$, $V_T = \{$*freedom*, *slavery*, *is*$\}$, and the following productions in P:

1. $S \rightarrow NP + VP$	5. $V \rightarrow V' \,/\, NP'$—
2. $NP \rightarrow N$	6. $NP' \rightarrow V$
3. $VP \rightarrow V + NP$	7. $V' \rightarrow NP$
4. $NP \rightarrow NP' \,/\, —V$	8. $N \rightarrow \{$*freedom*, *slavery*$\}$
	9. $V \rightarrow$ *is*

This grammar can easily generate the sentence *freedom is slavery*, but it can also generate the interrogatory sentence *is freedom slavery?* This latter is derived as follows: $S \overset{1}{\Rightarrow} NP+VP \overset{3}{\Rightarrow} NP+V+NP \overset{4}{\Rightarrow} NP'+V+NP \overset{5}{\Rightarrow} NP'+V'+NP \overset{6}{\Rightarrow} V+V'+NP \overset{7}{\Rightarrow} V+NP+NP \overset{2,2}{\Rightarrow} V+N+N \overset{8,9}{\Rightarrow}$ *is*+*freedom*+*slavery*. Figure 2.5

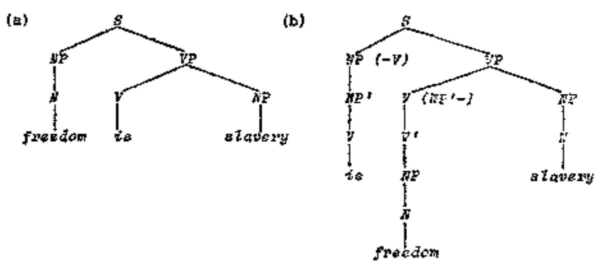


Fig. 2.5. Structural descriptions for the declarative (a) and the interrogative (b) sentences in Example 2.3.

shows the structural descriptions of the two sentences, that is, the phrase markers, to which the rewrite contexts have been added where necessary.

Permutations of elements can be realized with context-sensitive grammars, as may be seen in these illustrations. But it is also clear from the example that this is done in a highly unsatisfying way. The resulting phrase markers are very strange. Thus we see that the interrogative sentence (b) is composed of a *NP* and a *VP*, but that the *NP* is ultimately realized as *is*, and the *VP* as *freedom slavery*, thus in conflict with our dearest intuitions. Context-sensitive grammars supply the need for more than one phrase marker per sentence as badly as do context-free grammars.

Deletions, too, cannot in general be treated by context-sensitive grammars. It is possible, of course, to indicate the context in which a deletion occurs, but this necessarily leads to a type-0 production rule because the string is shortened. Correspondence, on the other hand, can be treated by context-sensitive production rules. There it is simply a matter of adding an element within a given context. Number correspondence, for example, in the sentence *the painter mixes the paint* could be dealt with by a production such as $Num \rightarrow sg/\ NP+sg+V$—, in which *Num* stands for the syntactic category *number*. If we are able to derive the string $NP+sg+V+Num$, application of this production will yield $NP+sg+V+sg$ in which the correspondence of number is realized (a similar argument holds for the plural). This is in fact the method used by Chomsky (1965) in dealing with the question of correspondence. At the suggestion of McCawley (1968b), Peters and Ritchie have proven (1969b) that such a use of context-sensitive production rules does not augment the weak generative power over that of context-free grammars. The advantage lies exclusively in the augmentation of descriptive adequacy.

Coordination and ambiguities yield the same problems for context-sensitive grammars as for context-free grammars (see the preceding paragraph). Some relationships among sentences, such as active/passive or declarative/interrogative, can to a certain extent be handled by context-sensitive grammars, namely, those concerning permutations and the addition of new elements. But just as in the example given in Figure 2.5, this leads to phrase markers which are in conflict with linguistic intuition.

To resume, nothing is known of the weak generative power of context-sensitive grammars in connection with natural languages, but the descriptive adequacy of context-sensitive grammars is hardly higher than that of context-free grammars. It seems justified to conclude that natural languages fall outside the class of context-sensitive languages, and that type-0 description is required.

2.5. RECURSIVE ENUMERABILITY AND DECIDABILITY OF NATURAL LANGUAGES

The step toward type-0 models for natural languages must not be taken lightly. The most important reason for caution is the decidability or recursiveness of natural languages. In Volume I, Chapter 7 we showed that the class of type-0 languages is equivalent to the class of sets accepted by Turing machines. Thanks to this equivalence, it was possible to show that type-0 languages are recursively enumerable sets (Theorem 7.3 in Volume I). A recursively enumerable language is a language for which a procedure exists to enumerate the sentences of that language, each in a finite number of steps. We have seen, however, that the complement of a type-0 language is not always recursively enumerable, and that consequently it is not generally true that type-0 languages are *decidable* (recursive). There is no algorithm by which a decision may be made, for every string, as to whether or not it belongs to the language. Such algorithms do exist for languages of types 1, 2, and 3.

With the introduction of type-0 grammars, therefore, we run the risk of generating undecidable languages. This, from a linguistic as well as from a psycholinguistic point of view, is a rather unattractive situation. We shall give three reasons for choosing a theory of natural languages in such a way that the languages generated are not only recursively enumerable, but also decidable.

(1) Native speakers will in general be as capable of judging that a sentence belongs to their language, as of judging that that is not the case. In other words, native speakers have an intuitive algorithm for the *recognition* of their language, and not only for *accepting* it. The formalization of this intuitive datum requires that the natural language be decidable in the model. One may object that there are also many unclear cases, for which, in this respect, there are no strong intuitions. But, as was said earlier, it is more elegant to ascribe this to psychological circumstances. The statement does not alter the intuitive fact that a judgment of ungrammaticality is just as direct as a judgment of grammaticality. If on the ground of this objection we drop the recursive enumerability of the complement of the language (the ungrammatical strings), on the ground of the same objection we must also drop the recursive enumerability, and therefore the type-0 character, of the language itself. It is also the case that intuitions of ungrammaticality are *strong*, i.e. the native speaker can often say what makes the string ungrammatical.

(2) A non-decidable language is unlearnable, even if the learner benefits from an informant. For the precise meaning of "learnability" and "informant" we refer the reader to the discussion in Volume I, Chapter 8, paragraph 3. In short this means that there is no algorithm by which an (observationally) adequate grammar can be derived from a sequence of strings marked "grammatical" and "ungrammatical". If there is no learnability in terms of an algorithm, there is certainly no learnability in terms of human cognitive capacities, given the finite character of the latter. The incontrovertible learnability of natural languages pleads that natural languages be considered as decidable sets.

(3) There remains the methodological principle, discussed in paragraph 2.1, that the strongest possible model must be chosen for a natural language. On the basis of this principle, the first step after the rejection of context-sensitive models is the decidable subset of type-0 languages. This is all the more urgent, since, as we have seen, the rejection of recursiveness in natural languages goes hand in hand with the rejection of recursive enumerability.

But to do so would mean to renounce the possibility of writing a generative grammar for the language, and therefore also the possibility of providing every sentence with an explicit parsing. This would come very near abandoning linguistics as a science.

Therefore the rules of the grammar should be chosen in such a way that the decidability of the language is maintained. This limits the choice considerably, and is not easily realized, as we shall see in Chapter 5. Furthermore, in setting phrase-structure grammars aside, we should take care not to "throw the baby away with the bath". The linguistic advantages of such grammars still hold (cf. 2.3.1), and ought, as far as possible, to be taken over into a more adequate theory of natural languages.

3

MIXED MODELS I: THE TRANSFORMATIONAL GRAMMAR IN *ASPECTS*

A transformational grammar is a pair $TG = (B, T)$, in which B is a *base grammar* and T is a set of *transformations*. In general B is a context-free grammar. Transformations are rules which have tree-diagrams as their input and output; when used in conjunction with the base grammar, they can raise the generative power to type-0 level.

Arguments of various kinds are advanced to support the use of this form of grammar in the description of natural languages. We shall mention a few of these arguments. By way of the B-component, the advantages of phrase structure grammars are simply taken up into a more complete linguistic theory. The transformational component T, on the other hand, makes it possible to assign more than one phrase marker to a sentence, and as we have seen in section 2.3.3, there is considerable need of such a possibility. Moreover, the type-0 character of the grammar is thus limited to the replacement of tree-diagrams with other tree-diagrams, allowing the recursiveness of the grammar to be kept under control. Semantic considerations also support the division of a grammar into two components. The semantic interpretation is determined entirely, or at least for the greater part, by the BASE STRUCTURE or DEEP STRUCTURE, that is, the phrase marker generated by B; the morphology of the sentence, on the other hand, can be described better in terms of the output of T, the SURFACE STRUCTURE. Still another argument is provided by the expectation, based on general language theory, that languages will tend to differ with respect to T, and to agree with respect to B, which would be considered the proper mechanism for the description of UNIVERSALS (the validity of this expectation is the subject of paragraph 3 in Chapter 5).

Transformational grammars differ quite considerably, however, in (i) the choice of base grammar, (ii) the choice of transformations, (iii) the distinction between *B* and *T*, i.e. the degree to which base and transformation rules may be applied "pell mell", (iv) the ratio between the size of *B* and that of *T*: few base rules may call for compensation in many transformation rules, and, within certain limits, vice-versa, and (v) the importance of *B* or *T* for semantic interpretation.

The diversity of transformational grammars, however, does not alter the fact that all of them are *mixed models*, that is, models in which a grammar of limited generative power (not more than type-1) is coupled with a limited set of rules for changing *P*-markers.

Most transformational grammars have evolved from Chomsky's formulation in *Aspects of the Theory of Syntax* (1965) (from this point we shall simply refer to the work as *Aspects*). In the present chapter we shall discuss the model presented in *Aspects*, first informally (3.1), then with a formal treatment of the structure of transformations (3.2). In the final paragraph of the chapter (3.3), we shall briefly discuss how certain considerations, principally semantic in nature, have led to changes in the original model. The changes proposed fall primarily into categories (iii), (iv) and (v) mentioned above. As the results of this are still very temporary, and as this book deals primarily with matters of syntax, our discussion of these points will not be very extensive. In Chapter 4 we shall treat a few alternative proposals concerning (i) and (ii). Those transformational grammars are in many respects very different from the *Aspects* model.

3.1. THE ASPECTS MODEL, AN INFORMAL DISCUSSION

In *Aspects*, a grammar consists of three components, a *syntactic* component, a *phonological* component, and a *semantic* component.

The syntactic component has the recursive qualities necessary for the generation of an infinite set of sentences. The phonological and semantic components describe respectively the aspects of sound and meaning of the structure generated by the syntactic component. Notice that the word "grammar" is used in *Aspects* in the wider sense (see Chapter 1, paragraph 1), including phonology and semantics. Grammar in the narrower sense, the subject of this book, correspond largely to that which is called the syntactic component in *Aspects*; there is complete correspondence when we do not consider morphology.

In this sense, the grammar in *Aspects* is a pair (B, T) of base grammar and transformations. We shall now discuss its principal properties in an informal way.

3.1.1. *The Base Grammar*

The productions of the base grammar are of two kinds: CATEGORIAL RULES and LEXICAL RULES. The categorial component is composed of context-free rewrite rules. They form a grammar with category symbols (*S*, *NP*, *Pred P*, *VP*, *V*, *N*, etc.) as the nonterminal vocabulary, and with grammatical formatives (*sg*, *pl*, *past t*, etc.), a so-called DUMMY SYMBOL Δ, and the boundary symbol # as the terminal vocabulary. For reasons which will become clear later, every derivation begins with $\#S\#$ instead of simply with S, but as long as there is no chance of confusion we shall omit the boundary symbols. The categorial rules, moreover, have the following two properties: (1) *Recursivity*: S (or actually $\#S\#$) appears in one or more productions to the right of the arrow, so that S can again be introduced into a derivation; there are no other recursive rules in B.[1] (2) The rules of the categorial component are applied in a certain *order*. This is done cyclically: when one arrives at the end of the list, one must start again at the beginning, and if there is an S which has not yet been rewritten, it is first to be

[1] This means that if, for a certain element X_n, the categorial component allows the following derivation $X_n \Rightarrow \omega_1 X_1 \psi_1 \Rightarrow \ldots \Rightarrow \omega_1 \ldots \omega_n X_n \psi_n \ldots \psi_1$, it holds necessarily that $X_i = S$ for some i, $i = 1, \ldots, n$.

dealt with. This ordering is inspired by the so-called "sequential grammars" of Ginsburg and Rice (1962). The restriction on the order of application is formulated in *Aspects*, but not used. Peters (1966) showed that sequential context-free grammars are weakly equivalent to unordered ("ordinary") context-free grammars. We shall ignore this restriction in the further discussion.

Example 3.1 gives part of a base grammar. Like the other examples in this chapter, it is meant only as an illustration. These examples are given to clarify certain properties, and not as serious proposals for a transformational grammar of English.

EXAMPLE 3.1. The base grammar contains nine productions, the nonterminal vocabulary consists of the following elements, *S*, *NP*, *VP*, *V*, *N*, *D*, *Num*, and the terminal vocabulary is made up of the following elements, *it*, *sg*, *pl*, Δ, and *Q* (a "question" morpheme). The productions are:

1. $S \rightarrow (Q)+NP+VP$	5. $NP \rightarrow \Delta$
2. $VP \rightarrow V+(NP)$	6. $Num \rightarrow \{sg, pl\}$
3. $NP \rightarrow (D)+N+Num$	7. $V \rightarrow \Delta$
4. $NP \rightarrow it+S$	8. $N \rightarrow \Delta$
	9. $D \rightarrow \Delta$

By these production rules the tree-diagram in Figure 3.1 can be generated. Between parentheses in the diagram morphemes are given which can replace the dummy symbols. The way in which this is done is determined by the lexical rules, which will be discussed later. Let us suppose for the moment that the replacement has

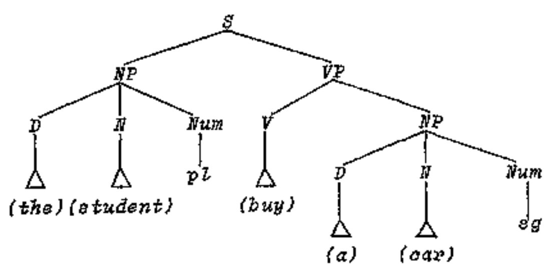


Fig. 3.1. Tree-diagram generated by the categorial rules in Example 3.1.

already taken place. Let us also suppose that the transformational component, applied to this diagram, successfully gives a terminal derivation (see paragraph 1.2 of this chapter). We can then call Figure 3.1 the DEEP STRUCTURE of the sentence *the students buy a car*. (It should be mentioned that no rules are given in this example for dealing with tense. When this enters the discussion in *Aspects*, Chomsky rewrites S to $NP + PredP$, where *PredP* stands for *predicate phrase*. This latter can in turn be rewritten as $Aux+VP$, and *Aux* as *pres t*, *past t*, etc., or it can be replaced by an auxiliary verb. The place of an indication of tense in the phrase marker, however, is still very arbitrary.)

In *Aspects* functional relations such as SUBJECT OF, PREDICATE OF, and DIRECT OBJECT OF are defined in terms of categorial properties of deep structures. For this definitions of DIRECT DOMINANCE and GRAMMATICL RELATION are necessary. Let $A \rightarrow \omega B \psi$ be a categorial rule in the base grammar (A and B are category symbols, and ω and ψ are possibly empty strings of terminal and/or nonterminal elements). Suppose that the base rules allow the derivations $\omega \stackrel{*}{\Rightarrow} \gamma$, $\psi \stackrel{*}{\Rightarrow} \delta$, and $B \stackrel{*}{\Rightarrow} \beta$, in which β is a non-empty string of terminal elements and γ and δ are possibly empty strings of terminal elements. In this case $A \stackrel{*}{\Rightarrow} \gamma\beta\delta = \alpha$ is a terminal derivation. It may be said then that (1) in this derivation A DIRECTLY DOMINATES $\omega\beta\psi$, because $\omega\beta\psi$ is derived from A in only one rewrite, and that (2) β has the GRAMMATICAL RELATION $[B, A]$ to α. In the example given in Figure 3.1, *student* has the grammatical relation $[N, NP]$ to *the student pl*, but *car* has no grammatical relation to *buy a car sg*, for there is no production $VP \rightarrow \omega N \psi$ in the grammar. Chomsky gives the following functional definitions. The grammatical relation $[NP, S]$ is "subject of". In the example, the noun phrase *the student pl (the students)* is the subject of the sentence *the student pl buy a car sg (the students buy a car)*. The relation $[VP, S]$ is "predicate of". Thus in the example, *buy a car* is the predicate of the sentence *the students buy a car*. The relation $[NP, VP]$ is "direct object of". Thus in the example, *a car* is the direct object of *buy a car*. Finally, the relation $[V, VP]$ is "main verb of". In the example, *buy* is the main verb of *buy a car*.

In paragraph 3.3. of the preceding chapter we met the ambiguities concerned precisely with such grammatical relations. *John watched the eating of shrimps,* for example, was ambiguous because *shrimps* could be taken either as the subject or as the direct object. It is possible on the basis of the just given definitions to express these two interpretations. The grammar in Example 3.1 can generate the phrase markers shown in Figure 3.2; they show two different deep structures for *John watched the eating of shrimps*. In Figure 3.2a, *shrimps* is the *subject* of *shrimps eat* according to the definitions, given the relation [*NP*, *S*] within the embedded clause. In Figure 3.2b, *shrimps* is the *direct object* of the embedded clause, given the relation [*NP*, *VP*]. Furthermore, quite in agreement with the intuition, the main clause has *John* as subject and *watched* as main verb, while the noun phrase which contains the subordinate clause is the direct object of the main clause. This representation is

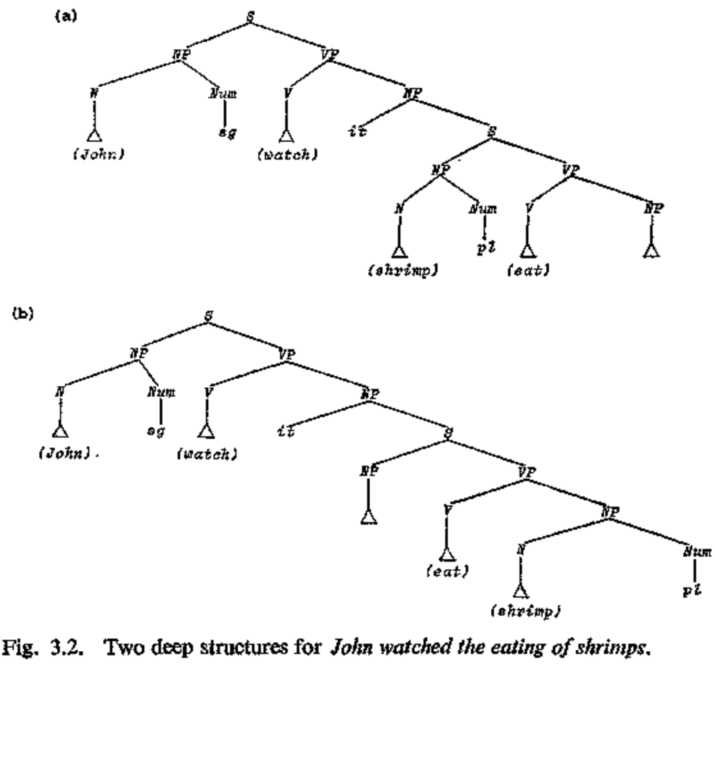


Fig. 3.2. Two deep structures for *John watched the eating of shrimps.*

satisfying up to this point, though it must, of course, be complemented by such transformations of the deep structures that ultimately the same terminal string, *John sg watch sg the eating of shrimp pl*, will be derived for both deep structures. But before going on to the discussion of transformations, we must still treat the lexical rules.

The *lexical rules* replace the dummy symbols with lexical formatives. A lexical formative consists of three "parts": (i) a phonological part, in which the sound properties of the formative are established; for the sake of simplicity we indicate this by *spelling* the morpheme: *shrimp, eat*, etc.; (ii) a syntactic part or set of syntactic features to which we return presently; (iii) a semantic part or set of semantic features, which will not be discussed here.

The conditions for replacing the dummy symbol Δ with a given lexical formative are couched in the syntactic features of that formative. When the tree-diagram satisfies these conditions the replacement may be performed.

A first condition for the replacement of a dummy symbol by a formative is that the formative be of the correct lexical category. Consider the sentence *the students buy a car* from Example 3.1, and notice the insertion of the formative *buy*. A condition for the insertion of a formative in that place is that it must be of the category V. Thus the dummy symbol in question cannot be replaced by a formative such as *magazine*. In order to exclude such strings as **the students magazine a car* while maintaining the possibility of a sentence like *the students buy a magazine*, the lexicon specifies that *magazine* has the category feature $[+N]$, and *buy* the category feature $[+V]$.

However not all the lexical formatives with the characteristic $[+V]$ may replace the dummy symbol. Thus in the example, the verb *laugh* is excluded, as we see in the ungrammatical string **the students laugh a car*. Obviously *buy* has a characteristic which *laugh* has not: *buy* is a transitive verb, while *laugh* is intransitive. Thus Chomsky distinguishes SUBCATEGORIES within a category; in this case the subcategories are those of transitive and intransitive verbs. Transitiveness and intransitiveness are syntactic features called (STRICT) SUBCATEGORIZATION FEATURES. Transitiveness can simply and efficiently be denoted as follows [+—*NP*]. This means that a formative with this feature can (+) appear in the place (—) immediately before a noun phrase (*NP*) in the deep structure. It is clear that the dummy symbol above *buy* in Figure 3.1 is in just such a place, and in this respect, therefore, may be replaced by *buy*.

But this still is not sufficient. Beside category and subcategory features, lexical formatives also need SELECTIONAL FEATURES. The verb *doubt*, just as *buy*, has the features [+*V*] and [+—*NP*], but the string **the students doubt a car* is nevertheless ungrammatical. The nature of the object obviously determines the kind of transitive verb which may be selected. Thus *doubt* may not be followed by a physical object like *car*. This may be expressed formally by assigning the selectional feature [——[+phys.obj.]] to *doubt*. This means that *doubt* cannot (—) occur in the place (—) directly before a phrase which has (+) the property "physical object". The verb *buy* is positive with respect to the same selectional feature.

Thus in the *Aspects* model every lexical formative receives a string of three kinds of features: category, subcategory and selectional features. For *buy*, for example, the string is as follows:

buy: [+*V*], [+—*NP*], [+—[+phys. obj.]], ...

The set of features of a lexical element is called the COMPLEX SYMBOL in *Aspects*, and abbreviated as *C*. The complex symbol of a lexical element contains the conditions under which that element may replace a given dummy symbol.

By way of a number of general rules, the so-called LEXICAL REDUNDANCY RULES, complex symbols can be simplified. Thus a formative with the property [+phys. obj.] is also an *N*. It is sufficient to take only the feature [+phys. obj.] into the complex symbol. A general lexical redundancy rule specifies that all formatives with this feature are at the same time [+*N*]. Much attention is paid to lexical structure in *Aspects*; redundancy rules of various kinds are treated, but we shall not deal with them here.

If lexical insertion is not performed by means of context-free rewrite rules, what kind of grammar is the base grammar *B*? Chomsky calls lexical insertion a transformation, and thus stated, *B* is already a transformational grammar. The reason for calling lexical insertion a transformation is that a phrase marker with certain features (specified in the complex symbol) is replaced by another phrase marker (in which Δ is replaced by a lexical formative). Such substitution transformations, however, are completely local operations on the phrase marker. In fact they do not take the weak generative power of the grammar beyond the reach of a context-free grammar. In other words, lexical insertion could also be realized by means of complicated context-free rules (cf. Peters and Ritchie, 1973). We have also seen that the other modification with respect to the ordinary context-free form, namely the ordering of rules, likewise does not lead to raising the generative power of context-free grammars. It holds, therefore, that the base grammar *B* is weakly equivalent to a context-free grammar; as for the categorial part of the grammar, moreover, there is a high degree of strong equivalence. The output of *B* consists of tree-diagrams with category symbols as nonterminal nodes and lexical formatives as terminal elements, as well as the special boundary symbol ≠ and the dummy symbol Δ (not all dummy symbols need be replaced by lexical formatives; remaining dummy symbols can later be transformationally removed). If the transformation rules do not block when such a diagram is presented as input, we call the diagram a DEEP STRUCTURE of the sentence which will later be derived transformationally. The "language" generated by *B* has the usual notation $L(B)$, and the analyzed language generated by *B*, i.e. the set of phrase markers, is denoted by $A(B)$.

3.1.2. *The Transformational Component*

The function of this component is the transformation of deep structures, by way of derived structures, into SURFACE STRUCTURES. Surface structures are tree-diagrams with terminal strings from which sentences of the language can be derived by morphonological operations. We shall freely call such surface strings *sentences*. It is quite clear that a good deal will be necessary to derive the sentence *John watched the eating of shrimps* from the diagrams in Figure 3.2. Some of the structures generated by B even resist operation by the transformational component, and the transformations are said to *block*. At the end of this paragraph we shall give a more exact description of the conditions under which this occurs. The subset of $A(B)$ for which the transformations do not block is the set of deep structures generated by the transformational grammar. The transformational component, thus, also has the function of *filter*.

The transformational component is a finite ordered sequence of transformations: $T = (T_1, T_2, ..., T_k)$. Each transformation T_i consists of two parts: (1) a STRUCTURAL CONDITION which indicates the domain of the transformation. It defines the conditions which the tree-diagram must satisfy if the transformation is to be applied. In particular, one may find in the structural condition the way in which the tree-diagram will have to be subdivided into terms or *factors* (these are parts of the tree-diagram which will be further defined below). As we shall see, the structural conditions also establish other conditions. (2) A *set of* ELEMENTARY TRANSFORMATIONS. Three types of elementary transformation are described in *Aspects*, the elementary *adjunction*, *substitution* and *deletion* of a factor or string of factors. The transformation consists of the simultaneous performance of such elementary operations, once the tree-diagram has been factorized according to (1). The substitution or deletion of a factor is limited by the PRINCIPLE OF RECOVERABILITY OF DELETIONS: when a string of factors disappears, some trace of it must be left behind. This can happen in two ways. One possibility is that a replica of the string of factors is present elsewhere in the derived tree-diagram. Another possibility is that every grammatical category has a finite number of deletable terminal strings, determined in advance; deletion will therefore cause no complete loss of information. For the moment we shall not discuss this principle, but will return to it in the formal treatment of transformations in paragraph 2 of the present chapter, and in Chapter 5.

The principle is of essential importance in determining the generative power of a transformational grammar.

The transformations are applied in order. We speak of a TRANSFORMATIONAL CYCLE as the operation of going through the list of transformations once. The cycle begins with the subsentences most deeply embedded in the deep structure. These are the parts of the tree-diagram which themselves are tree-diagrams with *S* as root, but in which no further *S* occurs. For every "subtree" with *S* as root, the cycle may be applied only if it has already been applied to every subsentence of *S*. The final cycle deals with the "top *S*" of the deep structure. It is therefore said that a transformational derivation works "cyclically from the bottom up".

A very informal example of such a cyclic application is the derivation of *John watched the eating of shrimps* from the deep structure of Figure 3.2b. The first transformational cycle begins with the subtree for Δ *eat shrimp pl*. In going through the list, we remove the Δ, and nominalize *eat shrimp pl* as *eating of shrimp pl*. In the second cycle, which deals with the main clause, *the* is substituted for *it*, and *sg* is adjoined to *V*. The final surface structure is shown in Figure 3.3.

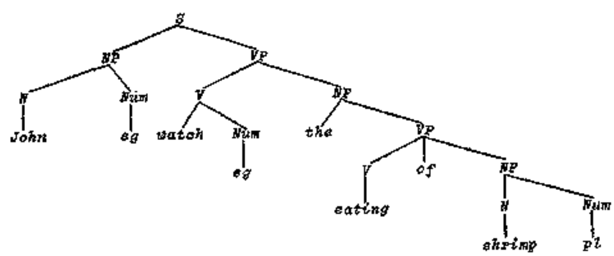


Fig. 3.3. Surface structure for *John watched the eating of shrimps*.

In more (but still in many ways incomplete) detail, we shall now discuss how a transformational grammar might handle a Dutch or German interrogative sentence. In Dutch and German the interrogative is formed by exchanging the positions of subject and (auxiliary) verb. Thus the declarative sentence *de aannemer* *bouwt het huis* (*the contractor builds the house*; German: *der Bauunternehmer baut das Haus*) becomes *bouwt de aannemer het huis?* in the interrogative (*does the contractor build the house?*; German: *baut der Bauunternehmer das Haus?*). The Dutch and German interrogative form is especially suitable for explaining some notions which will be needed in the formal analysis of transformations (paragraph 2 of the present chapter).

The base grammar in Example 3.1 can generate a deep structure for the interrogative sentence *bouwt de aannemer het huis?* If we do not take number and congruence of number into consideration, we can accept Figure 3.4 as a representation of this deep structure.

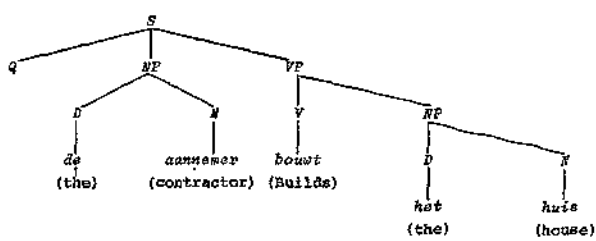


Fig. 3.4. Deep structure of *bouwt de aannemer het huis?* (abbreviated)

The Dutch question transformation T_Q has the factorization Q_1 - NP_2 - V_3 in its structural condition; Q_1, NP_2, and V_3 are single numbered factors. Does the deep structure of Figure 3.4 satisfy this condition? The question is whether the tree-diagram can be subdivided into subtrees in such a way that Q is the root of one subtree, NP is the root of the next subtree to the right, and V is the root of the subtree to the right of that. This is indeed possible; the factorization is represented in Figure 3.5. The tree thus satisfies the structural condition and is correspondingly factorized.

The elementary transformations can now be applied. There is only one of these in our question transformation, the substitution of the V-factor for the Q-factor, or in other words, the third factor comes to take the place of the first. The elementary substitution transformation T_s thus concerns the pair of factors (1,3) for a

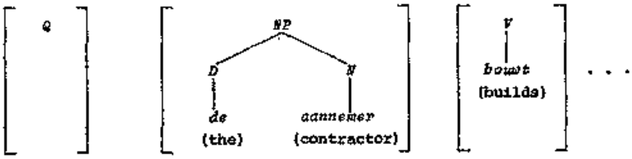


Fig. 3.5. Factorization of the deep structure in Figure 3.4, according to the structural condition of the question transformation.

question transformation. The two parts of the question transformation, the structural condition and the set of elementary transformations, can be summarized in the following notation: $T_Q = (Q_1 - NP_2 - V_3, T_s\ (1{,}3))$. The regrouping of the factors yields the tree-diagram in Figure 3.6. If no more transformations remain to be performed, this tree-diagram is the surface structure of the sentence.

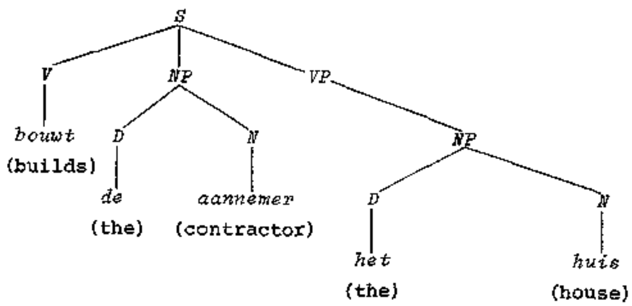


Fig. 3.6. Surface structure for *bouwt de aannemer het huis?*

A complete question transformation for Dutch or German is, of course, more complicated than outlined here. If there is an auxiliary verb, in effect, it is not the main verb, but the auxiliary verb which changes places with the subject; thus the declarative sentence *de aannemer heeft het huis gebouwd (the contractor has built the house)* becomes *heeft de aannemer het huis gebouwd? (has the contractor built the house?)* in the interrogative. There are also other conditions for the question transformation, more difficult to define, not only for Dutch but also for English. Take the dubiously grammatical sentence, for example, **are you undoubtedly ill?* (the Dutch equivalent, **bent u ongetwijfeld ziek?*, mentioned in Kraak and Klooster (1968), has the same difficulties as the English). Differences of opinion might exist on the advisability of the path $S - VP - NP$ in the diagram in Figure 3.6; one would prefer to eliminate the node with *VP*. Such an operation would be called TREE PRUNING, and can be accomplished, as we shall see in paragraph 2.2. of this chapter, by more formal means than those treated in the present paragraph.

Does the transformation satisfy the principle of recoverability? It does in fact. The *Q* disappears from the tree-diagram, but *Q* is the only element in its category. This case shows clearly what recoverability actually takes in. It means that if the transformation of which a given structure is the output is known (Figure 3.6, for example, is the result of a question transformation), then the input structure (Figure 3.4) can be reconstructed.

A distinction is made between OPTIONAL and OBLIGATORY TRANSFORMATIONS. Obligatory transformations *must* be applied, if at a given point in the cycle its structural conditions are fulfilled. Optional transformations *may* be applied under such circumstances.

We have mentioned above that transformations may act as filters. An example of this is the derivation of a relative clause. Consider the sentence *the postman who brought the letter asked for a signature*. This sentence is derived from (a) *the postman asked for a signature* and (b) *the postman brought the letter*. For the purposes of demonstration it is not very important whether (a) and (b) occur in the deep structure of the sentence in conjunction (linked by *and*) or in the form of an embedded constituent. We opt for the latter possibility, and will proceed to illustrate it. We suppose that the sentence is derived from a deep structure with the following terminal string (irrelevant details are overlooked): *the postman # the postman brought the letter # asked for a signature*. The two boundary symbols occur here because of the rewriting of #S# for the embedded sentence; they are mentioned here explicitly because they play a role in the transformation. The structural condition for this relative clause transformation is NP_1 —

$\# - NP_2 - V_3 - NP_4 - \#$, $NP_1 = NP_2$. This means that the tree-diagram must be able to be factorized as indicated, and moreover that the terminal strings of NP_1 and NP_2 are identical. The transformational modification now consists of a number of elementary transformations which yield the following facorization: NP_1 — *who* — V_3 — NP_4, *the postman who brought the letter.* However, there is nothing in the base grammar to prevent the generation of the following terminal string: *the postman # the dustman brought the letter # asked for a signature*, for every grammatical sentence can also be generated as an embedded sentence. This structure, however, is transformationally blocked, because of the identity condition $NP_1 = NP_2$ in the structural condition for the relative clause transformation. If NP_1 = *the postman* and NP_2 = *the dustman*, this condition is not satisfied. A transformational derivation is said to block when there is still one or more boundary symbol in the terminal string at the end of the last transformation cycle. This would be the case with this last example, as the complement transformation would fail. The input structure is "filtered out"; it is not a deep structure.

3.1.3. *Schematic Summary*

Figure 3.7 shows a diagram of the grammar in *Aspects*. It shows that the model generates a deep structure and a surface structure for every sentence in the language. The deep structure contains syntactic information which is necessary and sufficient for a complete semantic interpretation of the sentence. The surface structure gives all syntactic information which is needed for the determination of the morphological and phonological form of the sentence. In the *Aspects* model, these two structures are derived for every sentence in the language, as are all intermediary diagrams which occur in the transformation cycle. The STRUCTURAL DESCRIPTION Σ of a sentence is defined in this model as the pair (δ, ω), where δ is the deep structure, and ω is the surface structure. If, for a given sentence, two or more δ exist, but only one ω, the sentence is said to be DEEP STRUCTURE AMBIGUOUS. An example

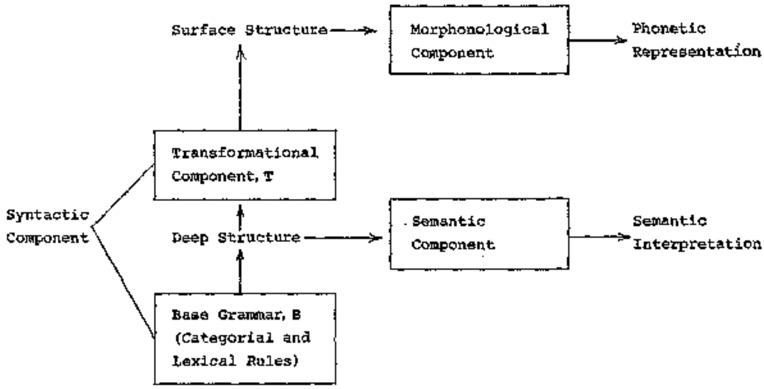


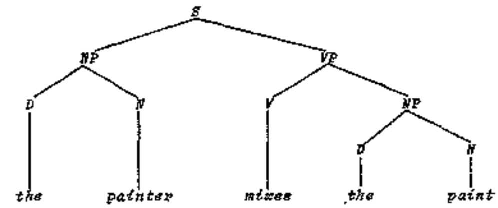
Fig. 3.7. Schema of the *Aspects* model.

of this is *John watched the eating of shrimps*, which has two deep structures in Figure 3.2, and one surface structure in Figure 3.3. The examples in section 2.3.3 under (3), in which context-free grammars failed to represent ambiguity, are deep structure ambiguous; they could be treated adequately by a transformational grammar. If, for a given sentence there are more than one δ, and also two or more ω, the sentence is said to be SURBACE STRUCTURE AMBIGUOUS. The sentence *malicious boys and girls tease the little children* is an example of this.

3.2. TRANSFORMATIONS, FORMAL TREATMENT

3.2.1. *The Labelled Bracketing Notation*

The input and output of transformations are tree-diagrams. The visual advantage of a two-dimensional tree-diagram is a technical disadvantage when it must figure in a written transformation rule. We would prefer to symbolize transformations, like the production rules of a phrase structure grammar, as rewrites of strings. Consequently, we need a string notation which is isomorphous with the tree notation. The common system for this uses "labelled brackets" for the representation of tree-diagrams; the notation is therefore called LABELLED BRACKETING NOTATION. An example of a labelled bracketing is given in Figure 3.8. For every constituent of the tree-



$({}_S({}_{NP}({}_D the)_D({}_N painter)_N)_{NP}({}_{VP}({}_V mixes)_V({}_{NP}({}_D the)_D({}_N paint)_N)_{NP})_{VP})_S$

Fig. 3.8 Tree-diagram and labelled bracketing for the sentence *the painter mixes the paint*

diagram there is a pair of brackets, left and right, each of which is labelled according to the syntactic category of the constituent concerned. The representation of a sentence in labelled bracketing notation is called a LABELLED BRACKETING. But not every labelled bracketing concerns a sentence. In the following we wish to use the notion in a very general sense. We define it, therefore, as follows: For a grammar a LABELLED BRACKETING is every finite string of elements from $V_T \cup V_N \cup L \cup R$, where L is the set of labelled left brackets, $L = \{(_A, A \in V_N\}$, and R is the set of labelled right brackets, $R = \{)_A, A \in V_N\}$. We state without proof that for every tree-diagram in a grammar (see definition in Volume I, section 2.2), there is one and only one labelled bracketing. The inverse does not hold. Every labelled bracketing which corresponds to a tree-diagram is WELL-FORMED. Labelled bracketings which are not well-formed would be, for example, $({}_S a)_V$, $({}_S a({}_S$, $({}_S({}_V a)_V$, and so forth. For grammars one can also define the concept directly as follows. (In the rest of this chapter we shall number the most important definitions to facilitate reference.)

DEFINITION 3.1. A WELL-FORMED LABELLED BRACKETING is every string ω over $V_N \cup V_T \cup L \cup R$, for which either

(1) $\omega \in V_N \cup V_T$ or,

(2) $\omega = ({}_A\psi)_A$ or,

(3) $\omega = \psi\varphi$

where ψ and φ are well-formed labelled bracketings.

(This is called a *recursive definition*; note that although the concept itself is used in the definition, the latter is not tautological.)

A well-formed labelled bracketing is said to be CONNECTED in cases (1) and (2). Thus $({}_S a)_S({}_{NP}({}_N a)_N)_{NP}$ is a well-formed labelled bracketing which is not connected, while $({}_{NP}({}_D a)_D({}_N a)_N)_{NP}$ is connected and consequently also well-formed. A TERMINAL LABELLED BRACKETING is a labelled bracketing with elements exclusively from $V_T \cup L \cup R$.

In order to speak of the terminal string of a tree-diagram, we must be able to remove the brackets. We must, therefore, define the *debracketing function.*

DEFIUITION 3.2. The DEBRACKETIZATION $d[\omega]$ of the labelled bracketing ω is the string which remains when all labelled brackets are removed from ω;

Thus $d[({}_{NP}({}_D a)_D({}_N a)_N)_{NP}] = $ aa.

3.2.2. *A General Definition of Transformations*

The replacement of tree-diagram with tree-diagram in diagram notation becomes the replacement of connected well-formed labelled bracketing with connected well-formed labelled bracketing in labelled bracketing notation. For the general definition of transformations, which is much broader than the definition given in *Aspects* which will be formalized in paragraph 3.2.4 of this chapter, we shall deal only with the rewriting of terminal labelled bracketings. This is in complete agreement with the linguistic use of transformations. (Notice that the deep structure of a sentence corresponds to a terminal labelled bracketing.)

Before presenting the definition, we must first treat two questions. In the first place we must realize that transformations are not ordinary rewrite rules, but *rule schemas*. We have seen rule schemas already, such as $NP \rightarrow NP^n + and + NP$, $n > 0$, in section 2.3.3. A rule schema stands for a possibly infinite set of rewrite rules. Many structural conditions are of this sort. For the Dutch question transformation, we found the condition (much simplified) $Q_1 - NP_2 - V_3$. Every tree-diagram which fulfills this condition lies in the domain of the question transformation. If the grammar generates an indefinite number of noun phrases, there is an indefinite number of tree-diagrams which satisfy this condition. The question transformation is a summary of an infinity of rewrite rules over terminal labelled bracketings. A transformation, thus, indicates how a *set* of terminal labelled bracketings can be rewritten. Let us call such a set a TREE TYPE. The definition of transformations must therefore show that tree types are rewritten as tree types. The fact that transformations are rule schemas is a direct consequence of the linguistic usage of applying transformations to *complete* tree-diagrams. If it were permitted to apply transformations to incomplete tree-diagrams (cf. Figure 2.2 in Volume I), that is, before a terminal derivation is obtained, it would not be necessary to define transformations over *terminal* labelled bracketings. As the tree can usually be completed in various ways, transformations must be schemas.

In the second place, it can occur in linguistics that a particular transformation is applicable in more than one place in the tree-diagram. Suppose that we have a structure which can be factorized as $A - B - A - B - A$, and a transformation whose structural condition is the factorization $A - B - A$. In such a case the transformation could be applied either to the first three factors or to the last three, possibly with differing results. This, however, will rarely be the case in linguistics, especially since every transformational cycle concerns only a strongly limited domain in the tree-diagram. On the other hand it does happen that the structural condition is satisfied in two different places in the tree-diagram, without overlapping (in practice this occurs particularly in phonology; cf. Chomsky and Halle (1968)). With the condition $A - B - A$, we see this in a factorization such as $A - B - A - X - A - B - A$, where X is an arbitrary string of factors. In general, then, a transformation is a *nondeterministic rule*. It transforms a given tree type into a finite set of tree types. This is entirely analogous to the transition rules of non-deterministic automata (cf. Volume I, sections 4.2, 5.2, and 6.1).

Suppose that $W(V_N, V_T)$ is the set of terminal connected well-formed labelled bracketings over nonterminal vocabulary V_N and terminal vocabulary V_T. W is then a set of terminal (complete) tree-diagrams. Let w stand for a possibly infinite subset of W; thus $w \subset W$, and w is a tree type. Let us indicate any finite set of tree types by f. The output of a transformation, as we have just seen, must be such a finite set. The entire set of such finite sets f over $W(V_N, V_T)$ is noted as $F(W(V_N, V_T))$, or simply $F(W)$. This represents "the set of finite sets of tree types". Transformations, then, can be defined as follows:

DEFINITION 3.3. A TRANSFORMATION over (V_N, V_T) is a pair (w, f), where w is a subset of $W(V_N, V_T)$, and f is a subset of $F(W)$.

Equivalent formulations of this are: A transformation maps a subset of W in the subsets of F, and: A transformation T is a subset of the cartesian product of W and F, $T \subset W \times F$.

One way to write a transformation is in the form $w \to f$, just like the notation for production rules. (Notice that this notation differs from the informal notation given in the preceding paragraph. We shall return to his subject in paragraph 2.4 of this chapter.) Thus, the Dutch-German question transformation can be written as:

$$T_Q\colon ({}_SQ({}_{NP}X)_{NP}({}_{VP}({}_VY)_VR)_{VP}U)_S \to \{({}_S({}_VY)_V({}_{NP}X)_{NP}RU)_S\}$$

The subset of W appears before the arrow. The variables X, Y, R, and U stand for well-formed labelled bracketings, and R and U are possibly empty. Because in principle, for each of these variables, an infinity of terminal labelled bracketings can be chosen, the terms to the left of the arrow stand for an infinite set of terminal trees. Notice also that the term to the left is connected: if U is well-formed, then $({}_S$ necessarily corresponds to $)_S$. The variables themselves need not be connected. Thus X can stand for $({}_N John)_N and ({}_N Peter)_N$, where the leftmost $({}_N$ does not correspond to the rightmost $)_N$. The term to the right of the arrow is a finite set with only one element. That element stands for a tree *type*, and thus for a (possibly infinite) set of terminal trees. Its variables (Y, X, R, and U) mean that if a given terminal labelled bracketing is chosen for the term at the left, the same labelled bracketing must be chosen for the same variable in the term at the right.

Let us show that T_Q is applicable to the terminal connected labelled bracketing

$$({}_S Q({}_{NP}({}_D de)_D({}_N aannemer)_N)_{NP}({}_{VP}({}_V bouwt)_V({}_{NP}({}_D het)_D({}_N huis)_N)_{NP})_{VP})_S$$

The variables here have the following values: $X = ({}_D de)_D({}_N aannemer)_N$, $Y = bouwt$, $R = ({}_{NP}({}_D het)_D({}_N huis)_N)_{NP}$, and $U = \lambda$. The transformation changes the labelled bracketing to

$$({}_S({}_V bouwt)_V({}_{NP}({}_D de)_D({}_N aannemer)_N)_{NP}({}_{NP}({}_D het)_D({}_N huis)_N)_{NP})_S.$$

By drawing the tree-diagram for this, the reader will see that tree pruning has taken place, that is, the superfluous VP node in Figure 3.6 has been removed. If we stipulate in the grammar that auxiliaries belong to category V, then the question transformation given here will also provide the correct solution for Dutch and German sentences with auxiliary verbs. The main verb will be found in factor R, and will remain in place during the transformation. Finally, let us point out that the debracketization of this labelled bracketing is precisely the sentence *bouwt de aannemer het huis?*

3.2.3. *The Interfacing of Context-free Grammars and Transformations*

Before returning to the formal facets of transformations in the *Aspects* model (paragraph 3.2.4), we shall first show that it is possible, by the use of the debracketing function d, to give a very simple representation of a transformational grammar.

Let G be a context-free grammar. The language generated by G

is $L(G)$, and the analyzed language is $A(G)$. It is obvious that $L(G)$ is obtained by debracketing the elements of $A(G)$.

It is possible to write a grammar G' in such a way that $L(G') = A(G)$. The sentences of $L(G')$ will be precisely the structural descriptions of the sentences generated by G. This may be seen in the following. Take $G = (V_N, V_T, S, P)$. $G' = (V_N, V'_T, S, P')$ is constructed as follows.

(i) $V'_T = V_T \cup L \cup R$, in which $L = \{(_A \mid A \in V_N\}$ and $R = \{)_A \mid A \in V_N\}$. Thus the sets of labelled left and right brackets are added.

(ii) For every production $A \rightarrow \alpha$ in P, P' will contain a production $A \rightarrow (_A\alpha)_A$.

We shall illustrate, by way of an example, and without proof, that if G' is thus constructed, $L(G') = A(G)$.

EXAMPLE 3.2. Let G have the productions listed below in column (1), and G' the productions listed in column (2).

(1)	(2)
$S \rightarrow NP+VP$	$S \rightarrow (_S NP+VP)_S$
$VP \rightarrow V+NP$	$VP \rightarrow (_{VP} V+NP)_{VP}$
$NP \rightarrow D+N$	$NP \rightarrow (_{NP} D+N)_{NP}$
$D \rightarrow the$	$D \rightarrow (_D the)_D$
$N \rightarrow \{people, animals\}$	$N \rightarrow \{(_N people)_N, (_N animals)_N\}$
$V \rightarrow help$	$V \rightarrow (_V help)_V$

It is not difficult to derive the sentence *the people help the animals* from grammar G. If the corresponding production rules of G' are applied in the same order, we obtain

$$(_S(_{NP}(_D the)_D(_N people)_N)_{NP}(_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$$

as may easily be verified. This sentence in $L(G')$ is precisely the structural description of *the people help the animals* in $L(G)$. If x' is a sentence in $L(G')$, then $x = d(x')$, the debracketization of x', is a sentence in $L(G)$.

In this way, a transformational grammar $TG = (B, T)$ such as in *Aspects*, with a context-free base grammar, can now simply be considered as a triad (B', T, d), in which B' is the context-free grammar which generates as its sentences the structural descriptions of the sentences generated by *B*. The transformational component *T*, will then indicate how such *sentences* (and not tree-diagrams) are to be rewritten. In this case, transformations will replace strings with strings. If a transformation replaces a sentence with a shorter string, we are dealing with a type-0 rule (which is neither of type-1 nor of type-2). Finally, the debracketing function, *d*, acts to remove the brackets after application of the transformations. It still holds, however, even for this (B', T, d) model, that the transformations are not ordinary type-0 rules, but rule schemas. Unfortunately, little is known of the generative power of rule schemas, and of their place (or lack of it) in the hierarchy of grammars.

3.2.4. *The Structure of Transformations in Aspects*

The general definition of transformations (Definition 3.3) includes much more than what is used in *Aspects*, and more than is necessary on empirical linguistic grounds. Every substitution of a tree-diagram for a tree-diagram is included in the general definition of transformation, but in paragraph 3.1.2 we saw that in *Aspects* only three elementary transformations were admitted: adjunction, substitution, and deletion of a factor or string of factors, and this within the limitations of the principle of recoverability. The formulation of this in *Aspects* is quite informal, however, and it is impossible to see precisely what can be done with transformations as long as a much more precise definition is not given. Peters and Ritchie (1973) were the first to perform a formalization of the *Aspects* model, and the results they obtained were surprising, as we shall see in Chapter 5. Without attempting to be exhaustive, we shall present the essence of this formalization. In order to be clear and concise in this, we shall first introduce the concept of *elementary factorization*, which was not used by Peters and Ritchie.

DEFINITION 3.4. The ELEMENTARY FACTORIZATION of a terminal labelled bracketing φ is the ordered set of p elementary factors $[\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p]$, such that $\varphi = \varepsilon_1\ \varepsilon_2 \ldots \varepsilon_p$, where

(i) φ is a component of a connected terminal labelled bracketing (cf. Definition 3.1), and where ELEMENTARY FACTOR is defined in (ii) and (iii):
(ii) ε_i contains one and only one terminal element;
(iii) the leftmost symbol of ε_i is not a right bracket, and the rightmost symbol of ε_i is not a left bracket.

In this way, φ is divided into the smallest possible "terminal" factors, and the boundaries between factors are precisely the phrase boundaries. Thus, the elementary factorization of $\varphi = (_S(_{NP}(_D\mathit{the})_D(_N\mathit{people})_N)_{NP}(_{VP}(_V\mathit{help})_V(_{NP}(_D\mathit{the})_D(_N\mathit{animals})_N)_{NP})_{VP})_S$ is $[(_S(_{NP}(_D\mathit{the})_D, (_N\mathit{people})_N)_{NP}, (_{VP}(_V\mathit{help})_V, (_{NP}(_D\mathit{the})_D, (_N\mathit{animals})_N)_{NP})_{VP})_S]$. In this example, $\varepsilon_2 = (_N\mathit{people})_N)_{NP}$. Notice that not every labelled bracketing has an elementary factorization. This is the case, for example, for labelled bracketings which contain no terminal elements.

DEFINITION 3.5. A FACTORIZATION of a terminal labelled bracketing φ is defined if φ has an elementary factorization $[\varepsilon_1, \varepsilon_2 ..., \varepsilon_n]$. A factorization then is an ordered set of m FACTORS $[\psi_1, \psi_2, ..., \psi_m]$, such that $\varphi = \psi_1\psi_2 ... \psi_m$, in which $\psi_1 = \varepsilon_1\varepsilon_2 ... \varepsilon_i$, $\psi_2 = \varepsilon_{i+1}\varepsilon_{i+2} ... \varepsilon_j, ... , \psi_m = \varepsilon_k\varepsilon_{k+1} ... \varepsilon_{n-1}\varepsilon_n$. In other words, a factorization is a partition of an elementary factorization.

The example given with the preceding definition allows the following factorization, inter alia:

$[(_S(_{NP}(_D\mathit{the})_D, (_N\mathit{people})_N)_{NP}(_{VP}(_V\mathit{help})_V,$
$(_{NP}(_D\mathit{the})_D(_N\mathit{animals})_N)_{NP})_{VP})_S]$

In this factorization, $\psi_2 = \varepsilon_2\varepsilon_3 = (_N\mathit{people})_N)_{NP}(_{VP}(_V\mathit{help})_V$. Another factorization of the same labelled bracketing is

$[(_S(_{NP}(_D\mathit{the})_D(_N\mathit{people})_N)_{NP}(_{VP}(_V\mathit{help})_V,$
$(_{NP}(_D\mathit{the})_D(_N\mathit{animals})_N)_{NP})_{VP})_S]$

Here $\psi_2 = \varepsilon_4\ \varepsilon_5 = (_{NP}(_D\mathit{the})_D(_N\mathit{animals})_N)_{NP})_{VP})_S$.

We can now try to define a very special factorization of a terminal labelled bracketing φ. The factorization should on the one hand not cut through connected well-formed substrings in φ. It should be remembered that these are strings which are either surrounded by corresponding brackets, or strings consisting of just one terminal element (cf. Definition 3.1). This condition means that each connected substring of φ is in its entirety part of a factor in the factorization. On the other hand, the factorization should be as fine as possible, i.e. contain as many factors as possible. As an example, let us consider the case where $\varphi = ({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N})_{NP}({}_{VP}({}_{V}help)_{V}({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N})_{NP}$. This φ has an elementary factorization since φ is a part of a well-formed terminal labelled bracketing. The elementary factorization is (with numbering): $[\varepsilon_1 = ({}_{NP}({}_{D}the)_{D},\ \varepsilon_2 = ({}_{N}people)_{N})_{NP},\ \varepsilon_3 = ({}_{VP}({}_{V}help)_{V},\ \varepsilon_4 = ({}_{NP}({}_{D}the)_{D},\ \varepsilon_5 = ({}_{N}animals)_{N})_{NP}]$. There is only one way to factorize φ in such a way that, on the one hand, each connected labelled bracketing, also the largest, is part of a factor, and, on the other hand, there is a maximum number of such factors. That is the factorization $[\psi_1, \psi_2, \psi_3]$ in which $\psi_1 = \varepsilon_1\varepsilon_2$, $\psi_2 = \varepsilon_3$, $\psi_3 = \varepsilon_4\varepsilon_5$: $[\psi_1 = ({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N})_{NP},\ \psi_2 = ({}_{VP}({}_{V}help)_{V},\ \psi_3 = ({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N})_{NP}]$. Such a factorization of φ is called the *unique factorization* of φ. (See the more detailed treatment of the notion "standard factorization" in Peters and Ritchie, 1973.) A broad definition of this will be sufficient here.

DEFINITION 3.6. The UNIQUE FACTORIZATION of a terminal labelled bracketing φ (defined if φ has an elementary labelled bracketing) is the factorization in which
(i) every substring of φ which is a connected well-formed labelled bracketing is as a whole a part of a factor;
(ii) the factorization is the most minute for which (i) holds, i.e. of all factorizations which fall under (i) the unique factorization counts the largest number of factors.

We offer a few examples of unique factorizations (2) of labelled bracketings (1).

(1) Labelled Bracketing	(2) Unique Factorization
$\varphi_1 = ({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N}$	$[({}_{NP}({}_{D}the)_{D}, ({}_{N}people)_{N}]$
$\varphi_2 = ({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N})_{NP}$	$[({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N})_{NP}]$
$\varphi_3 = ({}_{VP}({}_{V}help)_{V}({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N})_{NP})_{VP}$	$[({}_{VP}({}_{V}help)_{V}({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N})_{NP})_{VP}]$
$\varphi_4 = ({}_{VP}({}_{V}help)_{V}({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N})_{NP}$	$[({}_{VP}({}_{V}help)_{V}, ({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N})_{NP}]$
$\varphi_5 = ({}_{VP}({}_{V}help)_{V}({}_{NP}({}_{D}the)_{D}({}_{N}animals)_{N}$	$[({}_{VP}({}_{V}help)_{V}, ({}_{NP}({}_{D}the)_{D}, ({}_{N}animals)_{N}]$

In the unique factorization of φ, as we have pointed out, every connected labelled bracketing in φ is part of a factor. The first example in column (1), $\varphi_1 = ({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N}$, has the following connected parts: *the*, $({}_{D}the)_{D}$, *people*, and $({}_{N}people)_{N}$. Such a connected part is the same as that which we have called a subtree in paragraph 3.1.2 of this chapter. Each of these parts appears uncut in one of the factors in column (2). The *interior* of a factor is defined as the largest connected part of that factor.

DEFINITION 3.7. The INTERIOR I (ψ) of a factor ψ in a unique factorization is the largest connected labelled bracketing in that factor.

The interior of $({}_{NP}({}_{D}the)_{D}$ is not *the*, but $({}_{D}the)_{D}$; that of $\psi = ({}_{N}people)_{N}$ is not *people*, but $I(\psi) = ({}_{N}people)_{N}$. The interior of φ_2 in column (1) is the labelled bracketing itself. Notice that every factor of a unique factorization has an interior, for every factor contains at least one terminal element. If there is no greater connected unity, that element is the interior. This definition leads directly to the following.

DEFINITION 3.8. The LEFT-HAND EXTERIOR $E_l(\psi)$ of a factor in a unique factorization is the part of the factor to the left of the interior; the RIGHT-HAND EXTERIOR $E_r(\psi)$ of a factor in a unique factorization is the part of the factor to the right of the interior.

The left-hand exterior $E_l(\psi)$ of $({}_{NP}({}_{D}the)_{D}$ is $({}_{NP}$, the righthand exterior $E_r(\psi)$ is λ, because the interior is $({}_{D}the)_{D}$. The lefthand exterior of a factor such as $animals)_{N})_{NP})_{VP})_{S}$ is λ and the righthand exterior is $)_N)_{NP})_{VP})_S$, because *animals* is the interior. The exterior thus consists of the labelled brackets which remain after the interior is removed.

We have just seen that for $\psi = (_N people)_N$, $I(\psi) = (_N people)_N$. This interior has the general form $(_{A_1}(_{A_2}\ldots(_{A_m}\omega)_{A_m}\ldots)_{A_2})_{A_1}$, where ω contains no corresponding exterior brackets. In this example $m = 1$, $A_1 = N$, and $\omega = people$. We call ω the KERNEL of $I(\psi)$, denotated by $K(\psi)$. The kernel of $(_{NP}(_D the)_D(_N people)_N)_{NP}$ is $(_D the)_D$ $(_N people)_N$, in which $(_D$ and $)_N$ are not corresponding brackets. The kernel of $(_{NP}(_N people)_N)_{NP}$ is *people*. If the kernel is removed, that which remains to the left and to the right of it will be called respectively $U_l(\psi)$ and $U_r(\psi)$. Thus for $\psi = (_{NP}(_D the)_D(_N people)_N)_{NP}$, $I(\psi) = \psi$, $K(\psi) = (_D the)_D\,(_N people)_N$, $U_l(\psi) = (_{NP}$ and $U_r(\psi) =)_{NP}$. For $(_{NP}(_N people)_N)_{NP}$, $U_l(\psi) = (_{NP}(_N$ and $U_r(\psi) =)_N)_{NP}$. U_l and U_r always form a symmetric pair. Summing up:

$$\psi = E_l(\psi)I(\psi)E_r(\psi) = E_l(\psi)U_l(\psi)K(\psi)U_r(\psi)E_r(\psi).$$

We shall now define the *content* of a unique factorization as the string of interiors of the factors.

DEFINITION 3.9. The CONTENT $C(\varphi)$ of φ, given the unique factorization $[\psi_1,\ldots,\psi_n]$ of φ, is the string $I(\psi_1)\ I(\psi_2)\ldots I(\psi_n)$, where $I(\psi_i)$ is the interior of ψ_i.
The content is thus defined only if φ has a unique factorization.

Once again our examples are taken from the labelled bracketings in column (1) on page 69. Their contents are given in column (3).

(3) Content
$C(\varphi_1) = (_D the)_D(_N people)_N$
$C(\varphi_2) = (_{NP}(_D the)_D(_N people)_N)_{NP}$
$C(\varphi_3) = (_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP}$
$C(\varphi_4) = (_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP}$
$C(\varphi_5) = (_V help)_V(_D the)_D(_N animals)_N$

The content of a connected labelled bracketing is the labelled bracketing itself, as is the case for φ_2 and φ_3 in column (3).

Just as the content is defined as a string of interiors, we define the REST, $R(\varphi)$, as the string of exteriors of a unique factorization which remains after the interiors have been removed; thus $R(\varphi_1) = ({}_{NP}$, $R(\varphi_2) = \lambda$, $R(\varphi_3) = \lambda$, $R(\varphi_4) = ({}_{VP}$, and $R(\varphi_5) = ({}_{VP}({}_{NP}$.

We are now able to define the elementary transformations of deletion, substitution and adjunction.

DEFINITION 3.10. The ELEMENTARY DELETION of a labelled bracketing φ, $T_d(\varphi)$, is defined as $R(\varphi)$.

The deletion of φ is thus that which remains after the content of φ has been removed. $T_d(\varphi)$, then, can only be defined if $T_d(\varphi)$ has a content. Examples of this (with reference to column (1)) are: $T_d(\varphi_1) = R(\varphi_1) = ({}_{NP}$, $T_d(\varphi_2) = R(\varphi_2) = \lambda$, and so forth.

DEFINITION 3.11. The ELEMENTARY SUBSTITUTION $T_s(\psi, \varphi)$ is the replacement of the interior of ψ with the content of φ, thus $T_s(\psi, \varphi) = E_l(\psi)C(\varphi)E_r(\psi)$.
Substitution is defined only if ψ has an interior, that is, if it is a factor of the unique factorization of a labelled bracketing, and if φ has a content, that is, if it itself has a unique factorization.

Take, for example, $\psi = ({}_{N}people)_{N})_{NP}$ and $\varphi = ({}_{VP}({}_{V}help)_{V}({}_{NP}({}_{D}the)_{D}\,({}_{N}animals)_{N}$. Here $E_l(\psi) = \lambda$, $E_r(\psi) =)_{NP}$, and $C(\varphi) = ({}_{V}help)_{V}({}_{D}the)_{D}({}_{N}animals)_{N}$. Therefore $T_s(\psi,\varphi) = ({}_{V}help)_{V}({}_{D}the)_{D}\,({}_{N}animals)_{N})_{NP}$.

DEFINITION 3.12. The ELEMENTARY LEFT-ADJUNCTION $T_l(\psi, \varphi)$ is defined as $E_l(\psi)U_l(\psi)C(\varphi)K(\psi)U_r(\psi)E_r(\psi)$. The ELEMENTARY RIGHT-ADJUNCTION $T_r(\psi, \varphi)$ is defined as $E_l(\psi)U_l(\psi)K(\psi)C(\varphi)U_r(\psi)E_r(\psi)$. The conditions on ψ and φ for T_l and T_r are the same as in the preceding definition.

As an example of *elementary right-adjunction* we construct the following. Let $\varphi = ({}_{PP}({}_{Prep}in)_{Prep}({}_{NP}({}_{N}Norway)_{N})_{NP})_{PP})_{NP}$, in which *PP* stands for "prepositional phrase" and *Prep* for "preposition", and $\psi = ({}_{S}({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N})_{NP}$. We then have the following values for the various terms of the transformation: $E_l(\psi) = ({}_{S}$, $E_r(\psi) = \lambda$, $U_l(\psi) = ({}_{NP}$, $U_r(\psi) =)_{NP}$, $K(\psi) = ({}_{D}the)_{D}({}_{N}people)_{N}$, and $C(\varphi) = ({}_{PP}({}_{Prep}in)_{Prep}({}_{NP}({}_{N}Norway)_{N})_{NP})_{PP}$. Then $T_r(\psi, \varphi) = ({}_{S}({}_{NP}({}_{D}the)_{D}({}_{N}people)_{N}({}_{PP}({}_{Prep}in)_{Prep}({}_{NP}({}_{N}Norway)_{N})_{NP})_{PP})_{NP}$.

Finally, a convention introduced by Peters and Ritchie, the REDUCTION CONVENTION, should also be mentioned in this connection. One of the two following cases can occur as part of a labelled bracketing, either through peculiarities of the base grammar, or through the transformations.

(i) $({}_A\lambda)_A$, where λ is the null-string. This could occur, for example, through a deletion transformation.

(ii) $({}_{A_1}({}_{A_2}\cdots({}_{A_n}({}_{A_1}\omega)_{A_1})_{A_n}\cdots)_{A_2})_{A_1}$, where ω is a well-formed labelled bracketing. This is called the *nesting* of A_1 in A_1.

In (i) a labelled bracketing is obtained which is not well-formed (cf. Definition 3.1), and in (ii) the labelled bracketing is redundant, because it is said twice that ω belongs to category A_1. The reduction convention states that substrings of type (i) are to be removed as soon as they occur, and that the interior pair of brackets $({}_{A_1},)_{A_1}$ are to be removed when cases of type (ii) occur. Since this is a general convention concerning labelled bracketings, we shall not specifically write "reduced labelled bracketing" when the reduction has taken place. We shall omit the adjective "reduced", for by convention every labelled bracketing is reduced.

Every transformation, such as the question transformation and the complement transformation, is presented as a combination of elementary transformations. For a complete definition of transformations according to the *Aspects* model, two matters must still be worked out: in the first place the manner in which elementary transformations are combined into such a transformation for a given labelled bracketing, and in the second place, the conditions on which the transformation may be applied, that is, the structural condition which the labelled bracketing must satisfy and the general principle of recoverability.

The combination of elementary transformations $\{T_{el_1}, T_{el_2}, \ldots T_{el_p}\}$ for a given labelled bracketing φ can be further defined by indicating the factors which these elementary transformations concern. They may have to do with only one factor when T_{el} is a deletion, or they may concern two factors in cases of substitution and adjunction. For an elementary substitution, for example, $T_s(\psi, \chi)$, the factors to which ψ and χ correspond must be indicated.

This may be done most clearly on the basis of the elementary factorization of φ. For the *Aspects* model, moreover, the whole discussion can be limited to labelled bracketings which are *connected*. Let $[\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n]$ be the elementary factorization of the connected labelled bracketing φ. The elementary transformations are notated as follows.

(i) *Deletion*. $T_d(\varepsilon_{h\to i})$. This means that the factor which consists of the series of elementary factors $\varepsilon_h, \varepsilon_{h+1}, \ldots, \varepsilon_i$ is deleted (if deletion is defined for that factor).

Let $\varphi = (_S(_{NP}(_D the)_D(_N people)_N)_{NP}(_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$, with elementary factorization $[\varepsilon_1 = (_S(_{NP}(_D the)_D$, $\varepsilon_2 = (_N people)_N)_{NP}$, $\varepsilon_3 = (_{VP}(_V help)_V$, $\varepsilon_4 = (_{NP}(_D the)_D$, $\varepsilon_5 = (_N animals)_N)_{NP})_{VP})_S]$. This means that $T_d(\varepsilon_{1\to 2})$ is the deletion of the factor $\varepsilon_1\varepsilon_2$, or $(_S(_{NP}(_D the)_D(_N people)_N)_{NP}$. The interior of this factor is $(_{NP}(_D the)_D(_N people)_N)_{NP}$, and therefore $T_d(\varepsilon_{1\to 2}) = (_S$. The effect of $T_d(\varepsilon_{1\to 2})$ on the original labelled bracketing is thus: $(_S(_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$. For the same φ, we see that $T_d(\varepsilon_{2\to 3})$ is not defined. The factor $\varepsilon_2\varepsilon_3$ is $(_N people)_N)_{NP}$ $(_{VP}(_V help)_V$; it has no interior because it is not a factor in a unique factorization.

(ii) *Substitution*: $T_s(\varepsilon_{h\to i}, \varepsilon_{j\to k})$. This indicates the replacement of the interior of the factor $\varepsilon_h\varepsilon_{h+1} \ldots \varepsilon_i$ with the content of the factor $\varepsilon_j\varepsilon_{j+1} \ldots \varepsilon_k$, if defined.

For φ in our example, $T_s(\varepsilon_{1\to 2}, \varepsilon_{4\to 5})$ means the substitution of $(_{NP}(_D the)_D(_N people)_N)_{NP}$, i.e. the interior of $\varepsilon_1\varepsilon_2 = (_S(_{NP}(_D the)_D$ $(_N people)_N)_{NP}$, by $(_{NP}(_D the)_D(_N animals)_N)_{NP}$, i.e. the content of $\varepsilon_4\varepsilon_5 = (_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$. This yields $(_S(_{NP}(_D the)_D$ $(_N animals)_N)_{NP}(_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$.

(iii) *Adjunction*: $T_l(\varepsilon_{h\to i}, \varepsilon_{j\to k})$ or $T_r(\varepsilon_{h\to i}, \varepsilon_{j\to k})$. For T_l (and similarly for T_r), this means the replacement of factor $\varepsilon_h\varepsilon_{h+1} \ldots \varepsilon_i$ by $E_l(\varepsilon_{h\to i})U_l(\varepsilon_{h\to i})C(\varepsilon_{j\to k})K(\varepsilon_{h\to i})U_r(\varepsilon_{h\to i})E_r(\varepsilon_{h\to i})$, where $\varepsilon_{j\to k}$ is the factor $\varepsilon_j\varepsilon_{j+1} \ldots \varepsilon_k$.

For the labelled bracketing φ in the example, $T_l(\varepsilon_{1\to 2}, \varepsilon_{4\to 5})$ means that φ will be replaced by $(_S(_{NP}\ (_{NP}(_D the)_D\ (_N animals)_N)_{NP}$ $(_D the)_D(_N people)_N)_{NP}(_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$.

Each of the elementary transformations in $\{T_{el_1}, \ldots, T_{el_p}\}$ is of one of these three forms. Notice that in substitution and adjunction the new element is already present in the original labelled bracketing. A transformation, therefore, can introduce no new element from outside the labelled bracketing. This is a relatively restrictive formalization of the *Aspects* model.

We must see to it at this point that the elementary transformations do not "clash". This would occur if the factor $\varepsilon_{h \to i}$ of the one elementary transformation is identical with or overlaps the factor $\varepsilon_{j \to k}$ of another elementary transformation. In that case, what would happen if both elementary transformations were applied at the same time is not defined. In *Aspects* the general solution which was given in section 3.2.2. is not followed. It should be remembered that that solution consisted in defining the output of a transformation nondeterministically as a set. For a deterministic solution, a general condition must be placed upon transformations, namely that the factors concerned may not overlap. If, in formal terms, $\varepsilon_{h_m \to i_m}$ is the first factor of an elementary transformation T_{el_m} in the combination $T_{el_1}, \ldots, T_{el_p}$, where $m = 1, 2, \ldots, p$, then the NON-OVERLAP CONDITION means that $1 \leqslant h_1 \leqslant i_1 < h_2 \leqslant i_2 < h_3 \leqslant i_3 \ldots < h_p \leqslant i_p \leqslant n$, where n is the number of elementary factors in the labelled bracketing.

DEFINITION 3.13. An ELEMENTARY TRANSFORMATIONAL MAPPING with n terms, $M = \{T_{el_1}, T_{el_2}, \ldots T_{el_p}\}$ for a labelled bracketing φ is defined when

(i) φ has an elementary factorization with n elementary factors;

(ii) each of the elementary transformations T_{el} in M is defined;

(iii) M satisfies the non-overlap condition.

It is the labelled bracketing which is obtained by applying $T_{el_1}, \ldots, T_{el_p}$ to φ at the same time. (This definition is somewhat rough; for one more detailed, see the original article, Peters and Ritchie, 1973.) This labelled bracketing is also called the *value* of the transformational mapping. It is determined by convention that if (ii) does not apply, the value of the transformational mapping $M(\varphi)$ is equal to φ.

The notion of "transformational mapping" can now be extended to every factorization of φ:

DEFINITION 3.14. $M = \{T_{el_1}, T_{el_2}, \dots T_{el_p}\}$ is an m-term TRANSFORMATIONAL MAPPING for labelled bracketing φ, if there is a factorization $\psi_1, \psi_2, \dots, \psi_m$ of φ, and an n-term elementary transformational mapping $M' = \{T'_{el_1}, T'_{el_2}, \dots T'_{el_p}\}$, such that for each pair T_{el_m}, T'_{el_m}, it holds that $\psi_{h_m \to i_m} = \varepsilon_{h'_m \to i'_m}$ (notice that it is not necessary that $h_m = h'_m$ or $i_m = i'_m$), and in substitution and adjunction transformations it is true for every pair T_{el_m}, T'_{el_m} that $\psi_{j_m \to k_m} = \psi_{j'_m \to k'_m}$ (where it is again not necessary that $j_m = j'_m$ or $k_m = k'_m$).

The value of the m-term transformational mapping for φ is thus equal to that of the n-term elementary transformational mapping for φ; $M(\varphi) = M'(\varphi)$; The elementary transformations are in fact the same in both cases; only the units chosen for the m-term transformations are greater, or in any case not smaller. If one or more of the elementary transformations in M are not applicable to φ, then by convention $M(\varphi) = \varphi$, i.e. M leaves φ unchanged.

As the last step toward the definition of transformation according to the *Aspects* model, we shall now treat the structural condition and the principle of recoverability. In *Aspects* the structural condition consists of three kinds of data which the labelled bracketing must satisfy, (i) the *"is a" relation,* (ii) the *content-identity relation,* and (iii) the *debracketization relation.*

Suppose that φ has the (not necessarily elementary) factorization $[\psi_1, \dots, \psi_n]$. We may then say the following.

(i) $\psi_{h \to i}$ *is an A,* if the interior of the factor $\psi_h \psi_{h+1} \dots \psi_i$ $(1 \leqslant h \leqslant i \leqslant n)$ can be written as $(_{A_1}(_{A_2}\dots(_{A_m}\omega)_{A_m}\dots)_{A_2})_{A_1}$, where it is true of some A_i $(i = 1, \dots, m)$ that $A_i = A$, and where ω is well-formed.

Example: $(_{NP}(_{D}the)_{D}(_{N}people)_{N})_{NP}$ *is an NP,* $(_{NP}(_{N}people)_{N})_{NP}$ *is an NP,* but also *is an N.*

If φ has n factors, the notation for the fact that $\psi_{h \to i}$ is an A is: $A^n_{h \to i}$.

(ii) $\psi_{h \to i}$ *has the same content as* $\psi_{j \to k}$, if the content of the factor

$\psi_h \psi_{h+1} \ldots \psi_i$ is identical to that of the factor $\psi_j \psi_{j+1} \ldots \psi_k$, thus $C(\psi_{h \to i}) = C(\psi_{j \to k})$, where $1 \leqslant h \leqslant i \leqslant n$ and $1 \leqslant j \leqslant k \leqslant n$. Example: For $(_S(_{NP}(_D the)_D(_N people)_N)_{NP}(_{VP}(_V help)_V(_{NP}(_D the)_D(_N animals)_N)_{NP})_{VP})_S$, it holds that $C(\varepsilon_{1 \to 1}) = C(\varepsilon_{4 \to 4}) = (_D the)_D$.

If φ has n factors, the content-identity relation is written $C^n_{h \to i} = C^n_{j \to k}$.

(iii) $\psi_{h \to i}$ *has debracketization* x, if the debracketization of the factor $\psi_h \psi_{h+1} \ldots \psi_i$ is the terminal string x, thus $d(\psi_{h \to i}$ $(= x$.

DEFINITION 3.15. A STRUCTURAL CONDITION C for an n-term factorization $[\psi_1, \psi_2, \ldots, \psi_m]$ is a combination of n-term properties of types (i), (ii), and (iii).

Finally we shall define the principle of recoverability. This is necessary because we do not wish to call every combination of structural condition C and transformational mapping M a transformation. We wish to speak of transformation only when such a pair (C, M) leaves a "trace" after deletion or substitution. In *Aspects* this is presented in the following form. If the pair (C, M) and the result of the transformational mapping, φ', are given, then there is no more than a finite number of labelled bracketings φ, from which φ' can be derived by means of the mapping (C, M). In the case of more than one φ, we can speak of *structural ambiguity*. The guarantee of recoverability can be given in two ways. This first is that there be a copy in φ' of the string which has been deleted or replaced. The second is that the string which has been deleted or replaced is one of a finite number, determined beforehand, of deletable strings in that syntactic category. In Chapter 5 we shall see that the principle of recoverability is the pivot on which every argument on the generative power of the theory presented in *Aspects*, the theory of the universal base grammar, and the learnability of the language turns. The reason for the introduction of such a principle is to guarantee that an algorithm exists which assigns no more than a finite number of structural descriptions to every sentence in the language.

DEFINITION 3.16. A pair (C, M), in which C is an n-term structural condition and M is a *n-term* transformational mapping, satisfies the PRINCIPLE OF RECOVERABILITY if for every elementary deletion $T_d(\psi_{h\to i})$ and every elementary substitution $T_s(\psi_{h\to i}, \psi_{j\to k})$ in M, one of the two following conditions is met:

(i) After the application of (C, M), there is a copy left of the content of $\psi_{h\to i}$, i.e. there is a pair of natural numbers t and u such that the following property is an element of the structural C: $C^n_{h\to i} = C^n_{t\to u}$, and that M contains no elementary transformations by which $C^n_{t\to u}$ will come partially or completely to be omitted. That is, if M contains elementary transformation $T_d(\psi_{f\to g})$ or $T_s(\psi_{f\to g}, \psi_{v\to w})$ with $t \leqslant f \leqslant u$ or $t \leqslant g \leqslant u$ (and $\psi_{f\to g}$ thus overlaps $\psi_{t\to u}$), then M also contains elementary transformations $T_s(\psi_{y\to z}, \psi_{p\to q})$, $T_l(\psi_{y\to z}, \psi_{p\to q})$ or $T_r(\psi_{y\to z}, \psi_{p\to q})$ such that $p \leqslant t \leqslant u \leqslant q$ (i.e. $\psi_{t\to u}$ is contained in $\psi_{p\to q}$). This guarantees that the content of $\psi_{t\to u}$ nevertheless remains somewhere in the transformational mapping.

(ii) The structural condition C states that $d(\psi_{h\to i})$ is one of a finite number of terminal strings $x_1, ..., x_m$.

A transformation according to the theory presented in *Aspects* can now be defined as follows:

DEFINITION 3.17. A TRANSFORMATION is a pair (C, M), in which C is an n-term structural condition (cf. Definition 3.15), and M is an n-term transformational mapping (cf. Definition 3.14), which fulfills the principle of recoverability (Definition 3.16).

A factorization $[\psi_1, \psi_2 ..., \psi_n]$ is a PROPER ANALYSIS for the transformation (C, M) if each of the n-term properties of C holds for $[\psi_1, ..., \psi_n]$ and if the factorization satisfies the structural conditions specified in Definition 3.15. In this we allow that a factor may be empty. If φ does not have a proper factorization for the transformation $T = (C, M)$, then, by convention, $T(\varphi) = \varphi$, i.e. the transformation leaves φ unchanged. The value of an OPTIONAL transformation of the labelled bracketing φ is the two-term set $\{\varphi, \varphi'\}$ if φ has a proper factorization, and φ if that is not the case. In the former case φ may be changed "at will" to φ', or left unchanged.

In the following example we present a transformation according to the *Aspects* model. It is the passive transformation (*Aspects*, p. 104).

EXAMPLE 3.3. The English passive transformation is a nine-term pair, $T_p = (C, M)$. In other words, a proper factorization for T_p contains nine factors. We shall first give a rough characterization of T_p; the formal discussion will follow.

The nine factors are the following: U_1, NP_2, Aux_3, V_4, W_5, NP_6, X_7, $Pass_8$, Y_9, where U, W, X and Y are more or less arbitrary. T_p changes this string of factors to the string $U_1+NP_6+Aux_3+Pass_8+V_4+W_5+X_7+NP_2+Y_9$.

Formally, the structural condition C for the passive transformation is the following set of properties: $\{NP^9_{2\to 2}, Aux^9_{3\to 3}, V^9_{4\to 4}, NP^9_{6\to 6}, Pass^9_{8\to 8}\}$. (A careful reading of *Aspects* would perhaps demand that it be added that W_5 is not an NP, thus, $\sim NP^9_{5\to 5}$.) This means that in the nine-term factorization the second factor is an NP, the third is an Aux (for "auxiliary verb" including tense), the fourth is a V, the sixth is an NP, and the eighth is of the category *Pass* ("passive"-formative). The nine-term mapping M consists of the following elementary mappings: $M = T_s(\psi_{2\to 2}, \psi_{6\to 6}), T_r(\psi_{3\to 3}, \psi_{8\to 8}), T_d(\psi_{6\to 6}), T_s(\psi_{8\to 8}, \psi_{2\to 2})$. It is obvious that M satisfies the non-overlap condition, and that it is defined for every nine-term factorization in which ψ_2, ψ_3, ψ_6, and ψ_8 have an interior, and ψ_1, ψ_1, and ψ_2 have a content.

Let us now see if the following labelled bracketing has a proper factorization for T_p. $\varphi = ({}_S({}_{NP}\textit{the secretary})_{NP}({}_{PredP}({}_{Aux}({}_{Tense}\textit{pt})_{Tense}({}_{Aspect}\textit{have en})_{Aspect})_{Aux}({}_{VP}({}_V\textit{pass})_V({}_{Prt}\textit{on})_{Prt}({}_{NP}\textit{the mail})_{NP}({}_{Dir}\textit{to the director})_{Dir}({}_{Man}({}_{PP}({}_{Prep}\textit{by})_{Prep}({}_{Passive}\textit{be en})_{Passive})_{PP})_{Man})_{VP}({}_{Time}\textit{yesterday})_{Time})_{PredP})_S$. In this labelled bracketing, *PredP* stands for *predicate phrase*, *pt* for *past tense*, *Prt* for *particle*, *Dir* for *direction*, *Man* for *Manner*, *PP* for *prepositional phrase*. This labelled bracketing obviously supposes a much more extensive base grammar than we have treated here.

There is indeed a proper factorization for φ, namely, in the following nine factors:

$\psi_1 = \lambda$
$\psi_2 = (_S(_{NP}\textit{the secretary})_{NP}$
$\psi_3 = (_{PredP}(_{Aux}(_{Tense}\textit{pt})_{Tense}(_{Aspect}\textit{have en})_{Aspect})_{Aux}$
$\psi_4 = (_{VP}(_V\textit{pass})_V$
$\psi_5 = (_{Prt}\textit{on})_{Prt}$
$\psi_6 = (_{NP}\textit{the mail})_{NP}$
$\psi_7 = (_{Dir}\textit{to the director})_{Dir}(_{Man}(_{PP}(_{Prep}\textit{by})_{Prep}$
$\psi_8 = (_{Passive}\textit{be en})_{passive})_{PP})_{Man})_{VP}$
$\psi_9 = (_{Time}\textit{yesterday})_{Time})_{PredP})_S$

This factorization is a proper analysis because (1) the factorization has the features mentioned under C, namely, it has nine terms, ψ_2 *is an NP*, ψ_3 *is an Aux*, ψ_4 *is a V*, ψ_6 *is an NP*, and ψ_8 *is a Passive*, (2) the factorization allows definition of each of the elementary transformations in M, because ψ_2, ψ_3, ψ_6 and ψ_8 all have interiors, and ψ_6, ψ_8 and ψ_2 have contents.

The transformation $T_p = (C, M)$ gives rise to the following factors:

$\psi_1' = \lambda$ (nothing is said of ψ_1 in M)
$\psi_2' = (_S(_{NP}\textit{the mail})_{NP}$ (by $T_s(\psi_{2\to 2}, \psi_{6\to 6})$)
$\psi_3' = (_{PredP}(_{Aux}(_{Tense}\textit{pt})_{Tense}(_{Aspect}\textit{have en})_{Aspect}(_{Passive}\textit{be en})_{Passive})_{Aux}$ (by $T_r(\psi_{3\to 3}, \psi_{8\to 8})$)
$\psi_4' = (_{VP}(_V\textit{pass})_V$ (nothing is said of ψ_4 in M)
$\psi_5' = (_{Prt}\textit{on})_{Prt}$ (nothing is said of ψ_5 in M)
$\psi_6' = \lambda$ (by $T_d(\psi_{6\to 6})$)
$\psi_7' = (_{Dir}\textit{to the director})_{Dir}(_{Man}(_{PP}(_{Prep}\textit{by})_{Prep}$ (nothing is said of ψ_7 in M)
$\psi_8' = (_{NP}\textit{the secretary})_{NP})_{PP})_{Man})_{VP}$ (by $T_s(\psi_{8\to 8}, \psi_{2\to 2})$)
$\psi_9' = (_{Time}\textit{yesterday})_{Time})_{PredP})_S$ (nothing is said of ψ_9 in M)

The output of the transformation ψ' is thus: $(_S(_{NP}\textit{the mail})_{NP}(_{PredP}(_{Aux}(_{Tense}\textit{pt})_{Tense}(_{Aspect}\textit{have en})_{Aspect}(_{Passive}\textit{be en})_{Passive})_{Aux}(_{VP}(_V\textit{pass})_V(_{Prt}\textit{on})_{Prt}(_{Dir}\textit{to the director})_{Dir}(_{Man}(_{PP}(_{Prep}\textit{by})_{Prep}(_{NP}\textit{the secretary})_{NP})_{PP})_{Man})_{VP}(_{Time}\textit{yesterday})_{Time})_{PredP})_S$. The following sentence is thence derived: *the mail had been passed on to the director by the secretary yesterday.*

In somewhat less detail Peters and Ritchie also give definitions of *transformational cycle* and of *transformational derivation.*

A transformational cycle supposes an ordered list of transformations. We shall call this list $(T_1, T_2, \ldots T_k)$.

DEFINITION 3.18. A TRANSFORMATIONAL CYCLE with reference to $(T_1, \ldots T_k)$ is an ordered set of labelled bracketings $(\varphi_1, \varphi_2, \ldots, \varphi_{k+1})$ for which $T_i(\varphi_i) = \varphi_{i+1}$, $i = 1, 2, \ldots, k$.
Notice that it is not necessary that $\varphi_i \neq \varphi_{i+1}$. This is not the case, in particular, when φ_i has no proper factorization for T_i.

This definition is insufficient when the list also includes optional transformations. It should be remembered that the value of an optional transformation is a set of two labelled bracketings if φ has a proper factorization: $T(\varphi) = \{\varphi', \varphi\}$. We may maintain the definition, however, by the convention that if T_i is optional and $T_i(\varphi_i) = \{\varphi_i', \varphi_i\}$, then $\varphi_{i+1} \in \{\varphi_i', \varphi_i\}$. This means that if the list $(T_1, \ldots, T_k)$ contains optional transformations, the possibility exists that for a given labelled bracketing φ_1 there is more than one transformational cycle with reference to $(T_1, \ldots, T_k)$.

A transformational derivation is a certain series of transformational cycles. We shall first illustrate this with an example. It was stated in paragraph 3.1.1. of this chapter that every derivation in the base begins with $\#S\#$, and that every new S introduced is also surrounded by two boundary symbols. The following string represents a terminal labelled bracketing, derived from the base grammar:

$$\varphi = \#(_{S_5}\alpha_1\#(_{S_1}\alpha_2)_{S_1}\#\alpha_3\#(_{S_2}\alpha_4)_{S_2}\#\alpha_5\#(_{S_4}\alpha_6\#(_{S_3}\alpha_7)_{S_3}\#\alpha_8)_{S_4}\# \alpha_9)_{S_5}\#.$$

Each α in this string is a terminal labelled bracketing which contains neither $(_S$, nor $)_S$, nor $\#$. A transformational derivation is performed as follows. First the right hand brackets $)_S$ in φ are numbered in ascending order from left to right. In the example, this operation has already taken place: $)_{S_1},)_{S_2}, \ldots,)_{S_5}$. Then the corresponding left hand brackets are numbered correspondingly (this has also been done in the example). The first transformational cycle concerns the α between $(_{S_1}$ and $)_{S_1}$, in the present case α_2. The last labelled bracketing in the cycle we call α_2'. The first cycle in the transformational derivation will then have replaced α_2 with α_2'. The second cycle concerns the α between $(_{S_2}$ and $)_{S_2}$, i.e. α_4, which it replaces with α_4'. The third cycle concerns S_3 and replaces α_7 with α_7'. At that moment φ has been changed to

$$\varphi' = \#(_{S_5}\alpha_1\#(_{S_1}\alpha_2')_{S_1}\#\alpha_3\#(_{S_2}\alpha_4')_{S_2}\#\alpha_5\#(_{S_4}\alpha_6\#(_{S_3}\alpha_7')_{S_3}\#\alpha_8)_{S_4}\alpha_9)_{S_5}\#.$$

The following cycle concerns S_4. The first string of this cycle is $\beta_1 = \alpha_6\#(_{S_3}\alpha_7')_{S_3}\#\alpha_8$. Let the result of this cycle be called β_1'. The effect of this cycle is the replacement of φ' by φ'':

$$\varphi'' = \#(_{S_5}\alpha_1\#(_{S_1}\alpha_2')_{S_1}\#\alpha_3\#(_{S_2}\alpha_4')_{S_2}\#\alpha_5\#(_{S_4}\beta_1')_{S_4}\alpha_9)_{S_5}\#.$$

The last cycle concerns S_5. Denote the string between $(_{S_5}$ and $)_{S_5}$ by β_2. β_2 is then the initial string of the cycle; the terminal string is then β_2'. This finally yields $\varphi''' = \#(_{S_5}\beta_2')_{S_5}\#$.

DEFINITION 3.19. (rough definition) The labelled bracketing ω is the result of a TRANSFORMATIONAL DERIVATION from γ with reference to $(T_1, T_2, \ldots, T_k)$, if ω is obtained by applying the list $(T_1, \ldots, T_k)$ first to the subsentence farthest to the left in δ, i.e. the labelled bracketing α for which it holds that $(_S\alpha)_S$ is a component of δ, $)_S$ is the leftmost right hand bracket in φ, and $(_S$ corresponds to $)_S$; the list is then applied to the subsentence which is bordered on the right by the leftmost $)_S$ less one, and so forth until the rightmost $)_S$ is reached.

DEFINITION 3.20. For the transformational grammar $TG = (B, T)$, the labelled bracketings δ and ω are respectively called DEEP STRUCTURE and SURFACE STRUCTURE of the SENTENCE x, if

(i) δ is generated by B,

(ii) ω is the result of a transformational derivation from δ relative to T,

(iii) $\omega = \#(_S\psi)_S\#$, where ψ contains no $\#$,

(iv) $x = d(\psi)$.

The pair $\Sigma = (\delta, \omega)$ is called the STRUCTURAL DESCRIPTION of x. The LANGUAGE generated by TG consists of the strings in V_T^* for which such a pair (δ, ω) exists. The third condition in fact contains a formalization of the notion of BLOCKING. If at the end of a transformational derivation boundary symbols remain within the outer S-brackets, then neither deep structure nor surface structure nor sentence are defined. An example of such blocking for a relative clause was given in paragraph 3.1.2. When a derivation blocks, the labelled bracketing in question is filtered out.

The filtering function of the transformational component is limited, because only one pair of boundary symbols per subsentence can be removed. However this filtering function can be increased when the base grammar is modified in such a way that the boundary symbol can also be introduced elsewhere than around S. Proposals in this direction were also made in *Aspects* (p. 191).

In paragraph 3.1.1 of this chapter the dummy symbol Δ was presented as an element of V_T. But neither in *Aspects*, nor in Peters and Ritchie (1972), nor in the above do we find guarantees that Δ will not occur in ω. Chomsky suggests that the symbol be removed transformationally, while Peters and Ritchie allow it to appear in ω, supposing, apparently, that the morphological rules will deal with it. We shall leave this as an open question here.

In closing this paragraph, we would make a few general remarks on the formalization which has been presented. *Aspects of the Theory of Syntax* is an informal book which allows very divergent kinds of formalization. It is most unfortunate that efforts to formalize the conceptual framework of that work, perhaps the most widely read and often quoted in modern linguistics, were only made seven years after its first publication. The *Aspects* model is only an outline of a linguistic theory, and it is difficult, if not impossible, to determine whether or not the theory should be further developed in that direction. The aim of Peters and Ritchie was to define the notion of "transformational grammar" as precisely as possible without leaving the framework of *Aspects*. Despite the fact that this leads to formulations which at times are not very graceful from a mathematical point of view, such an undertaking is well founded. In effect, if such an extremely restrictive definition of transformation should show that the theory is still too strong, i.e. generates too much, there would be good reason to diverge from the given outline. In Chapter 5 we shall show that this is indeed the case.

3.3. LATER DEVELOPMENTS

One of the important principles of the *Aspects* model is that transformations do not change meaning; they are PARAPHRASTIC. In this the theory presented in *Aspects* is clearly different from that of *Syntactic Structures*, in which a transformational syntax was developed which was completely independent of semantic considerations. The criterion for the correctness of syntactic rules lay in the justification of the distinction between grammatical and ungrammatical. In *Aspects*, paraphrase relations come to play an important part. Chomsky does this following a proposal by Katz and Postal (1964): transformations are paraphrastic, that is, meaning-preserving. This is shown in the diagram of Figure 3.7. The semantic interpretation is determined exclusively by an input from the base grammar; deep structures carry all the syntactic information necessary for semantic interpretation, while transformations have no influence on this.

Soon after the publication of *Aspects* this point of view was called into doubt. Let us consider a few of the classic examples responsible for this.

(1) *Reflexive Pronouns.*
The sentence *Nixon voted for himself* goes back to the deep structure *Nixon voted for Nixon*. The relation is a paraphrastic reflexive transformation. But if that is the case for the preceding example, then *everybody voted for himself* must be based on the deep structure *everybody voted for everybody*. This relation, however, is clearly not paraphrastic.

(2) *Relative Constructions.*
The sentence *the postman who brought the letter, asked for a signature* goes back to *the postman brought the letter* and *the postman asked for a signature*, by way of a paraphrastic relative clause

transformation. However it is not the case that the sentence *all postmen who bring letters ask for signatures* is paraphrastically related to the pair *all postmen bring letters* and *all postmen ask for signatures.*

(3) *Coordinations.*
John is both shy and fresh is paraphrastically related to *John is shy* and *John is fresh.* The same coordination transformation, however, is not paraphrastic in the derivation of *no number is both even and odd* from *no number is even* and *no number is odd.*

(4) *Passives.*
The sentence *the target was not hit by the arrows* is based, via a paraphrastic passive transformation, on *the arrows did not hit the target.* The same transformation, however, is not paraphrastic if *the target was not hit by many arrows* is derived from *many arrows did not hit the target* (in one of the two possible readings of this sentence). There is a clear difference in meaning here (to which we shall return in greater detail in Chapter 4, paragraph 3).

The problems occur especially when quantifiers such as *many, all, every* are combined with negations or with a condition of identity of reference, i.e. where the deep structure contains two elements with the same denotation.

It is true that cases (1) to (4) show that some transformations of the *Aspects* model are not meaning-preserving, but they do consistently make the correct prediction concerning grammaticality. The supposed deep structure and corresponding transformations in all cases lead to grammatical sentences, while sometimes a change in meaning takes place, and sometimes not. We seem to return to the principle enunciated in *Syntactic Structures,* that transformations account for the grammaticality of sentences, but that they are not necessarily paraphrastic. This is precisely the conclusion drawn by Chomsky in his publications after *Aspects.* Transformations sometimes change meaning and consequently not only the deep structure is determinant for the semantic interpretation, but also the surface structure. Aside from this, however, the deviations from *Aspects* remained rather minor. There is still an "independent" syntax which generates the sentence and its structural description, and some aspects of this structural description form the input of the semantic component which gives the semantic interpretation of the sentence. Some change has been made on the question of which aspects of the structural description undergo semantic interpretation. In *Aspects* only the deep structure underwent semantic interpretation, but in Chomsky's later work certain features of the surface structure do also. This new approach is called *interpretative semantics*; it is a question of independently motivated syntactic structures which undergo semantic interpretation. Little is known of the form of such a semantic interpretation, but it is certain that a semantic structure has a different form than a syntactic structure.

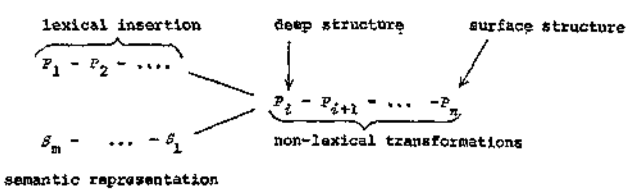
Although examples (1) to (4) show transformational changes of meaning while grammaticality is maintained, cases are known in which the application of transformations of the *Aspects* type causes the loss of grammaticality.

(5) *Each-Hopping.*
This transformation derives *the men each won a prize* from *each of the men won a prize*. Sometimes this leads to changes of meaning, as is the case, for example, when *the men each hate his brothers* is derived from *each of the men hates his brothers*. But the problem here is that when reflexive pronouns are present, this transformation leads to ungrammaticality. From *each of the men shaved himself* it should be possible to derive **the men each shaved himself* (see Hall-Partee (1971b) for a more detailed analysis of this and similar phenomena) and one is led to wonder whether the deep structures generated by the *Aspects* model are really adequate. The doubt is increased by examples of the following kind (Lakoff, 1970):

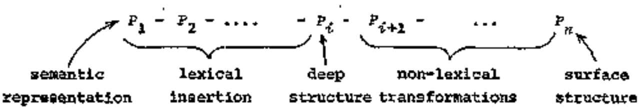
(6) *Ambiguities.*
The sentence *two dogs followed a hundred sheep* is ambiguous. It can mean that the two dogs followed a hundred sheep each, that each of the hundred sheep was followed by two dogs, or that a total of a hundred sheep was followed by a total of two dogs.

This ambiguity is not "lexical" (as is the case in *the tank is filled with water* where the ambiguity is caused by the fact that *tank* can have more than one meaning here). It is not a surface structure ambiguity either, for there is only one possible surface parsing for this sentence (cf. definitions of ambiguity in paragraph 3.1.3 of this chapter). Therefore there must be more than one deep structure for the sentence, i.e. there is more than one transformational derivation. The theory in *Aspects*, however, gives only one transformational derivation for it, i.e. one deep structure.

Examples like (5) and (6) suggest that the notion of "deep structure", as used in the *Aspects* model, is not adequate. Perhaps it is possible to maintain the paraphrastic character of transformations by making the grammar generate more adequate underlying structures. This may be attempted by specifying the range of quantifiers in the underlying structure (cf. Chapter 4, paragraph 3). A more radical approach is also possible, namely, by abandoning the interpretative character of the semantic component, or in other words, by abolishing to a certain extent, the distinction between semantic and syntactic rules. To clarify this, we return briefly to the *Aspects* model. In it, the categorial rules generate a structure P_1, with the dummy symbol and grammatical formatives as terminal elements. The lexical insertion rules transform these gradually into P_i, the deep structure, and the transformational cycles finally transform P_i into P_n, the surface structure. The deep structure P_i is interpreted semantically by means of the semantic rules. In other words, P_i leads successively to structures S_1, S_2, ..., S_m, where S_m is the semantic interpretation. All these "operations" are nothing other than formal relations among structures. They have no direction or temporal order. The *Aspects* schema may therefore be represented as below in (7).



The first proposal is to replace the above schema with (8):



The deep structure in (8) is derived from a sequence of semantic operations, which regulate, among other things, lexical insertion and hierarchical ordering. This approach gave rise to the term *generative semantics* as opposed to interpretative semantics. Much discussion, however, (Chomsky 1971, Katz 1971, Lakoff 1971, Chomsky (1972) has shown that formulations (7) and (8) are notational variants of each other which no empirical test can distinguish.

It is true that that which was left to semantics in the interpretative approach must be recuperated by "syntactical means" in the generative approach, for there is no longer any separate semantic component. On this point the generative semanticists have proposed a number of interesting modifications regarding the *Aspects* theory, concerning, among other things, the mechanism of lexical insertion. In the *Aspects* model, lexical insertion is accomplished by the replacement of a dummy symbol with a lexical element, if the phrase marker satisfies the restrictions defined in the complex symbol of that element. Throughout this process, however, the phrase marker remains unchanged. Generative semanticists, on the other hand, perform lexical insertion by replacing subtrees of a semantic interpretation with lexical elements. The terminal elements of such subtrees are abstract elements, "semantic primitives" which, for simplicity, are denoted by words. The classic example (though now somewhat bypassed, see Fodor 1970) is presented in Figure 3.9. It indicates how the word *kill* is inserted during the generation of *John kills Mary*. At a certain stage of derivation this sentence has the underlying structure shown in Figure 3.9a. This contains an explicit semantic interpretation of *kill*, and the meaning is represented as a nesting of the predicates *cause*, *become*, *not* and *alive*, all of which are semantic primitives. A number of transformations (*predicate* *raising*) change this structure through b. and c. to d., and the subtree under *Pred* can then be replaced by *kill*. This yields e. = P_t, the deep structure of *John kills Mary*. The surface structure f. = P_n then follows, details aside, by way of a *subject raising* transformation (more is said on this in Chapter 4, paragraph 3). All transformations here are paraphrastic; thus a. is synonymous with f., and even without the (optional) intermediate transformations the semantic primitives in a. can be replaced directly by the corresponding lexical elements. This, by means of a few obligatory transformations will yield the sentence *John causes Mary to become not alive*, which must then be synonymous with *John kills Mary*.

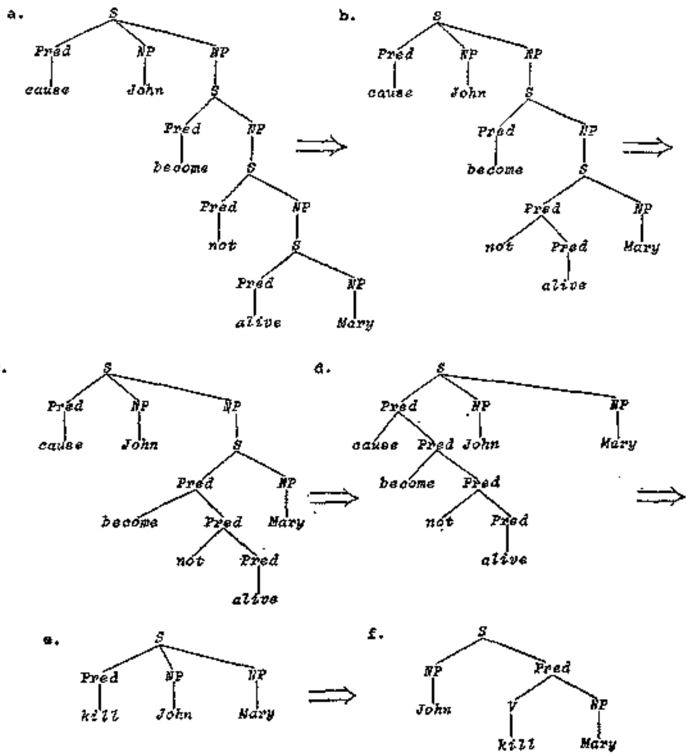


Fig. 3.9. Underlying structures for *John kills Mary*.

The analysis of structures more abstract than deep structures has several other advantages. These have to do with the range of quantifiers in natural languages (cf. Chapter 4, paragraph 3), with presuppositions (the sentence *the man who stole the money lives in Canada* presupposes that money was stolen, and *John never works after five o'clock* presupposes that John sometimes does work before five o'clock, although this does not follow logically), with topic-comment relations, and with focus. Topic-comment relations are usually marked by emphasis in the surface structure; thus *the letter has* ARRIVED is said when the listener expects comment on the letter, while the topic of *the* LETTER *has arrived* is *arrived*. Focus is that which the speaker himself thinks important in the sentence and which is in English usually marked by word order. The active/passive distinction is often a matter of focus; compare, for example, *the mayor opened the council meeting at eight o'clock* and *the council meeting was opened by the mayor at eight o'clock*, with *the mayor* and *the council meeting* as respective focuses.

All of these matters would fall under the semantic component in the *Aspects* model, if the difference between interpretative and generative semantics were limited to the difference between (7) and (8). But generative semanticists hold that the differences are greater. They argue that schema (8) is also unsatisfactory, for the rules of lexical insertion are not applied *en bloc*, but rather some lexical elements are inserted only after one or more non-lexical transformational cycles. If this proves to be the case, it will mean that the notion of "deep structure" as a distinct phrase marker in the derivation of a sentence will no longer be tenable.

Interpretative semanticists have a clear and detailed syntactic theory, but they are not very specific on the structure of semantic interpretation, though there is some tendency to correct this (see, for example, Jackendoff 1969). Generative semanticists, on the other hand, have enriched linguistics with many new semantic insights, but they are not very explicit on the syntactic mechanisms which would transform their underlying structures into sentences. What, for example, are the limitations on the alternation of lexical and other transformations? One new concept in this connection is that of DERIVATIONAL CONSTRAINT. A derivational constraint is a condition on the well-formedness of a transformational derivation as a whole, apart from the correctness of each individual transformational step. There is only one example of this in *Aspects*, the condition that transformations be applied cyclically and in a given order. At present, many other derivational constraints are being added to this, they are essentially conditions on pairs of (not necessarily directly consecutive) tree-diagrams in the transformational derivation. An example of a derivational constraint is the reduction of stress on the auxiliary verb; *Sam is happy*, for example, has the variant *Sam's happy* where the stress on *is* has been reduced. This optional transformation, however, may not be applied, if, somewhere in the transformational derivation, the element which follows the verb has been deleted. An example of this is *Max is happier than Sam is these days*, for which there is no stressless variant *Max is happier than Sam's these days*. This is therefore a condition on which the well-formedness of the entire derivation will depend. Lakoff (1971) remarks that in this respect generative semantics far outstrips the *Aspects* theory. This is indeed the case, and this new syntactic concept might well be justified from a linguistic point of view (although there is scarcely any agreement on the matter, cf. Chomsky (1972). But this new theory has in fact only *removed* limitations. A whole arsenal of new blocking mechanisms has been added to the filtering function of transformations in the *Aspects* model; with these new mechanisms, any enumerable set of sentences whatsoever can, in principle, be defined by a transformational grammar. Derivational constraints only raise the generative power of transformational grammars, and, as we shall see in Chapter 5, there is decidedly no need of that. Just as interpretative semantics is in need of more specific semantic rules, generative semantics needs a much more restricted syntax.

To summarize, we can state that it has proven impossible to maintain both the notion of "deep structure" as presented in *Aspects*, and the principle of paraphrastic transformations. When the former is set aside, generative semantics results, and when the latter is abandoned, interpretative semantics results. The question is to what extent the two trends may be variants from a formal point of view. But most of the syntactic modifications within the generative semantics group are enlargements with respect to the *Aspects* model, with all the serious disadvantages to be discussed in Chapter 5. As far as content is concerned, however, a short time of generative semantics has seen the growth of important insights into lexical structure, presuppositions, focus, and topic-comment relations.

4

MIXED MODELS II: OTHER TRANSFORMATIONAL GRAMMARS

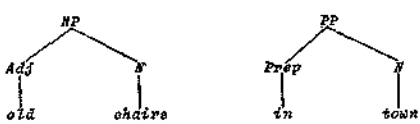
4.1. REASONS FOR FINDING ALTERNATIVE MODELS

The form of the transformations in a mixed model is largely determined by the nature of the base grammar. In the *Aspects* model the base grammar is a phrase structure grammar, and also after the publication of *Aspects*, transformational linguistics has tended to use phrase structure grammars as base grammars, and consequently transformations have retained the essentials of the originally indicated form. In Chapter 2, paragraph 5 it was mentioned that through the use of phrase structure grammars as base grammars the traditional advantages of phrase structure grammars could be taken into a more complete theory of natural languages. At the same time, many of the weaknesses of such grammars could be met by means of transformation rules. It was noticed that a number of the problems with phrase structure grammars are due to the impossibility of assigning more than one tree-diagram or phrase marker to a sentence at a time; in principle transformational grammars can solve this and many other problems.

But the formalism of phrase structure grammars, even within the framework of transformational grammars, still has a number of unattractive points, and this has led linguists to seek other bases which might be able to represent certain linguistic insights in a more natural way. This in turn has resulted in several alternative proposals concerning the structure of the transformational component of the grammar. To give some impression of those unattractive points, we shall mention a few linguistic notions which could not be built into a transformational grammar with a phrase structure grammar as base, unless accompanied by the necessary auxiliary constructions.

(1) *Endocentric versus Exocentric Constructions.*
These notions, first introduced by Bloomfield, are closely connected with that of "distribution". A construction is called ENDOCENTRIC if it contains a part which has the same distribution as the construction itself; the part can always take the place of the entire construction. Nearly any sentence in which *old chairs* occurs corresponds to an equally acceptable sentence in which only *chairs* occurs. Consider, for example, *take all the old chairs outside* and *take all the chairs outside*, or *old chairs creak* and *chairs creak*. *Old chairs* is an endocentric construction, the *head* of which is *chairs*. Some endocentric constructions have more than one head, as, for example, in *old chairs and tables*, where both *chairs* and *tables* are heads. All constructions which are not endocentric are exocentric. *In town* is an exocentric construction, because *John lives in town* corresponds to no sentence **John lives in* or **John lives town*.

A phrase structure grammar can express such relations only with difficulty. There is no natural distinction between tree-diagrams such as the following:



We must therefore establish a convention according to which *N* is always the head of the *NP* by which it is directly dominated, whereas the same does not hold for *Prep* and *PP*, or for *N* and *PP*. Such conventions are not superfluous; they are explicitly required for the correct representation of the structural conditions of certain transformations. It is, for instance, a condition for *tree pruning* (i.e. the removal of superfluous nodes in the tree-diagram, or of superfluous brackets $(_A$, $)_A$ in the labelled bracketing) that the head of the syntactic category *A* has been transformationally deleted. But then the head needs an independent definition for each possible constituent.

(2) *Dependency.*

Closely related to the preceding point is the fact that phrase structure grammars cannot give a simple representation of syntactic dependencies. Although *in town* is an exocentric construction, *town* is in a certain respect dependent on *in*, because it is connected with the rest of the sentence by means of the preposition. In *John lives in town, town* is related to *lives* by way of *in*, just as in the relative construction *the postman who brought the letter, brought the letter* is dependent on *who* in its relation to *the postman.* Such intuitive dependencies may be found in nearly every construction. It is not among the most difficult linguistic judgments to indicate the element through which a phrase is related to the rest of the sentence. The notion of "dependency" is extremely important to a number of linguistic theories, such as those of Harris and of Tesnière. Phrase structure grammars are remarkably unsuited for representing dependencies. They are designed for categorizing phrases hierarchically, and are good systems for expressing the "*is a* relation" (*old chairs* "*is a*" noun phrase, etc.; cf. Chapter 3, paragraph 2.4), not for representing dependencies.

(3) *The Sentence as Modifier and as Complement.*

In the relative construction mentioned above, *who brought the letter* is dependent on *the postman.* In the *Aspects* model, this construction is generated in the base grammar by means of sentence embedding, the recursive introduction of the symbol *S.* Precisely the same mechanism is used for the derivation of a sentence such as *I know that the postman brought the letter.* But in the first case, *the postman brought the letter* is an adjunction or modifier of *postman,* while in the second case it is the object-complement of the sentence. Intuitively there is a great difference between the addition of a modifier to a given sentence structure (Wundt calls this an *associative* relationship; cf. Chapter 2, paragraph 3.1) and the elaboration of part of that sentence structure, such as the object (Wundt calls this an *apperceptive* relationship). This distinction is completely neglected when a phrase structure grammar is used as the base grammar. In both cases, sentence embedding is used as the generative mechanism.

(4) *Functional Relations.*
Dependencies indicate the general lines of the functional relations within the sentence. In Chapter 3, paragraph 1.1, definitions of such functional relations were given as "subject of", "object of", etc. It should be remembered that functional relations are defined on the basis of the production rules, through the notion of "direct dominance". Such definitions are not only very indirect from the point of view of intuition, but they become impossible when a serious effort is made to express the *case relations* within the sentence in that way. An example should make this clearer. In the sentence *John gave the boy the money*, there are two noun phrases within the verb phrase, *the boy* and *the money*. Which of these is the "direct object"? For which of the two does the relation [*NP*, *VP*] hold? In the given definition it is implicitly supposed that there is only one noun phrase having the relation [*NP*, *VP*]. This may perhaps be the case, and the sentence may have a deep structure such as *(John (gave ((the money) (to the boy))))*, where *the boy* is no longer in the relation [*NP*, *VP*]. But if we wish to use such relational definitions for all case relations, and not only for the direct object ("objective case") and the indirect object ("dative case"), we reach an impasse. In the sentence *John went by train to Amsterdam*, *by train* is an "instrumental" case, and *to Amsterdam* is a "locative". But these are two coordinated prepositional phrases, both of which proceed from the rewriting of the verb phrase *VP*, according to no intrinsic order. Therefore relational definitions in terms of direct dominance can make no distinction between locative and instrument. Other problems with case relations such as "agent" and "dative" also occur in this connection. But case must be explicitly marked for the various parts of a structural description, because the semantic interpretation of the sentence is based precisely on this information. It is possible to realize this in the lexical insertion rules, by adding special syntactic

case markers to the complex symbol of a word. But then either an ambiguous situation will result in which some relations (such as "main verb of") will be defined configurationally and others (such as case relations) lexically, or the situation will be such that *all* relations will be defined lexically. From this we may conclude that a phrase structure grammar is not a natural means for expressing functional relations (we shall return to this in paragraph 5 of this chapter).

(5) *Hierarchy.*
In section 2.3.3 we pointed out that too much hierarchy may result from coordination. We showed that phrase structure grammars had to be extended with rule schemas (of the form $A \rightarrow B^n$) in order to avoid giving a pseudo-hierarchical description to a construction which is intuitively coordinative. In a transformational grammar with a phrase structure grammar as base this problem still remains unsolved (cf. Dik, 1968); extra mechanisms such as rule schemas are again required there. More generally, the use of phrase structure grammars easily leads to spurious hierarchy in linguistic descriptions. Every linguistic refinement leads either to the introduction of new nonterminal elements which are more or less "intuitive", or to an elaboration of the hierarchy by recursive sentence embedding (such as in Figure 3.9), in which case there is intuitively no longer a relationship between the length of the sentence and the extent of the hierarchy. Both options are unattractive from a linguistic point of view (we shall return to this in Chapter 5), but they are also unattractive from a psycholinguistic point of view. Many extremely subtle distinctions in the nonterminal vocabulary correspond to no "psychological reality" whatsoever (see Volume 3, section 3.1.), and there is no evidence that the native speaker, in understanding or producing sentences, constantly uses complicated hierarchies. The psycholinguist will have more use for a grammar in which all syntactic information is stored in the terminal vocabulary, than for a grammar which consists principally of the rewrite relations among highly abstract syntactic categories. The native speaker can then be described from

the point of view of a detailed lexicon which gives for every word the way in which it may be combined with other words as well as the functional relations which can be expressed with it. A model with an excess of hierarchy is psychologically unattractive.

In this chapter we shall discuss a number of alternative base grammars which, in varying degrees, avoid the difficulties mentioned in (1) to (5). A real comparison of the advantages and disadvantages of the various formulations is not possible at the present stage. The reason for this is that for most base grammars the form of the corresponding transformational component has at best only partially been elaborated. But even when the respective transformational components have been completely formalized, the decisive comparison must be based on the way in which the various transformational grammars can treat a number of "representative" linguistic problems in detail. On this point information is still scarce for all the alternative models.

Until a convincing comparison of the models can be made, the choice among them should be determined by the aims of the investigator. The practicing linguist will be inclined to use phrase structure grammars as base grammars because a great many problems and solutions in modern linguistics have been formulated within that framework. For the psycholinguist, however, such considerations are much less pressing and it might be more fruitful for him to use other types of grammars, more closely related to models of human linguistic behavior. The ideal situation would be one in which the mathematical relations among the various formulations were known in detail. The most workable formalization for a given linguistic or psycholinguistic problem could then be chosen without loss of contact with other formulations. In some cases such relations are already known, as we shall see in the following paragraph.

4.2. CATEGORIAL GRAMMARS

The history of these grammars goes back to the work of the Polish logicians Leśniewski and Ajdukiewicz, who developed a "theory of semantic categories", not for natural languages, but for artificial languages of logic, especially connected with the "Polish notation" in logic. Later categorial grammars came to be used for the description of natural languages, particularly through the work of Bar-Hillel. The predominating developments in the grammars discussed in the preceding chapter drew attention away from categorial grammars, and it is only since a relatively short time that they are seriously presented as bases for transformational grammars (by Lyons 1968, Miller 1968, Geach 1970, Lewis 1970, and others).

A categorial grammar *CG* is characterized by a finite VOCABULARY V, a small (finite) set of PRIMITIVE CATEGORIES Cp, including a special element S, a RULE OF RULES R which indicate how COMPLEX CATEGORIES can be derived from primitive categories, and, finally, a LEXICAL (ASSIGNMENT) FUNCTION A, which indicates the categories to which vocabulary elements belong. We shall first offer an example of this. Suppose that we have two vocabulary elements, *John* and *eats*, and two primitive categories, S and N. The following rule R_1 is then introduced; by it complex categories can be derived:

R_1: If C_1 and C_2 are categories, then $C_1\backslash C_2$ is also a category.

Because S and N are categories, $S\backslash N$ is also a category, and because $S\backslash N$ is a category, $(S\backslash N)\backslash N$, $N\backslash(S\backslash N)$, $(S\backslash N)\backslash S$, $S\backslash(S\backslash N)$, etc are also categories. Each of the words in the vocabulary is assigned one or more of these "category names" by the function A, for example, *John* is an N, and *eats* is an $N\backslash S$. We define the general LEFT CANCELLATION RULE as follows:

α reduces to β if $\alpha = C_1 + (C_1\backslash C_2)$ and $\beta = C_2$, where α and β are strings of categories.

In the example, *John eats* corresponds to the string $N + N\backslash S$, and the cancellation rule states that that string reduces to S. Therefore the string *John eats* belongs to category S. Given the categories of the vocabulary elements and the cancellation rule, we can determine the category of a phrase. If the string reduces to S, the phrase is said to be a SENTENCE in $L(CG)$.

If we wish to give a simple description of the sentence *John eats apples*, it must be possible for us to make additions also to the *right* of *eat*. For this we introduce rule R_2.

R_2: If C_1 and C_2 are categories, then C_1/C_2 is also a category.

The RIGHT CANCELLATION RULE, belonging to R_2, is defined as follows:

α reduces to β if $\alpha = (C_1/C_2) + C_2$, and $\beta = C_1$.

If both R_1 and R_2 hold, then complex categories such as the following may be formed: $N\backslash S$ (by R_1), $(N/S)/(N/S)$ (by R_2), $(N\backslash S)/N$ (by R_1 and R_2), and so forth.

Suppose that the function A assigns to *eats* both the above mentioned category $N\backslash S$ and the category $(N\backslash S)/N$. Let *apples* belong to the category N. Does the string *John eats apples* belong to $L(CG)$? This holds by definition if the categories of the string reduce to S. Figure 4.1 represents the reduction of this sentence to S. The dotted lines show how the reductions take place, and it is not difficult to see in this derivation the reflexion of a derivation in a context-free grammar; we shall return to this point later.

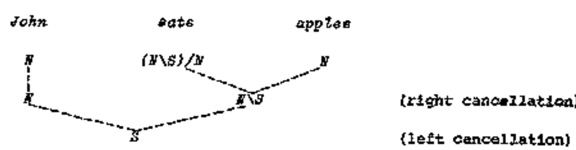


Fig. 4.1. Categorial reduction for the sentence *John eats apples*.

We can assign the category $(N\backslash S)/N$ to all transitive verbs, and the category $N\backslash S$ to all intransitive verbs. Verbs such as *eat* which can be both transitive and intransitive are assigned both categories. The notation for this is as follows: $A(eat) = \{N\backslash S, (N\backslash S)/N\}$. We can go on with other kinds of words; adjectives such as *fat* are N/N, adverbs such as *much* are $(N\backslash S)\backslash(N\backslash S)$. It would be an instructive exercise to reduce the sentence *fat John eats much* with these categories.

If a categorial grammar has only R_1 and the corresponding left cancellation rule, or only R_2 and the right cancellation rule, it is called UNIDIRECTIONAL; if it has both rules, it is called BIDIRECTIONAL. When used without further indication, *CG* will stand for a bidirectional categorial grammar; *UCG* will be used when express reference is made to a unidirectional categorial grammar. Bar-Hillel (1964), however, has proven that bidirectional and unidirectional categorial grammars are weakly equivalent.

At this point we can define categorial grammars formally.

A CATEGORIAL GRAMMAR is a system $CG = (V, C, R, S, A)$, in which V is a finite vocabulary, C is a finite set of primitive categories, R is a set of rules for the generation of categories, $S \in C$, and A is a function which assigns a set of categories (primitive or derived) to each of the elements of V.

A categorial grammar is UNIDIRECTIONAL if R contains one rule (R_1 or R_2), and BIDIRECTIONAL if R contains both R_1 and R_2. A string $x = a_1a_2 \ldots a_n$ in V^* belongs to category Y if there is a string of categories $C_1, C_2, \ldots, C_n$ such that (i) $C_i \in A(a_i)$ (i.e., C_i is an element of the set of categories which the function A has assigned to the vocabulary element a_i. Thus, for example, $N\backslash S \in A$ *(eat)*, because $A(eat) = \{N\backslash S, (N\backslash S)\backslash N\}$), and (ii) the string $C_1C_2 \ldots C_n$ reduces to Y. It is said that $x \in V^*$ is a SENTENCE if x belongs to category S. The LANGUAGE $L(CG)$ accepted or generated by CG is the set of sentences accepted or generated by CG. With categorial grammars, just as with automata, we speak rather of "accepting" than of "generating". When a sentence is presented as input, the categorial grammar passes through a series of reductions until the "final state" S is reached, just as an automaton reaches a state at which the sentence is accepted.

Bar-Hillel (1964), together with Gaifman and Shamir, has proven that categorial grammars are weakly equivalent to context-free grammars. We shall not give the proof here, we shall only show by means of an example how a weakly equivalent context-free grammar can be constructed for a given categorial grammar.

EXAMPLE 4.1. Take categorial grammar $CG = (V, C, R, S, A)$. An equivalent context-free grammar $CFG = (V_N, V_T, P, S)$ is constructed as follows. $V_T = V$, $V_N = C \cup W$, where W is the set of all categories which are assigned by the function A to the elements of $V \cup W$ is, of course, finite. The productions in P are composed as follows:

(i) If $C_1\backslash C_2$ is a complex category in W, then $C_2 \rightarrow C_1 + (C_1\backslash C_2)$ is a production in P.

(ii) If C_1/C_2 is a complex category in W, then $C_1 \rightarrow (C_1/C_2) + C_2$ is a production in P.

(iii) If C_i is a (possibly complex) category in W, assigned to vocabulary element a_i in V, then $C_i \rightarrow a_i$ is a production in P, for every a_i in V.

For the above example, let $CG = (V, C, R, S, A)$, with $V = \{John, eats, apples\}$, $C = \{S, N\}$, $R = R_1 \cup R_2$, and A as follows:

$A(John) = \{N\}$

$A(eats) = \{(N\backslash S)/N, N\backslash S\}$

$A(apples) = \{N\}$

Then the equivalent context-free grammar is $CFG = (V_N, V_T, P, S)$, with $V_N = \{S, N, (N\backslash S)/N, N\backslash S\}$, $V_T = \{John, eats, apples\}$, and the following productions in P:

$S \rightarrow N+(N\backslash S)$	(according to (i))
$N\backslash S \rightarrow (N\backslash S)/N+N$	(according to (ii))
$N \rightarrow John$	(according to (iii))
$N \rightarrow apples$	(according to (iii))
$(N\backslash S)/N \rightarrow eats$	(according to (iii))
$N\backslash S \rightarrow eats$	(according to (iii))

The reader can verify that the phrase marker in Figure 4.1 may be derived by this context-free grammar.

A categorial grammar is an ideal means for expressing endocentricity. It is not difficult to arrange a categorial grammar in such a way that an endocentric phrase has the same category as its head. Let us return to the example *old chairs* (from 4.1, under (1)). If we assign the category N/N to all the adjectives in the categorial grammar, *old chairs* is $(N/N)+N$, which reduces to N. Similarly, an adverb can be assigned the category $(N/N)/(N/N)$, and consequently the phrase *very old* will have the category string $((N/N)/(N/N))+(N/N)$, which reduces to N/N, the category of *old*.

But this advantage does not simply extend to dependencies in general. In the example given, it is not the case that a verb phrase is of the same category as the transitive main verb (the verb phrase is $N\backslash S$, and the main verb is $(N\backslash S)/N$). This is indeed in agreement with the fact that the verb phrase is not endocentric but exocentric; yet it would be preferable to have the dependent phrase in the more complex category, as is the case with endocentric constructions (Lyons (1968) also makes this proposal). But that does not hold here; the more complex category is that of the verb, while the dependent noun is of a primitive category. There are also arguments for a reversed approach; if a word has the function of "link" between two other words or phrases, as is the case with transitive verbs, prepositions or relative pronouns, it should have the more complex category, so that the dependent categories to the left and to the right of it might lend themselves to reduction.[1] If, however, the simpler category is assigned in general to the dependent element, the natural advantage of categorial grammars in the description of endocentric constructions is lost, and the head of the construction (or the independent element) no longer receives the same simple category as the entire construction. A categorial grammar can thus give adequate representation of either endocentricity or dependence, but not both at the same time.

There are many imaginable variations on the theme of "categorial grammar". Notice that a categorial grammar, as defined above, unites precisely *two* categories with every reduction. Such a grammar is strongly equivalent to a context-free grammar in Chomsky normal-form (cf. Volume I, Chapter 2, paragraph 3.1), as may easilv be verified on the basis of the construction in Example 4.1. One may also seek strong equivalence with other types of context-free grammars, for example, grammars in Greibach normal-form (cf. Volume I, Chapter 2, paragraph 3.2). To do this, the following rule must be introduced to replace rules R_1 and R_2.

R_3: If $C, C_1, C_2, \ldots, C_n$ are primitive categories, then $C/C_1C_2\ldots C_n$ is also a category.

[1] Such words resemble the *functors* of formal logic (cf. Curry 1961); there the dependent elements are the arguments.

The corresponding cancellation rule is the following:

> α reduces to β if $\alpha = (C/C_1C_2...C_n) + C_1 + C_2 + ... + C_n$ and $\beta = C$, where $C, C_1, ..., C_n$ are primitive categories.

EXAMPLE 4.2. Suppose that we have the following context-free grammar in Greibach normal-form: $CFG = (V_N, V_T, P, S)$, with $V_N = \{S, V, P, A, B, D, N\}$, $V_T = \{fast, in, John, park, runs, the, very\}$, and the following productions in P:

$S \rightarrow John + V + P$	$B \rightarrow fast$
$V \rightarrow runs + A + B$	$D \rightarrow the$
$P \rightarrow in + D + N$	$N \rightarrow park$
$A \rightarrow very$	

All the productions are of the form $A \rightarrow a\alpha$, with $a \in V_T$ and $\alpha \in V_N^*$. This grammar generates the sentence *John runs very fast in the park.* A strongly equivalent categorial grammar is $CG = (V_T, C, R_3, S, A)$, with $C = \{S, V, P, A, B, D, N\}$, and A as follows:

$A(fast) = \{B\}$	$A(in) = \{P/DN\}$
$A(very) = \{A\}$	$A(John) = \{S/VP\}$
$A(the) = \{D\}$	$A(park) = \{N\}$
	$A(runs) = \{V/AB\}$

Figure 4.2 shows the reduction of the sentence *John runs very fast in the park*. The dotted lines show the isomorphism between this diagram and a phrase marker for the derivation with a grammar in Greibach normal-form.

It follows from the fact that grammars in Chomsky normal-form are weakly equivalent to grammars in Greibach normal-form, that

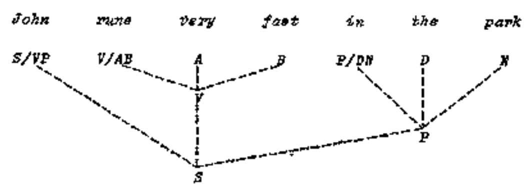


Fig. 4.2. Categorial reduction of the sentence *John runs very fast in the park* (Example 4.2).

the corresponding categorial grammars are weakly equivalent. This means that R_3 adds nothing to the weak generative power of categorial grammars. Because unidirectional and bidirectional categorial grammars are weakly equivalent, we can extend R_3 to a bidirectional rule R_4, without losing equivalence with context-free grammars:

R_4: If C, A_i, and B_j ($i = 1,..., n$, $j = 1,..., m$) are primitive categories, then $A_1A_2...A_n\backslash C/B_1B_2 ... B_m$ is also a category.

The corresponding cancellation rule is:

α reduces to β if $\alpha = A_1 + A_2 + ... + A_n + (A_1A_2...A_n\backslash C/B_1B_2...B_m) + B_1 + B_2 + ... + B_m$, and $\beta = C$.

In this way entire strings of categories to the left and to the right of the complex category can be eliminated.

Because of the weak equivalence of all these grammars, the choice among these possibilities is determined exclusively by consideration of the descriptive adequacy of the grammar. Thus bidirectional grammars have the advantage over unidirectional grammars that a natural representation of both left and right adjunctions is possible. Compare, for example, *the old chairs* and *the chairs here*. A left adjunction, such as *old*, to a noun, receives the category N/N; a right adjunction, such as *here*, receives the category $N\backslash N$. In both cases the categories related to the noun (N) *chairs* reduce to N. The linguist might prefer to derive these adjunctions transformationally (from *the chairs are old* and *the chairs are in this place* respectively) but it may not be taken for granted that every distinction between left and right adjunction can be expressed in the most satisfactory way by means of transformational derivation. Consider, for example, adverbial phrases of place and time which may occur in various places in the sentence, or tense morphemes which sometimes appear to the left, sometimes to the right of the verb: *John* will *come* and *John come*-s. The cancellation of entire strings of categories is an attractive point with respect to the base grammar, for by it various types of verbs can be characterized very well. The verbs in Table 4.1 give a (unidirectional) example of this.

TABLE 4.1. Complex Categories for Verbs.

Category	Verb	Basic Form	Sentence
S/N	*walk, sit, eat*	*walk (John)*	*John walks*
S/NN	*kill, eat*	*eat (John, apples)*	*John eats apples*
S/NNN	*give, send, tell*	*give (John, apples, children)*	*John gives apples to the children*
S/S	*continue, begin*	*begin (the bell rings)*	*the ringing of the bell begins*
S/NS	*say, think, know*	*say (John, it is raining)*	*John says that it is raining*
S/SN	*amaze, enjoy*	*amaze (the sun is shining, me)*	*that the sun is shining amazes me*
S/NNS	*tell, ask*	*ask (John, Peter, it is raining)*	*John asks Peter if it is raining*
S/NSS	*explain, relate to*	*relate to (I, John is hungry, John is growing)*	*I relate John's hunger to his growing*

It is possible in this way to express various functional relations, especially case relations. In Chapter 8, paragraph 1 we mentioned that in a phrase structure grammar such information can only be given in the lexicon. As a categorial grammar is essentially nothing more than a lexicon, it is not at all surprising that case relations can be formulated very naturally in it.

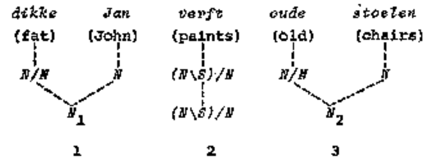
Some of the problems raised by phrase structure grammars (cf. Chapter 2, paragraph 3.3) were solved transformationally (cf. Chapter 3). One of those problems was that of discontinuous constituents (e.g. *I saw the man yesterday whom you told me about*). Context-sensitive grammars were able to deal with this, but not in a very convincing way, namely, by exchanging the grammatical categories of the various elements (cf. Chapter 2, paragraph 4). One might imagine a generalization of categorial grammars in which discontinuities could be formulated simply without changes of category. The generalization exists in that the *continuity restriction* implicit in the cancellation rule of R_3 is dropped. This means that $C/C_1C_2...C_n$ may be reduced to C if this complex category occurs in a string in which C_1, C_2, ..., C_n occur in this order, but not necessarily without interruption. More precisely, while retaining R_3 one can make the convention that reduction can take place if the string is of the form $\alpha_1 C_1 \alpha_2 C_2 ... \alpha_n C_n \alpha_{n+1}$ where α, α_2, ..., α_{n+1} are strings of zero or more categories, and the complex category occurs in one of the α's. The place of the complex category with respect to the primitive categories is therefore no longer important. For the sentence *I saw the man yesterday, whom you told me about*, the reduction of *I saw the man whom you told me about* is not hampered by the fact that *yesterday* breaks the sequence *man, whom*. The categories of these elements remain nevertheless unchanged. If the continuity restriction is abandoned, the argument for a bidirectional categorial grammar loses its value, for adjunction can then take place freely to the left and to the right. Even the interruption caused by the verb itself (e.g. in *John kills Peter*, $N + S/NN + N$) does not block the reduction. A number of typical transformational phenomena can thus be treated in this way. By dropping the continuity restriction, however, we raise the weak generative power of the grammar above that of a context-free grammar, but it is not known to what degree.

The power can also be raised by dropping the *ordering restriction* in the cancellation rule. In that case, one would allow that $C/C_1C_2...C_2...C_n$ reduce to C, when the complex category is followed by some permutation of C_1, C_2, ..., C_n. Interchanges, such as of the positions of particle and object *(John put on his coat, John put his coat on)*, can be expressed categorially in this way. In this case also, the degree to which the generative power is increased is unknown. If both the continuity restriction and the ordering restriction are dropped, the grammar is called an UNRESTRICTED categorial grammar. In such a grammar, only the elements which can occur in a sentence at the same time are specified; the order in which they may occur is not specified. In other words, every permutation of elements yields a new sentence. Unrestricted categorial grammars are equivalent to the systems called *set-systems* in connection with context-free grammars. A set-system has rewrite rules, the output of which does not consist of strings, but of *unordered* sets of elements; their formal structure was studied by Curry (1961) and by Šaumjan and Soboleva (1963). Obviously the generative power of such grammars is considerably greater than that of restricted categorial grammars and that of context-free grammars. Although they solve a number of problems (discontinuity, interchanges), they also raise new problems, such as the way to deal with restrictions on word order.

Linguistic literature offers no serious attempt whatsoever to define a transformational component for a categorial base grammar. If the base is restricted in the usual way, the transformational component will tend to function in the same way as that of the *Aspects* model, that is, in the adjunction, deletion, and substitution of subtrees. In terms of categorial grammars, a subtree is a category, primitive or complex, and consequently transformations will consist of the rewriting of strings of categories as strings of categories. The structural condition of a transformation will then specify whether a given string of elements is appropriate for the transformation. It therefore contains a string of categories; the string to be transformed has a proper analysis for the transformation if it can be reduced to that string of categories. Transformations thus consist of substitutions, adjunction and deletions, as well as *categorial changes*. The following example should make this more clear; it has no linguistic pretensions, however.

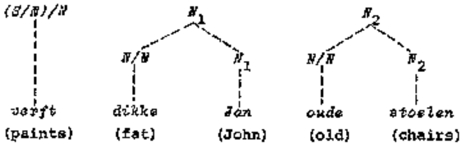
EXAMPLE 4.3. Suppose that the structural condition for the German or Dutch question transformation T_Q (cf. Chapter 2, paragraph 2.2) is $N + (N\backslash S)/X + X + Y$, where X and Y may be empty.
1 2 3 4

Does the Dutch string *dikke Jan verft oude stoelen (fat John paints old chairs)* fall into the domain of T_Q? Let us suppose that the base grammar assigns the following categories to the string: *dikke (fat)* $= N/N$, *Jan* (John) $= N$, *verft (paints)* $= (N\backslash S)/N$, *oude (old)* $= N/N$, *stoelen (chairs)* $= N$, and that the grammar is based on R_1 and R_2 with the corresponding cancellation rules. In that case the sentence can indeed be reduced to the structural condition of T_Q, as follows:



in which $X = N$, and $Y = \lambda$.

The transformation then changes the places of 1 and 2, and alters the category of 2 as follows: $(N\backslash S)/X \rightarrow (S/X)/N$. In general this will yield the new string $\underset{2}{(S/X)/N} + \underset{1}{N} + \underset{3}{X} + \underset{4}{Y}$. Applied to the example, this will yield:



The change of category is necessary to retain the possibility of reduction to S after the transformation. Concerning the structure of transformations we can state that on the one hand it is very easy to indicate the structural condition for a transformation: the domain of a transformation can be given as a string of (possibly complex) categories. But on the other hand it is sometimes necessary to make category changes which are hardly natural or attractive. This occurs especially with deletion transformations, where rather arbitrary category changes are needed.

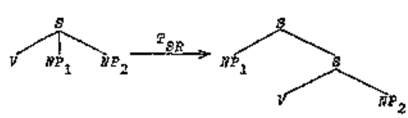
To close this paragraph, we shall summarize the advantages and disadvantages of the use of categorial grammars for natural languages. Among the advantages, categorial grammars are particularly well suited for representing word order, and for the hierarchical description of phrase structure. Categorial grammars can give a satisfactory formulation of the concept of "endocentric construction". The nonterminal vocabulary is very limited, but functional relations can nevertheless be expressed simply. Unrestricted categorial grammars can easily deal with discontinuities and free word order. Syntax is contained completely in the lexicon. Concerning the transformational component, categorial grammars can formulate structural conditions very elegantly, and in such a way that one can easily determine whether or not a given string satisfies such a condition.

Among the disadvantages, categorial grammars do not represent syntactic dependence satisfactorily, and lexical elements are often assigned multiple or very complex categories. The transformational component presents problems which as yet cannot be treated adequately, such as the arbitrary nature of category changes, especially in cases of deletion.

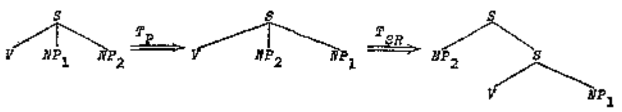
Until now the actual use of categorial grammars has been limited to the solution of problems in other grammars. We shall meet examples of this in Chapter 4, paragraphs 3 and 5. In Chapter 6, paragraph 2 we shall mention a contribution to the development of probabilistic categorial grammars.

4.3. OPERATOR GRAMMARS

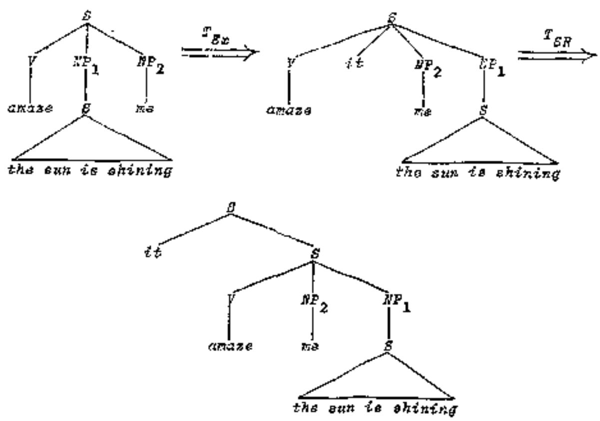
Closely related to categorial grammars, various proposals may be found in linguistic literature to represent base grammars as systems of logic. This implies replacing the subject-and-predicate construction of the *Aspects* theory with constructions of the form predicate-and-arguments, also called operator-and-operands or functor-and-arguments. Some authors take the main verb as the operator, and noun phrases as arguments (Harman 1970); the elementary syntactic rule is thus $S \rightarrow V + NP_1 + NP_2 + \ldots + NP_n$. The role of transformations is essentially the re-ordering of noun phrases. One such substitution is called SUBJECT RAISING (mentioned above in Chapter 3, paragraph 3), T_{SR}. The following diagram shows this transformation for the case where there are two noun phrases:



This gives the usual subject-and-predicate relation. Harman shows that it is in many ways better to place this T_{SR} at the end of the cycle. This can be illustrated with the treatment of the passive transformation. This moves the first NP to the end of the string, and when that is done before the T_{SR} takes place, the following transformational sequence is obtained:



A condition for this is that NP_1 contains something like *by* in the underlying form, but Harman gives no details on this. Table 4.1 shows that the verb *amaze* can have a sentence and a noun phrase as its arguments. The basic form for the sentence *that the sun is shining amazes me* is *amaze (the sun is shining, me)*. The same base form can also underlie the sentence *It amazes me that the sun is shining*. Harman shows that this sentence can be obtained by a transformational substitution of arguments followed by *subject raising*. The substitution transformation here is called *extrapolation*, T_{Ex}; it moves NP_1, leaving *it* behind. The cycle is as follows:[1]



[1] The triangles in these diagrams stand for the subtrees the internal structure of which is left unspecified.

Elements other than the main verb, especially *quantifiers* and *negation*, may also be described as operators. The argument of these operators is *S*. Quantifiers also contain *variables*. We shall show that there are important reasons for the introduction of such operators in the description of underlying structures. In Chapter 3, paragraph 3 we noticed that quantifiers (*all, every, many, some, a few, one*, etc.) lead to problems in the *Aspects* theory. The range or domain of a quantifier can be changed by transformations, and the result of this may be that the transformation is not strictly paraphrastic. In the following example, the passive variant of (1) according to the *Aspect* model is (2), but there is a noticeable difference in meaning between (2) and one of the possible readings of (1).

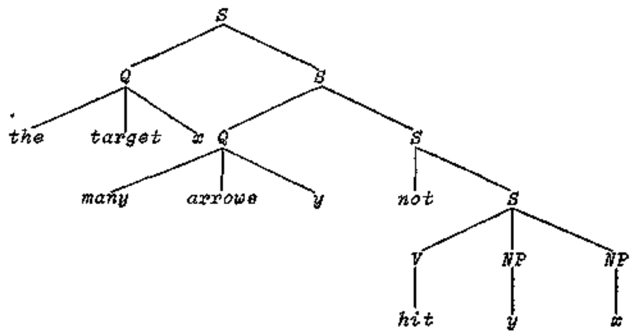
(1) *many arrows did not hit the target*
(2) *the target was not hit by many arrows*

We would paraphrase (1) with (3), and (2) with (4):

(3) *there are many (arrows which did not hit the target)*
(4) *it is not the case that (many arrows hit the target)*

In these paraphrases, we have placed the range of the operator (respectively *many* and *not*) between parentheses. From this we can see the difference immediately: in (3), and therefore also in (1), the operator *not* lies within the range of the operator *many*, whereas in (4), and therefore also in (2), the operator *many* lies within the range of the operator *not*. Chomsky pointed out this difference in *Aspects* (p. 224), but did not account for it in terms of different deep structures for (1) and (2).

Harman suggests the deep structures (5) and (6) (cf. Figure 4.3) for sentences (1) and (2) respectively, introducing the quantifier with the rule $S \rightarrow QS$, and limiting the range of the quantifier to the following *S*. The variables in these constructions make it possible to indicate the identity of certain elements in the deep structure. This is a very general need in linguistics, and is not limited to the treatment of quantifiers. Another example is pronominalization: *John washes John* is not synonymous with *John*



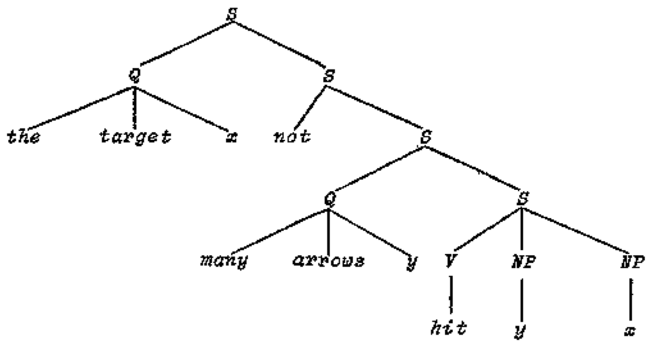


Fig. 4.3. Deep structures for *many arrows did not hit the target* (5), and *the target was not hit by many arrows* (6).

washes himself. The first sentence may be changed to the second only if the first *John* in *John washes John* is identical with the second *John.* This can be indicated in the deep structure by means of variables. Still another example is coordination. In *Clara takes her book and goes to school,* the subject of *goes* is *Clara.* Deletion of a second mention of *Clara* has taken place on the basis of its identity with the first mention. This also may be represented in the deep structure by means of variables, as is the case in (5) and in (6).

Beside verbs, quantifiers, and negation, several other linguistic categories also lend themselves to treatment in terms of operators. Seuren (1969) made one of the first proposals in this field. He treated not only quantifiers but also *qualifiers* as operators. The latter category takes in the already mentioned *negation*, but also *question* ("I ask if *S*"), *imperative* ("I request that *S*"), *assertion* ("I assert that *S*"), *suggestion* ("I suggest that *S*"), as well as *tense*, sometimes in combination with modal verbs (*can*, *must*, etc.). Nesting of operators is possible in Seuren's base grammar: the result of one operator is the argument of another operator. Thus negation may lie within the range of the question operator and not vice versa. Seuren calls the smallest non-nested element the NUCLEUS; in essence this is a string of subject, main verb, objects, and prepositional phrases. Thus Seuren omits Harman's very starting point, namely the definition of the main verb as an operator and the definition of other phrases as arguments. In this respect Seuren's position is quite remarkable, because in dealing with the nucleus he discusses some of its properties which would justify formulation in terms of operators in that case as well. They are, in particular, the relative lack of importance of the order of elements in the nucleus, and the dominant role of the main verb in the selection of the various phrases within the nucleus. In fact Seuren's nucleus grammar could easily be described as a categorial grammar without ordering restriction (cf. Chapter 4, paragraph 2), in which the main verb has a complex category as in Table 4.1, and the restrictions on the other phrases in the nucleus are stated in the category of the main verb. Categorial grammars were developed precisely for the representation of operator-operand constructions.

Harman's analysis only shows *how* operators can be used in a base grammar. Seuren develops the operator approach in much greater detail, but his is a hybrid system in which operators occur outside the nucleus, but not inside it. Neither author gives a systematic treatment of a transformational component.

The recent operator grammar presented by Harris (1970a) is much more comprehensive. It shows a striking degree of agreement with the work of the generative semanticists, although, due to historical and terminological circumstances, there is no question of any interaction. This is extremely unfortunate since Harris supports his system with an abundance of linguistic analyses which are also very essential to the generative-semantics point of view, and which are often of the same tenor as the arguments advanced in that camp.

Some difficulty in reading Harris' work is caused by the distributional framework from which he works. His method consists of the isolation of certain distributional dependences among syntactic elements, followed by the systematic description of them. This method of "working back" from the surface contrasts with the generative method in which the grammar is considered as a sentence-generating system. But Harris' operator grammar can also be represented as a generative grammar. In consonance with the general approach of this book, we shall attempt to give a generative summary of the Harris model. For further detail from the linguistic point of view, we refer to the original publication (Harris 1970a).

We shall begin the description with the construction of a KERNEL SENTENCE. This is a very simple sentence, with a minimum of operators; for the moment we shall limit the number of operators to one. The verb is an important operator, and we refer the reader again to Table 4.1, which was composed on the basis of Harris' survey of verb types. In Harris' notation, a verb which is a predicate over two noun phrases is of the category V_{NN}; it should be followed in the base by two noun elements, $V_{NN}+N+N$. This may also be expressed in a rewrite rule, $S \rightarrow V_{NN}+N+N$. Lexical insertion might yield, for example, *stroke*+*John*+*the dog*. Only one transformation is performed on such a string. Harris calls this transformation GLOBAL PROJECTION; it is identical with *subject raising*, in which the first argument changes places with the operator. This will yield a PROTO-SENTENCE, such as *John*+*strokes*+*the dog*. There is a system of morphophonemic rules by which the protosentence is given a morphemic realization, in this case, the kernel sentence *John strokes the dog*.

The kernel sentences of a language are finite in number. The generative power of the grammar resides in two groups of transformations, PARAPHRASTIC TRANSFORMATIONS, and EXPANSIONS (*incremental transformations* in Harris' terminology). Paraphrastic transformations operate on proto-sentences, and, as their name indicates, do not lead to changes in meaning. At most they lead to changes in the relationship between the sentence and the speaker or the hearer (as in *topic-comment* and *focus* relations). An example of this is the passive transformation, which, operating on the proto-sentence *John strokes the dog*, yields *the dog is stroked by John*. This proto-sentence can in turn be realized as a sentence by means of the morphophonemic rules. But this sentence is not a kernel sentence.

Expansions do not operate on proto-sentences, but on strings in the base. Expansions are obtained by taking one operator as the argument of another operator; this can, in its turn, have other operators as arguments and so forth. This embedding process can continue indefinitely. It is in this way that the hierarchic nesting of predicates comes about. Take the operator *relate to*, for example. It belongs to the category V_{NVV}, and we substitute *I* for *N*, *being hungry* for *V*, and *growing* for *V*. This gives us the basic form *relate to* (*I*, *being hungry*, *growing*). By the global projection (subject raising), this form can be transformed into a proto-sentence without further expansion: *I*+*relate to*+*being hungry*+*growing*, which the morphophonemic rules make into *I relate being hungry to growing*. It is possible, however, further to expand the operators *being hungry* and *growing*, both of which are of the type V_N, for example, as follows: *being hungry (John)* and *growing (John)*. The nested basic form will then be:

relate to (I, being hungry (John), growing (John)).

The global projection is then successively applied to each operator from the inside out. This yields the proto-sentence:

I relate to ((John being hungry) (John growing))

Morphophonemic rules transform the proto-sentence into:

I relate John's being hungry to John's growing.

The proto-sentence can undergo more paraphrastic transformations, and other proto-sentences can be constructed which lead to other morphophonemic realizations, such as, for example, *I relate*

John's being hungry to the fact that he is growing or *I relate John's hunger to his growing*. But there is only one sentence which can be derived from a base structure without paraphrastic transformations. If the base structure has undergone no expansion, that sentence is the kernel sentence; if expansion has taken place, the sentence has a special status in Harris' grammar: it is an element of the REPORT LANGUAGE. The report language consists of the sentences which are generated without paraphrastic transformations. Kernel sentences are the simplest sentences in the report language; all the other sentences in it have one or more expansion in their generation history. The idea of a report language is rather surprising in itself. It means that a natural language contains a sub-language in which precisely the same things can be said as in the language as a whole, as paraphrastic transformations retain meaning. The sentences of this sub-language, abstraction made of morphophonemic variations, consist of nestings of predicates.

Lexical insertion takes place first in the base. But paraphrastic transformations and other morphophonemic rules can replace the lexical elements of the proto-sentence,[1] as we have seen in the examples. Just as in generative semantics, lexical insertion need not precede transformations.

The outline given to this point of the formal framework of the Harris theory is summarized in Figure 4.4. The schema in the figure has been simplified in that diagonal lines have been omitted. The morphophonemic rules, and especially the paraphrastic transformations, can yield the same phonemic form for different proto-sentences; when this occurs, the sentences are ambiguous. Such is the framework of the Harris operator grammar. Two remarks will now be made, on the language and on the report language generated by that grammar.

(i) The grammar generates more than the sentences usually called "grammatical", for there is no lexical or transformational mechan-

[1] Harris describes the paraphrastic transformations as a part of the morphophonemic system. It is not clear to what extent it is more desirable to have the paraphrastic transformations precede other morphophonemic rules than to mix them.

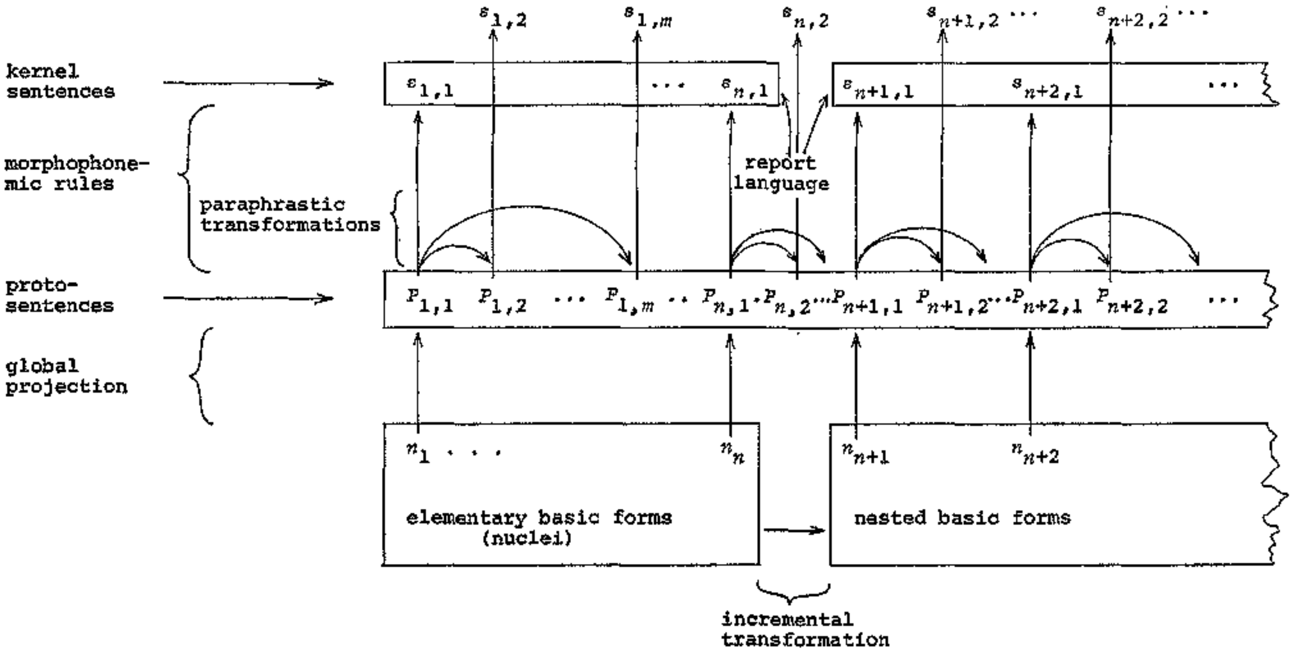


Fig. 4.4. Schema of the Harris Operator Grammar.

ism to block such sentences as *the apple eats the boy* or *I am knowing*. It lacks the selectional features mentioned in Chapter 3, paragraph 1.1. For Harris, however, this is anything but a disadvantage. In his opinion, restrictions of this sort are made by the *universe of discourse*, the state of affairs of the world of which we speak, or at least our knowledge of it. Suppose, for example, that neurophysiologists discovered that "knowing" is a neurological process; for them, then, a sentence like *I am knowing* could be completely acceptable. As for *the apple eats the boy*, there could be huge apples on the other side of the moon, with skin pores so enormous that an astronaut could easily disappear into them. The point Harris wishes to make is that language must be communicative in *every* domain of discourse. Only when the linguist takes a particular domain (the language of weather reports, the language of chemistry, etc.) for further analysis can such selectional features be introduced meaningfully. They can then be integrated into the system of paraphrastic transformations, and can lead to the blocking of some sentence forms.

(ii) The report language is minimal in more than one respect. While all information can be expressed in the report language, it will often be in a very "meager" form. The syntactic structure of the report language is very simple, as we have seen, but the language is also minimal from a lexical point of view, as the number of operators in the base is held to a minimum. Harris wishes the report language to have the smallest possible set of operators which are as general as possible. All other operators are derived inasmuch as they are introduced by means of the paraphrastic transformations for the replacement of a nesting of elementary operators. An example might be the following: adverbs of measure can generally modify verbs as well as adjectives: *he works a little* and *this bike is a little clumsy*. Some adverbs of measure, however, cannot easily be combined with verbs. An example of this is *very* in **the dog limps very*. This is a reason to exclude the adverb *very* from the set of primitive operators. In its place we can have *to a great extent*, which has the same meaning, but not the same distributional limitation as *very*. We can say *the dog limps to a great* *extent*, although this is hardly a very elegant sentence. The word *very* will therefore not appear in the report language; it can be inserted by a paraphrastic transformation if *to a great extent* has an adjective in its range, such as in a basic form like *to a great extent (is large (the house))*, where *is large* is an adjectival operator. After global projection, a paraphrastic transformation yields *the house is very large*. This transformation is not applicable to *to a great extent (limps (the dog))*, where *limps* is a verbal operator. Various restrictions on the occurrence of lexical elements in certain contexts, treated in the *Aspects* theory by means of subcategorization and selectional features, can therefore also be handled in the Harris system, though not in the base. Thus remark (i) should be refined: only in the report language are there no selection restrictions on the combination of lexical elements.

This whole approach very much resembles that of the generative semanticists. The information carried by a sentence can be described by means of a hierarchy of elementary predicates. Parts can be replaced transformationally by "derived" lexical elements. However the information contained in such a derived element is completely contained in the hierarchy of predicates which it replaces. The derivation of *John kills Mary* in Figure 3.9 would fit very well, as far as form is concerned, into the Harris operator grammar. The only difference is that in the figure, the terminal elements other than *John*, *kill*, and *Mary* are abstract semantic primitives, while in the Harris model they would represent ordinary morphemes. It is noteworthy, however, that in generative semantics an appropriate verbal form can always be found for a semantic primitive, and this strengthens the impression that such a limited *report language* does indeed exist. It would be very interesting to see research done on the extent to which transformational lexical insertion, such as in Figure 3.9, is *optional*. If that extent proved to be great, a report language could also be defined on the basis of generative semantics, and all information could be expressed in it with a minimal vocabulary and an extremely limited syntax.

Finally, we shall discuss a few classes of operators in the Harris grammar. *Verbs* are operators, as shown in Table 4.1. Some verbs are composed of more than one word *(relate to)*, but little linguistic analysis is available on this class of operators. One might wonder if Fillmore's verb analyses (cf. Fillmore 1969, et al.) satisfy the distributional restrictions required by Harris. For the present it is not known which verbs the report language should contain.

The progressive form has an elementary operator, *be in the process of*, which is the basis of such sentences as *I am writing*. We have already mentioned *to a great extent* as a measure operator. Adverbs of manner have the form *be of x manner*, in which *x* is *slow, quick*, etc. The subjunctive mood is generated by means of a *demand* operator. Time relations among subsentences are all based on the operator *be before*: the sentence *John comes after Peter* is based on *be before (come (Peter), come (John))*. Similarly, the operator for comparatives is *be more than*. Reference-identity, which is a condition for certain deletion transformations, such as *Clara takes the book and (Clara) goes to school*, are performed by means of a *sameness operator*, which is also important in the derivation of adjectives and relative clauses. The operator *and* is also involved in these derivations. The operators *and* and *or* are the only ones in the Harris system which are not predicates. These operators can be demonstrated together in the following adjective derivation:

Nested basic form: *Same* (N_1, N_2) (*and*$((V_N(N_1), V_N(N_2)))$
Example: *Same* (dog_1, dog_2) (*and*((*limps*(dog_1), *be old* (dog_2)))
After global projection: N_1V_N *and* N_2V_N *and* N_1 *is the same as* N_2
Example: dog_1 *limps and* dog_2 *is old and* dog_1 *is the same as* dog_2

This is report language; a paraphrastic relative transformation yields: *the dog which is old limps*, which, after an adjective transformation, yields: *the old dog limps*.

In the Harris grammar, adjectives are essentially subcategories of verbs. This also holds for plural morphemes, and some conjunctions are also taken as verbs. Thus, *because* is an operator of the form V_{VV}, a predicate with two arguments. Adverbs and negation can likewise be treated as verbs. We can thus characterize operators in the Harris model concisely as consisting of predicative verbs with nouns or verbs as arguments, as well as of *and* and *or* which are operators but not predicates. Finally, *tense* is also an operator;[1] it is perhaps the last or one of the last to be applied to a basic form. Harris is not very explicit about the ordering of operators, however, and in general there are many formal questions in his model which remain unanswered.

The most detailed treatment is still that of the base grammar. The other subsystem, the morphophonemic rules including paraphrastic transformations are given little attention, not only from a linguistic, but also from a formal point of view. The manner in which the domain of a transformation is defined remains an open question, as does the precise form of a morphophonemic rule. The restrictions on alternation of transformations which introduce lexical elements and other transformations are also unknown. Such restrictions decidedly do exist, but they are not formulated by Harris any more than they are in generative semantics (for a similar criticism of the latter, see Fodor 1970). The base of the Harris grammar is beyond doubt a context-free grammar, but little can be said of the generative power of this operator grammar without a more precise definition of the morphophonemic system.

The advantages of operator grammars might therefore be summarized as follows. They are very well suited for detailed representation of functional relations, for the argument on which a given predicate operates is always explicitly stated. Moreover all these predicates are elementary operations, and consequently, in semantic analysis, no further atomization of these predicative relations is necessary. The far reaching distinction between a base with expansions as recursive rules, and a paraphrastic morphophonemic component, is also attractive, at least as an empirical challenge. Can one indeed distinguish a report language with sentences in a simple logical form, on which the whole language is based? The nonterminal vocabulary in an operator grammar is very limited, and the base is extremely simple; various divergent

[1] It is by *tense* that a kernel sentence can be based on more than one operator.

phenomena such as verbs, quantifiers and tense are all treated uniformly.

Many of the disadvantages of operator grammars can be traced to the fact that the transformational component has undergone only a rudimentary elaboration. Both Harman and Seuren give only incidental information on the subject. In the Harris system the morphophonemic rules, including the paraphrastic transformations, are a closed book. We do not know how word order is determined (abstraction made of the "global projection"), nor do we know how words form groups and subgroups, how discontinuous constituents come into being, or what the internal structure is of operators which consist of more than one morph. There is still no formal basis either for the distinction between exocentric and endocentric constructions, nor for a separate treatment of sentence adjunction. However it does seem that in the Harris grammar this latter, as well as the endocentric relation, is characterized by a derivation in which the operator *and* plays a role (as we have seen in the derivation of *old dog*). The Harris grammar, moreover, is very limited in the treatment of selection restrictions, and it has not been shown that lexical insertion can be handled adequately within this system without great formal difficulties.

4.4. ADJUNCTION GRAMMARS

An adjunction grammar might be called a "grammar of modifiers". The idea is that the sentence has a very simple frame which can be made more complicated only by the addition of modifiers or adjuncts. Inversely, one can successively cancel the adjuncts of a given sentence, without losing the status of "sentence". In *crowds from the whole countryside demanding their rights surrounded the palace*, we can first cancel *whole* and *their*, then *from the countryside* and *demanding rights*, leaving finally *crowds surrounded the palace* With every cancellation the string still remained a sentence. One could also say that a sentence contains endocentric constructions, and the modifiers of all those constructions can be cancelled, so that only a grammatical string of heads remains. The adjunct which is cancelled is either a single word (*whole*, *their*) or a string (*from the countryside*, *demanding rights*). The latter are exocentric phrases, as is the remaining "sentence frame". We call such sentence frames CENTER STRINGS (this corresponds roughly to Seuren's *nuclei*); the modifiers are called ADJUNCTS. In natural languages we see that adjuncts are not only added to elements of the center string or its expansions, but also to the center string as a whole (in such a case they are called SENTENCE ADJUNCTS). Is it possible to characterize a language completely, with a finite set of center strings, a finite set of adjuncts, and a system of rules which regulates the way in which adjuncts and center strings are joined?

This thought was developed in very much linguistic detail by Harris (1968), and may be seen as the beginning of his operator grammar, which in fact contains an elaboration in detail of the internal relational structure of center strings and adjuncts. Harris' work since 1959 in *string analysis* (cf. Harris 1970b) provides a good deal of linguistic justification for adjunct grammars, and moreover, there is a detailed formal treatment of them. It was Joshi who developed a formal theory for adjunct grammars, much as Peters and Ritchie formalized the *Aspects* theory. Joshi also expanded Harris' original work on a number of important linguistic points. The Joshi adjunct grammar and the *Aspects* model are the only linguistically interesting mixed models, the transformational components of which have been formalized, and the generative powers of which are known. The Joshi grammar is usually called MIXED ADJUNCT GRAMMAR: *MAG*. One of the best and most extensive computer programs for syntactic analysis (Sager 1967) is based on an adjunction agrammar.

From a linguistic point of view, one of the most important contributions made by Joshi (Joshi et al 1972a, b; Joshi 1972) is the addition of a new category of segments, the REPLACERS, to the existing classification of center strings and adjuncts. Replacers are themselves center strings, but they may also replace an element in another center string. This is a form of sentence embedding. In the sentence *John tells that his bike has been stolen*, *that his bike has been stolen* is the replacer for *S* in *John tells S*.

However, Joshi's center strings, adjuncts and replacers are not strings of words, but rather of grammatical categories. A few examples of center strings are the following: *NtV* (*John will walk*; *t* stands for *tense*), *NtVN* (*John eats apples*), *NtVNPN* (*John gave the book to Charles*; *P* stands for "preposition"). These are ordinary elementary sentence forms, much like those in Table 4.1 and those in Seuren's grammar, with the exception that tense here plays a role in the center string, while this was not the case for Seuren's nuclei.

A MIXED ADJUNCT GRAMMAR *MAG* has a base which contains an adjunct grammar *AG*, and a transformational component *T*. We shall begin with a description of the base.

There are various proposals for the base. The differences reside in the characterization of the three types of segments: center string, adjuncts, and replacers. We shall follow the simplest and most graceful proposal, namely, to have the three sets coincide; adjuncts and replacers are also center strings, and every center string can, in principle, act as an adjunct or as a replacer.

The center strings which figure in a given grammar, and the conditions under which they may be adjuncts or replacers are established in the JUNCTION RULES and the REPLACING RULES, which comprise the categorial component of the base grammar. We define as follows. The categorial component of the base grammar is an ADJUNCT GRAMMAR $AG = (\Sigma, J, R)$. Σ here represents a finite set of center strings; each center string is a finite string of elements over a vocabulary C of categories, of which S is an element. J represents the finite set of JUNCTION RULES, and R represents a finite set of REPLACING RULES.

A JUNCTION RULE indicates (i) which center string is the "host", (ii) which center string is the adjunct, (iii) to which element of the host the adjunct is adjoined, and whether this occurs to the right or to the left of that element. In formal terms, a junction rule is a triad $u = (\sigma_i, \sigma_j, l_k)$, or $u = (\sigma_i, \sigma_j, r_k)$, where $\sigma_i, \sigma_j \in \Sigma$, and σ_i is the host, σ_j is the adjunct, and l_k and r_k indicate respectively that the adjunct is added to the left or to the right of the k^{th} element of σ_i.

Suppose, for example, that $u = (NtVN, NtV, l_4)$. This adjunction will then yield the following compound string: $NtV((NtV)N)$.

If $u' = (NtVN, NtV, r_4)$, then the string $NtV(N(NtV))$ follows, and if $u'' = (NtVN, NtV, r_3)$, we have $Nt(V(NtV))N$. The parentheses indicate the element of the host to which the segment is adjoined.

A junction rule is not only applicable to the center string indicated, but it may also be applied to all strings derived from the center string. The center strings of u and u' are the same, $NtVN$. We first apply u to $\sigma = NtVN$, then derive $\sigma' = NtV((NtV)N)$. We can now apply u' to σ', because it is derived from the correct center string. When u' is applied, NtV must be added to the fourth element of the original center string, thus to N. The result is $\sigma'' = NtV((NtV)N(NtV))$. The fourth element has now received adjuncts both to the left and to the right. If successive adjuncts are added at the same side of the element, by convention they are always inserted directly next to the element. Other adjuncts already present move one place over. A special case of adjunction, sentence adjunction, will be discussed shortly, in the treatment of the replacing rules.

A REPLACING RULE indicates (i) which center string is the host (that center string containing at least one S) (ii) which nucleus is the replacer of S, (iii) which S is replaced, if there is more than one S. In formal terms, a replacing rule is a triad $r = (\sigma_i, \sigma_j, k)$, which means that σ_j replaces the k^{th} S-element. As in practice there is often only one S in the center string, the k $(= 1)$ may be omitted. Thus $r = (\sigma_i, \sigma_j)$ means that σ_j replaces the only S in σ_i.

If, for example, $r = (NtVSPN, NtV)$, this will yield the string $NtV[NtV]PN$. The brackets indicate that the segment is inserted by replacement. As was the case with the junction rules, it holds that the replacing rule is applicable not only to the indicated center string ($NtVSPN$ in the example), but also to all its derivations, regardless of whether these are the result of adjunction or replacement. A condition for the application of this is, of course, that the S with the correct number k in the original center string be maintained. No further replacing rule may therefore be applied to the string $NtV[NtV]PN$, because no further S from the original center string is available. However this string may still undergo further adjunctions.

Until now, we have only spoken of adjuncts as additions to an element in the center string. But there are very good linguistic arguments for the admission of another type of adjunct, the SENTENCE ADJUNCTION. Compare, for example, the sentences *John works hard* and *John works sometimes. Hard* refers here to *work*, but *sometimes* refers to the whole sentence *John works*. Therefore the sentence *sometimes John works* is more acceptable than *hard John works*. The first sentence, moreover, can be paraphrased as *John is a hard worker*, but the second cannot be paraphrased as **John is a sometimes worker*. Thus *sometimes* is a sentence adjunction. The adjunct grammar contains a special rule for sentence adjunction according to which an adjunct is assigned not to an element, but to a whole string. The rule has the form $(\sigma_i, \sigma_j, l_s)$ or $(\sigma_i, \sigma_j, r_s)$, which means that the adjunct σ_j is added to the center string σ_i as a whole, either on the left or on the right. One might wonder why a sentence adjunction cannot be accomplished by means of a standard junction rule, in which an adjunct is assigned to an element S in the center string. The answer is simply that it would be impossible in that way to add sentence adjunctions to the original center strings in Σ, for these are not the result of rewrites of S, as was the case in phrase structure grammars. Given the sentence adjunction rule, moreover, it is not necessary to make adjunction to an S element possible by means of the standard adjunction rule. This is excluded by the following convention: the elements S are not numbered with the other elements of a center string; they are numbered separately for the use in replacing rules only.

In an adjunct grammar, it holds for replacers and adjuncts in the same way as for hosts that they need not be the center string indicated in the rule, but only that they be derived from it. If, for example, NtV can be a replacer for the S in $NtVSPN$, then every derivation of NtV can also play the role of replacer. Suppose, for example, that there is a junction rule $(NtV, NtVN, r_1)$; in that case not only NtV, but also $(N(NtVN))tV$ can replace the specified S, and this would yield the string $NtV[(N(NtVN))tV]PN$.

Thus the two types of rules form a recursive system over the set of center strings Σ. A string is said to be *derived from a center string* in Σ, if it is obtained from that center string by the application of zero or more adjunctions and/or replacements. Although this is not essential, for the rest of the discussion a derivation can best be imagined as going "from the bottom up" and not "from the top down", as was the case in phrase structure grammars. In other words, a string which is inserted as an adjunct or a replacer may not be altered in further derivations. The very last phrase of the derivation is therefore the insertion of all the "prefabricated" strings in the "matrix center string".

The LANGUAGE generated by an adjunct grammar is the set of strings in $(C-S)^*$, i.e. strings of categories in which S does not occur, and which can be derived from the center strings in Σ by means of R and J. This language is *preterminal* because the strings contain no lexical elements. Obviously the center strings in which S does not occur are also sentences of this preterminal language.

A system of lexical insertion rules is appended to this categorial adjunct grammar, much as was the case in the *Aspects* theory. These rules replace the preterminal category symbols with terminal lexical elements. A lexical element is "attached" to the category elements of the center string at the moment the latter is inserted into another string. The last operation of insertion applies, of course to the "matrix" center string itself. The great advantage of this method is that it makes it possible in a simple way to indicate the selection restrictions which concern (i) the relations within a center string or nucleus (quite in agreement with Seuren's approach in that regard), (ii) the relations among elements in the host or the adjuncts already present in it, and elements of the inserted center string. The form of the lexical rules in the Joshi grammar, however, has not been further elaborated, and we shall leave it out of the following discussion. We will now round up the discussion of the base grammar by another example.

EXAMPLE 4.4. Let $AG = (\Sigma, J, R)$, with

$\Sigma = \{\sigma_1 = NtVN, \sigma_2 = NtVS\}$
$J = \{u_1 = (NtVN, NtVN, r_4)\}$
$R = \{r_1 = (NtVS, NtVN)\}$

We use u_1 first. This may be done only if we fill in the lexical elements for the two center strings $NtVN$ at the same time. Let us suppose that the lexical rules allow that u_1 concerns the following terminal strings:

$$u_1 = (\begin{array}{cccccccc} N & t & V & N, & N & t & V & N \\ | & | & | & | & | & | & | & | \\ \textit{John} & \textit{inf} & \textit{read} & \textit{article} & \textit{Charles} & \textit{pt} & \textit{write} & \textit{article} \end{array}, r_4),$$

where *inf* stands for "infinitive" and *pt* for "past tense". This yields:

$$\begin{array}{cccccccc} N & t & V & (N & (N & t & V & N)) \\ | & | & | & | & | & | & | & | \\ \textit{John} & \textit{inf} & \textit{read} & (\textit{article} & (\textit{Charles} & \textit{pt} & \textit{write} & \textit{article})) \end{array}$$

To this we apply replacing rule r_1, i.e. we use this result, which is a derivation of $NtVN$, for the replacement of the S in $NtVS$. To do so, however, we must first insert the lexical elements into $NtVS$. Suppose that the lexical rules allow the following insertion:

$$\begin{array}{ccc} N & t & V\ S. \\ | & | & | \\ \textit{John} & \textit{pt} & \textit{try} \end{array}$$

The replacing rule r_1 will then yield:

$$\begin{array}{ccccccccccc} N & t & V & [N & t & V & (N & (N & t & V & N))] \\ | & | & | & | & | & | & | & | & | & | & | \\ \textit{John} & \textit{pt} & \textit{try} & [\textit{John} & \textit{inf} & \textit{read} & (\textit{article} & (\textit{Charles} & \textit{pt} & \textit{write} & \textit{article}))] \end{array}$$

This string, which no longer contains an S, comprises the deep structure of the sentence *John tried to read the article that Charles had written.* It is obvious that a complete transformational procedure will be necessary to derive this sentence. But before treating the transformational component, we shall first make a remark on the relation between adjunct grammars and phrase structure grammars.

In the form described here, adjunct grammars are equivalent to a subset of context-free grammars. This is easy to see if we consider adjunct grammars with replacing rules only, and no junction rules. Replacing rules replace an S in a center string with another center string which in turn can contain an S. The corresponding context-free rule is $S \rightarrow \sigma$, where σ stands for a string in Σ (such as *NtV*, *NtVSPN*, *StV*, etc.). If for every σ in Σ a rule $S' \rightarrow \sigma$ is added, with S' as the start symbol of a context-free grammar, and $V_N = (S, S'\}$, then that context-free grammar is equivalent to the adjunct grammar. The context-free grammar will then generate phrase markers as in Figure 4.5. It is obvious that the

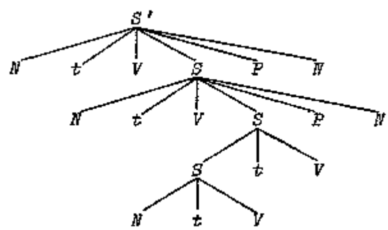


Fig. 4.5. Phrase structure representation of an adjunct grammar structure.

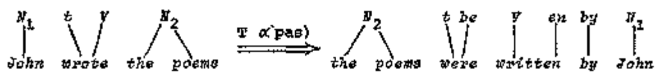
recursive element S is always *directly* derived from S in such a grammar. This is called a context-free grammar with *depth* 1. Such grammars are a strict subset of the set of context-free grammars. Not all context-free languages can be generated by such context-free grammars. But Joshi (to be published) shows that a very minor change in context-free grammars of depth 1 is sufficient to allow them to generate the entire set of context-free languages. It is more important here, however, to know if there is linguistic need of recursive hierarchies of greater depth. This remains an open question, but it is indeed noteworthy that the phrase markers to be found in generative semantic literature generally show no greater depth than 1 (i.e. the recursive element is always introduced by a direct rewrite of itself), or they can easily be reduced to that depth. In Figure 3.9a, for example, we are actually dealing with depth 1, if we replace a pair of rules such as $S \rightarrow Pred + NP$, $NP \rightarrow S$ with the pair $S \rightarrow Pred + NP$, $S \rightarrow Pred + S$. The linguistic need for more hierarchic structure than in sentence embedding is perhaps not very great.

The relations between context-free grammars and adjunct grammars do not change when we also take junction rules into consideration. We can give a junction rule the form of a context-free rewrite rule by the introduction of *dummy* category symbols which will be found in the correct place in the host. For example, the junction rule $u = (NtVN, NtV, I_4)$ can be simulated by the pair $S \rightarrow N + t + V + (S') + N$, $S \rightarrow N + t + V$. The dummy category elements (S' or S'' etc.) are always directly dominated in the tree-diagram by another dummy element or by S. It is indeed the case that the adjunct grammar in this respect is much more elegant than the context-free grammar. Not only does an adjunct grammar clearly show the element to which an adjunct is added, which is not the case for a phrase structure grammar, but it also does not cause senseless multiplication of hierarchic relations when more than one adjunct is added to a single element. This was already mentioned in Chapter 2, paragraph 3.3, and in Chapter 4, paragraph 1, where it was shown that coordination without the use of rule schemas leads to a spurious hierarchic structure. A junction rule can be applied repeatedly without complicating the hierarchy.

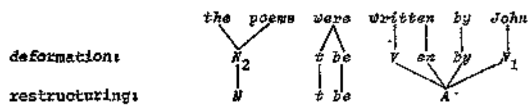
There are also various formulations of the *transformational component* of a mixed adjunct grammar. In its simplest form, this component contains two sorts of transformations, α transformations and β transformations.

The α transformations are applicable only to elements of Σ (center strings) or strings derived from center strings by means of α transformations. These transformations can delete, adjoin, or substitute elements. Substitution and adjunction can be performed only with elements which are already present and with elements of a finite set given beforehand. The result of a transformation of a string is called the DEFORMATION of that string. The adjunct grammar corresponds closely to the *Aspects* model as far as α transformations are concerned, but it is easier to determine the domain of a transformation with an adjunct grammar. In dealing with the *Aspects* model, Peters and Ritchie needed a very complicated formulation to establish that a labelled bracketing had a proper analysis for a given transformation. It is a much simpler matter with Joshi's α transformations. An α transformation is defined for a given center string. It indicates the deformation of the center string and its RESTRUCTURING. This latter means simply that by convention the deformation should be considered in the following as some specified center string. An example should illustrate this.

EXAMPLE 4.5. The passive α transformation replaces the center string N_1tVN_2 with the string N_2t *be* V *en by* N_1. Thus, for example



The result of this transformation is not a center string, but it is established in the definition of passive transformation that the resulting string may be considered to be the center string N t *be* A, that is, the preterminal string for sentences such as *the poems were ugly*. The restructuring is as follows:



The restructuring also shows the further transformations which may be performed on the string; in this case, it is any transformation which is applicable to the center string N t *be* A. An example of this is the relative clause transformation, which changes *the poems were ugly* to *the poems which were ugly*. In precisely the same way, this transformation changes *the poems were written by John* to *the poems which were written by John*.[1]

Restructuring is not necessarily limited to one center string, and also only a *part* of the deformation may be brought under a new center string. Thus, another restructuring for the deformation

[1] The restructing resulting from the passive transformation demands more detail than we give here: we can derive *the ugly poems*, but not **the written by John poems*. The linguistic elaboration of the transformations in a mixed adjunct grammar is still at its beginnings.

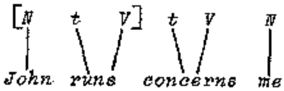
of the passive transformation is the following:

Here *N t be A* is a center string, and *X* is that which remains of the deformation after restructuring. Further transformations, then, only concern the center string. There are certain restrictions on restructuring. Thus the element *t* can never be replaced by *N* or *V*, but only by *t*. The element *V* can only be restructured as *A*, if it is followed by *en* (as in the example) or *ing* (as in *John is writing*) in the deformation. But there is little known about such empirical limitations on restructuring. Attempts have been made to describe the possible restructurings of a deformation by means of a categorial grammar (Hirschman, 1971), in the same way as the *domain* of a transformation is treated in Chapter 4, paragraph 2.

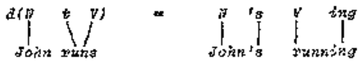
To summarize, we can say that α transformations consist of a center string, a deformation, and one or more restructurings of the deformation. The domain of an α transformation is the center string, or a deformation the restructuring of which is that center string.

The function of β transformations is the transposition of center strings, deformed or not, in the preterminal string. Transposition of a segment means that the point of adjunction of the segment is changed. The points are established in the *R* and *J* rules of the base grammar. A β transformation can therefore be considered as a rule which alters the base rule. It may be said that a β transformation replaces a base rule with a *pseudo rule*. When this occurs, the host is never changed, but when the point of adjunction is changed, the adjunct or replacer may be deformed.

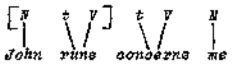
EXAMPLE 4.6. We derive the sentence *John's running concerns me.* The sentence is based on the center strings *NtV (John runs)* and *StVN (S concerns me)*. The insertion of the former into the latter, however, requires the nominalization of *John runs* as *John's running*. We are dealing here with the following replacing rule: $r = (StVN, NtV)$. This yields the string



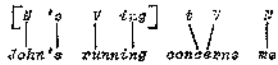
The β transformation now replaces r with $r' = (StVN,\ d(NtV))$, where



The change from r to r' thus changes



to



In this example the point of adjunction is not changed, but the replacer is deformed. When junction rules are applied, the point of adjunction often changes also, for example in *the poem which was ugly* → *the ugly poem*. There are also cases in which an adjunct is divided into segments with different points of adjunction. Take the center string *the proof was concise*, and suppose that we wish to add the adjunct *John proved the theorem* to *the proof*. We then need a β transformation which changes *the proof (John proved the theorem) was concise* to *(John's) proof (of the theorem) was concise*, where the adjunct is divided into two segments with different points of adjunction.

For further details on the transformational component, we refer the reader to the original publications. We might point out that a remark on the ordering of transformations may be found there: the α transformations work cyclically "from bottom to top" just as in *Aspects*, while the β transformations have no extrinsic order.

We wish to close this short discussion of mixed adjunct grammars with a few remarks on a general condition for transformations which Joshi calls the TRACE CONDITION. This condition resembles Chomsky's principle of recoverability of deletions (cf. Chapter 3, page 74, Definition 3.16), but differs from the latter from both linguistic and formal points of view. The trace condition defines a characteristic trace for every transformation. It thus holds that

that trace cannot be deleted in further transformations. What the trace of a transformation actually is is an empirical question which must be answered separately for each transformation (it is, by the way, also an empirical question whether every transformation does indeed have a trace, but even if this should not hold for some transformations, the mathematical results remain valid). For the sake of example, let us see what the trace of an English passive transformation is. To do this, we begin with the center string *John wrote the poem*, and we use the passive transformation for the derivation *the poem was written by John*. We then use the string thus obtained as a relative clause in the sentence *the poem was ugly*. A β transformation yields the string *the poem (which was written by John) was ugly*. The part between parentheses is that which remains of the passive deformation; it can still further be reduced. The *which deletion* transformation makes this into *the poem (written by John) was ugly*. But this *written by John* is still not a minimal trace of the original passive sentence. Suppose, for example, that we coordinate the sentence obtained with *the poem recited by John was ugly*. Transformationally this yields: *the poem (written by John) was ugly and the poem recited by John was ugly*. The "conjunction reduction" transformation gives *the poem (written) and recited by John was ugly*. We see now that the only thing which remains of the original passive deformation is the word *written*, or, in preterminal terms, $V + en$. If it is impossible to remove this transformationally (and that seems to be the case), we may say that the trace of the passive transformation is $V + en$ At first sight the trace condition seems, like the principle of recoverability, to be bound to the separate transformations. This, however, is not the case. The trace condition regards complete transformational derivations: the trace of a transformation may not be eliminated in the entire further derivation. If each transformation leaves at least one morpheme behind, it is obvious that there is an upper limit to the length of the deep structure of a sentence of a given length, for no more transformations can have taken place than the number of morphemes in the surface structure allows. In the following chapter we shall see that this, as opposed to the principle of recoverability, is the guarantee of recursiveness of the transformational grammar.

What, in summary, are the most important advantages and disadvantages of adjunct grammars? Their most noticeable advantage is the explicit distinction of head, adjunct and sentence complement, in contrast with phrase structure grammars in which the distinction between head and adjunct has no natural representation, and no distinction is made between the adjunction (of modifiers) and the replacement (in the form of sentence complements). The constructions which result from adjunction are all endocentric, and all others are exocentric. The mixed adjunct grammar is the only grammar in which this distinction is completely accounted for. The mixed adjunct grammar also offers a strikingly simple solution for the unrestricted coordination of adjuncts; this does not lead to false hierarchy, as is the case with phrase structure grammars. At the same time the amount of hierarchy in these grammars is kept to a minimum, as is the nonterminal vocabulary. This is attractive for theories on the native speaker. In particular, the idea of a small set of center strings or minimal sentence frames which are joined in series in speech is a challenge which psychologists have not yet answered. The formal properties of these grammars are known rather precisely, and, especially in Harris' and in Sager's work, there is a good deal of detailed linguistic "filling".

A mixed adjunct grammar works with rather large units, the center strings. The relations within the center string, consequently, receive very little attention. Functional relations among the elements of the center string, such as dependencies and case relations, can indeed be defined *ad hoc*, but they fit less naturally into the total formal system; this is precisely what Harris sought to work out in his operator grammar. Very little linguistic elaboration has, as yet, been accomplished on the transformational component, and less still on the morphological rules.

4.5. DEPENDENCY GRAMMARS

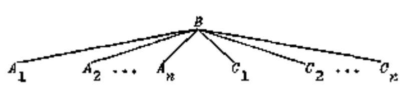
A phrase structure grammar is not very well suited for describing dependency relations among the elements of a sentence. This becomes very obvious in the treatment of endocentric constructions: a tree-diagram can distinguish neither head nor modifier. Categorial grammars are somewhat more successful in this, as we have seen; in them the head has the same category as the entire constituent. Adjunct grammars were developed especially for the description of head/modifier relations. Endocentric constructions, however, are not the only ones in which linguistic dependency among elements occurs; it can also very well appear in exocentric constructions. In the nuclei of the Seuren model, for example, nominal elements are dependent on the main verb: according to Seuren, it is the main verb which determines the selection restriction for the nominal elements in the nucleus, and not the inverse. Another example is the prepositional phrase, where the noun phrase is dependent on the preposition. Dependency is actually a distributive notion: the syntactic surroundings in which a word group can occur as a whole are determined principally by the independent element, the head of the word group, and the other words contribute very little in this regard. This holds for both endocentric and exocentric constructions. We have seen that the endocentric phrase *old chairs* can occur nearly everywhere *chairs* can occur alone. The word *old* scarcely limits that distribution at all. Correspondingly, the syntactic surroundings in which a prepositional phrase (such as *over the house*) can occur are much more limited by the preposition *(over)* than by the noun phrase *(the house)*; the preposition is the head of the construction.

Categorial grammars are suited for expressing only one type of dependency, either the endocentric or the exocentric type. Operator grammars offer a good representation of exocentric dependency, especially the dependency between the main verb and noun groups in the sentence. But endocentric dependencies are represented only indirectly; they go back transformationally to exocentric constructions in the base. Adjunct grammars, finally, give no detailed analysis of the relations within the center string or nucleus, and in this sense they fail to deal with exocentric dependencies.

DEPENDENCY GRAMMARS were developed especially to express such syntactic dependencies. Like all the other grammars in this chapter, they have the advantage of a very limited nonterminal vocabulary; it consists here only of preterminal syntactic categories, each of which can be replaced only by terminal elements.

A DEPENDENCY GRAMMAR $DG = (V_N, V_T, D, L, T)$ is characterized by a finite NONTERMINAL VOCABULARY V_N, a finite TERMINAL VOCABULARY V_T, a finite set of DEPENDENCY RULES D, a finite set of LEXICAL RULES L, and a set of START SYMBOLS. In the following discussion we shall suppose, for reasons which will be indicated later, that the set of start symbols contains only one element, T.

The DEPENDENCY RULES D indicate for each category in V_N which categories are dependent on it and in which relative position. The rule $B(A_1\ A_2\ \ldots\ A_n\ *\ C_1\ C_2\ \ldots\ C_m)$ means that $A_1, \ldots, A_n, C_1, \ldots, C_m$ are dependent on B in the indicated sequence, with B in the place of *. This can be represented graphically by placing B above the string and connected with the dependent elements as follows



The number of dependent elements in a dependency rule is equal to or greater than zero. If it is equal to zero, the rule is as follows: $B(*)$, which means that the element B can occur without dependent elements.

The LEXICAL RULES L are simple rewrite rules of the form $A \rightarrow a$, in which $A \in V_N$ and $a \in V_T$. Although one might expect that an adequate dependency grammar would need a more complicated form of lexical insertion, as was the case in *Aspects*, little is known on the subject. We shall return to this subject, but for the moment we retain this simple rewrite form for the lexical rules.

START SYMBOLS are categories which need not be dependent on another category; they can start a derivation. We do not use the symbol S for this, however, because all the elements in V_N are preterminal, and we prefer not to have lexical rules of the form $S \to a$. We suppose that there is only one start symbol, T, with the intuitive meaning of *sentence type* (interrogative, imperative, etc.).

EXAMPLE 4.7. $DG = (V_N, V_T, D, L, T)$, with $V_N = \{D, N, V, T, P\}$, $V_T = \{a, ass, boy, child, gave, ice\ cream, the, to\}$, with the dependency rules D and the lexical rules L as follows:

1. $T(*V)$	5. $T \to ass$
2. $V(N*NP)$	6. $V \to gave$
3. $P(*N)$	7. $N \to \{boy, child, ice\text{-}cream\}$
4. $N(D*)$	8. $P \to to$
	9. $D \to \{the, a\}$

With this grammar we derive the sentence *the boy gave the ice-cream to a child*. The start symbol occurs only in rule 1; this gives the string TV. By rule 2 the dependents of V are inserted, yielding $TNVNP$, and by rule 3 the dependents of P are introduced, yielding, $TNVNPN$. By applying rule 4 three times, we get $TDNVDNPDN$, the preterminal string from which the sentence desired can be derived by means of lexical rules 5 to 9. This derivation can be represented in a DEPENDENCY DIAGRAM, as in Figure 4.6.

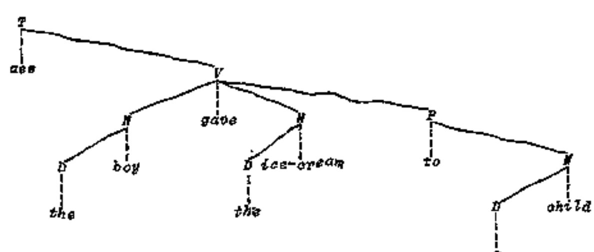


Fig. 4.6. Dependency diagram for the sentence *the boy gave the ice cream to a child.*

In such a diagram we can see the dependency relations from the top of the diagram to the bottom; the category elements in those relations are ordered horizontally according to their position in the preterminal string. The lexical elements are added, and connected to the diagram by dotted lines. The terminal element *ass* stands for *assertion.*

The DIRECT DEPENDENTS of an element are the elements which are mentioned in the dependency rule. In the example, V is directly dependent on T, and P is directly dependent on V. The INDIRECT DEPENDENTS of an element are the elements which are derived from that element in more than one step. In Figure 4.6 P is indirectly dependent on T. A CONSTITUENT is an element with all its direct and indirect dependents. In the figure, *a, a child, to a child, the boy gave the ice-cream to a child,* etc. are constituents. The HEAD of the constituent is the element which is independent of the other elements in the constituent. Thus *gave* is the head of *the boy gave the ice-cream to a child,* and *to* is the head of *to a child.*

The generative power of a dependency grammar resides in recursive rules, which insert the start symbol T, as, for example, in the rule $N(T^*)$. Gaifman (1965) has proven that dependency grammars are (weakly) equivalent to context-free grammars. The proof, which we shall not treat here, is indirect; it shows the equivalence of dependency grammars and categorial grammars which in turn are equivalent to context-free grammars. It is not difficult to construct an equivalent context-free grammar for a given dependency grammar (the inverse is much more complicated). A context-free grammar equivalent to the dependency grammar in Example 4.7 has the following production rules:

$S \rightarrow T + V'$
$V' \rightarrow N' + V + N' + P'$
$P' \rightarrow P + N'$
$N' \rightarrow D + N$
and rules 5 to 9.

The context-free phrase marker which corresponds to the dependency diagram in Figure 4.6 is given in Figure 4.7. The construction procedure is based on the insertion of an extra nonterminal symbol for each category which can have dependents itself (S for T, N' for N, and P' for P).

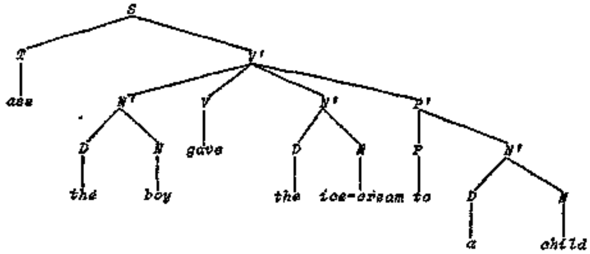


Fig. 4.7. Context-free phrase marker for *the boy gave the ice cream to a child.*

A comparison of Figures 4.6 and 4.7 shows that the former, in contrast to the latter, clearly represents the dependents without excess of hierarchic relations or nonterminal elements. The various relations within the nucleus, the dependency of the various word groups on the main verb or the "direction" of the selection restrictions, are particularly well represented. But on the other hand, the distinction between exocentric and endocentric is lost. The dependency diagram also does not allow one to deduce the type of a constituent. Robinson (1970) gives the following rule: an element with more than one direct dependent is the head of an exocentric construction. Thus in the example, *V* is the head of an exocentric construction. This condition, however, is sufficient but not necessary. The preposition *to* has only one direct dependent, but *to a child* is nevertheless exocentric. The intuitive interpretation of a dependency diagram is rather one of *selection*: the head determines the choice of the dependent elements. Such diagrams, like chemical structures, show the valence of each element (the number of direct dependents which an element can have) and the connected chains in which elements are ordered.

A dependency diagram is perhaps a fitting means for expressing case relations in the base. Caution is necessary in this respect, however, for, despite the work done by Fillmore and others (cf., for example, Fillmore, 1968), research on the formal properties of case relations is still in a very early stage of development. For a thorough linguistic analysis on the subject, in which the formal system of dependency grammars is used, we refer the reader to Anderson (1971). Without going into much linguistic detail, we shall at this point outline the formal means for the dependency representation of case, following the general lines of the work done by Robinson (1969, 1970).

It is possible to introduce a syntactic category for each case. Such syntactic categories may be rewritten as case morphemes. In English this will generally take the form of a preposition, but it can also take that of a suffix. Let us introduce the following nonterminal symbols, by way of example, without taking position as to the linguistic relevance of the case categories used: *A* for *agent*, where *A* can be rewritten as *by*; *I* for *instrument*, where *I* can be rewritten as *with* (or *by*; we shall return to this later); *Dt* for *dative*, where *Dt* can be rewritten as *to*; *L* for *locative*, where *L* can be rewritten as *in*, *at*; *O* for *objective*, where *O* can be rewritten as *of*. These case categories are introduced into the base as direct dependents of the verb, for example, by a rule such as $V(A*O\ Dt)$. Each of these categories can then be given an *N* as dependent, by rules such as $A(*N)$. Thus the underlying structure of a sentence such as *the boy gave the ice-cream to a child* is something like the diagram in Figure 4.8. The lexical elements are inserted for the sake of clarity.

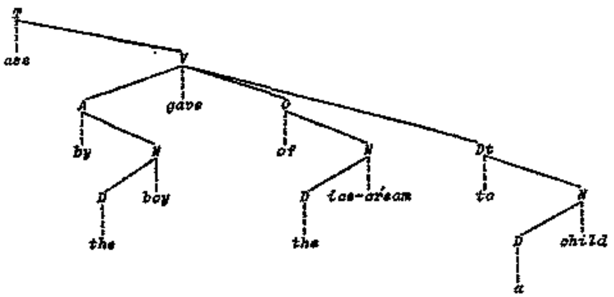


Fig. 4.8. Dependency diagram for *the boy gave the ice cream to a child* with case relations.

Lexical insertion rules, as we have already mentioned, are not simple rewrite rules, and we must now examine this matter more closely. The verb *give* has a case specification in the lexicon, namely, [*A*—*O* *Dt*]. The insertion of the verb is possible only if its syntactic case specification in the lexicon agrees with the structure of the dependency diagram (this is completely analogous to lexical insertion in the *Aspects* model). How does the insertion of a noun take place? It is pointless to give nouns case specifications in the lexicon. Case features are excluded because the majority of nouns can occur in a variety of different case roles. According to Robinson, it is the presence of a "case-related" feature which determines whether or not a given word may fill a given case function. Thus, for the dative and for the agent, it is probably possible only to choose words which are [+ animate]. By a "syntactic redundancy rule", the feature [+ animate] is added to the *N* which is directly dependent on the *A* or *Dt* in the dependency diagram. The two redundancy rules in question are: *N* →[+ animate]/ *A*[*—] and *N* → [+ animate]/ *Dt*[*—]. As usual, the surroundings in which the feature is added to *N* are specified to the right of the diagonal "/". Lexical insertion of a given noun may take place only if, according to the lexicon, the noun possesses the feature required. Thus in Figure 4.8 the feature [+animate] is added to the *N* which is directly dependent on *A*, by means of the syntactic redundancy rule. Lexical insertion of *boy* is allowed because in the lexicon, *boy* is specified as [+animate]. Inanimates such as *stone* and *comparison* are therefore excluded as agents. Similar conditions hold for the dative.

As for prepositions, we suppose that they are specified according to case: *by* has the feature [+*A*], *with* has [+*I*], etc. Their insertion is determined by these case features. But there are also other conditions for the insertion of prepositions. Thus *by* can also be used for *I*, provided that no *A* is specified; compare, for example, *the window was broken by the ladder* and *John broke the window with the ladder*. Prepositions, likewise, are often not realized in lexical insertion. The *by* of the *agent* appears only in passive constructions, and the *to* of the dative is dependent on position; compare *John gave the book to Peter* and *John gave Peter the book*. The objective preposition is realized even less often in lexical insertion. In this connection we naturally think of transformational mechanisms, but very little is known of their function in the present question.

We close this paragraph with a few remarks on the transformational component of a dependency grammar. Transformations replace dependency diagrams with dependency diagrams. They may be written not only in diagram form, but also in the labelled bracketing notation. For more detail on this, we refer the reader to Robinson (1970). Transformations must be able to delete, adjoin and substitute elements. The deletion of an element presents no problems, if that element has no dependents. The following convention can be introduced for the case in which the element to be deleted does have dependents. If *C* depends directly on *B*, and *B* on *A*, when *B* is deleted, *C* depends directly on *A*. Adjunction makes the element added (possibly together with the constituent dependent on it) dependent on a new head. Substitutions, however, raise all sorts of formal and empirical problems. The matter is still simple when substitution consists only in the interchanging of the dependents of an element, as in the exchange of positions of noun phrases in a passive transformation, for all the elements remain dependent on *V*. It remains an open empirical question, however, if exchanges of roles of head and dependent can take place when the elements are exchanged. In other words, is it possible in the surface structure to reverse such a semantically significant relation? Robinson takes as a working hypothesis that this is not possible.

Like the other grammars in this chapter, dependency grammars have the advantage of a limited nonterminal vocabulary. This, as we have seen, consists entirely of preterminal elements. This offers certain advantages in the transformational component, all the more striking when compared with transformations in the *Aspects* model. In the latter, the output often contains various superfluous category symbols, and *ad hoc* conventions are needed to eliminate them. One such convention, as we have seen, is the reduction convention (cf. Chapter 3, paragraph 2.4): if, after a transformation, such a labelled bracketing as $(_A(_B(_A\alpha)_A)_B)_A$ occurs, the inner pair of brackets $(_A,)_A$ are "automatically" removed. In diagram form,

Another convention is *tree pruning*: remove every embedded S from which no more than one branch leads. Thus the path — A — S — B — is simplified to — A — B —. But also other category symbols often fill no role whatsoever after transformation. Every transformational treatment of the adjective, for example, meets this problem. Robinson (1970) gives the following example of this (after Ross (1967)): Two stages in the transformational derivation of the adjective are given in Figure 4.9.

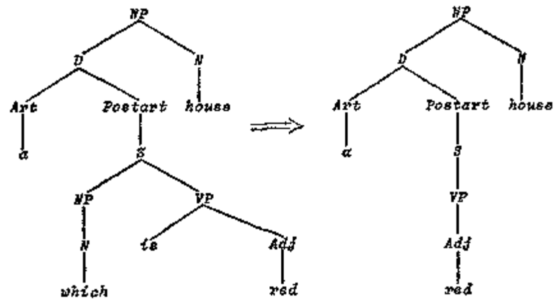


Fig. 4.9. Two stages in the transformational derivation of an adjectival construction (*Aspects* model).

Tree pruning does eliminate S from the path of categories which, after transformation, is dominated by *red*, but *Postart* and *VP* remain and have very little intuitive significance. Suitable solutions may, of course, be found for this, but this example shows that the use of an abstract nonterminal vocabulary demands transformational means with only formal and no intuitive significance. Reduction and tree pruning are pure artifacts of the rule system used, the phrase structure grammar. Robinson (1970) shows how the same two stages look in the dependency system. They are given in Figure 4.10.

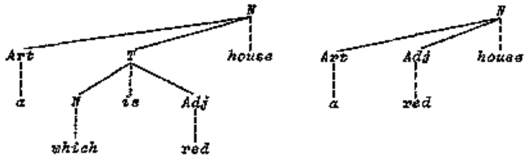


Fig. 4.10. Two stages in the transformational derivation of an adjectival construction (dependency system).

But the intuitive attractiveness of the transformational component can only be judged when we dispose of (a) a complete formalization of the dependency transformations, like the formalizations which already exist for the *Aspects* model and for adjunct transformations, and (b) a detailed analysis of a number of "representative" linguistic cases. At present neither is available. Concerning the formalization, we should point out that for dependency grammars, nothing has yet been done to define a principle of recoverability or a trace condition for the transformations.

To summarize, the advantages of a dependency grammar include the natural representation of distributional dependencies, the limited nonterminal vocabulary, the facility in formulating case relations, and the simple way of accounting for word order and constituent structure. Disadvantages include the limited possibility of distinguishing endocentric constructions from exocentric constructions, and, until now, the uncertainty on the structure of the transformational component and its generative power.

4.6. FINAL REMARKS

The grammars which have been discussed in this chapter differ in many respects. Other mixed models, moreover, are being published regularly (see, for example, Hudson (1971) and Huddleston (1971) for a formalization of Halliday's *systematic grammar*). The question as to which model is correct is pointless and without answer, for every kind of transformational grammar has its pro and con. The linguist and the language psychologist who seek a model will be guided in their choice by the nature of the phenomenon they wish to study, for some phenomena are naturally representable by one form of grammar, whereas others require a different formalism. All investigators, however, would be served by more detailed data on the weak and strong equivalence of the various transformational grammars. It is usually only for historical reasons that schools of linguistics tend to bind themselves to a particular formal system. If it were shown that different systems were equivalent to a considerable degree, there might be a chance to break through the isolation which is so characteristic of the formation of schools. Where differences are only notational conventions, mathematical linguistics could play an important boundary spanning role by showing how one system might be translated into the other. Where differences concern really substantial questions, only reflection on the possibility of notational translation will allow a judgment on the greater or lesser descriptive adequacy of one or another grammar, or show that the problems in question can be solved more or less independently of the formal system used. Unfortunately at the present stage knowledge about the formal equivalence of the various grammars is still very limited, especially as far as the transformational components are concerned. The practicing linguist has no choice but to acquire some skill in the use of the most important systems, lest he should no longer see the substantial forest because of the formal trees. A decidedly important reason for the use of a transformational mechanism of the *Aspects* type is the simple fact that so many modern linguistic studies are based on the *Aspects* model, and scientific communication is thereby facilitated. But this reason is not sufficient, for that system retains its weak points, and, on the other hand, many important linguistic discoveries have been formulated in other systems.

5

THE GENERATIVE POWER OF TRANSFORMATIONAL GRAMMARS

The conclusion of Chapter 2 stated that the step toward type-0 grammars for the description of natural languages should not be taken lightly. In Chapter 2, paragraph 5 it was argued that for linguistic purposes only grammars should be considered which generate *recursive* sets. In the present chapter we shall discuss the extent to which the *Aspects* model satisfies this condition. We shall also make comparisons with the mixed adjunct grammar, the only other transformational grammar of which the formal structure is known in detail. It will further be shown that the *Aspects* model does indeed generate a type-0 language; the discussion of this in Chapter 3, paragraph 2.3 was not carried out completely, when it appeared that transformations are rule *schemas*.

In the present chapter we shall first show that the *Aspects* theory gives no guarantee of decidability, and moreover, that a transformational grammar of that form, or even of a simpler form, can generate all type-0 languages, that is, all recursively enumerable sets (paragraphs 1 and 2). In paragraph 3 we shall show that this conclusion has serious consequences for linguistics. In paragraph 4, finally, we shall discuss the direction in which solutions to the problem may be sought.

5.1. THE GENERATIVE POWER OF TRANSFORMATIONAL GRAMMARS WITH A CONTEXT-SENSITIVE BASE

Peters and Ritchie (1973) give a strongly restricted definition of "transformation". The *Aspects* model would certainly tolerate wider definitions. Nevertheless these authors were able to prove that transformational grammars of that form can generate all recursively enumerable sets. In this paragraph we shall follow in some detail the proof that transformational grammars of the *Aspects* type with context-sensitive base grammars can generate all type-0 languages. The proof uses only the elementary deletion transformation, the cyclic character of the transformational component, and the principle of recoverability. Although the base of the *Aspects* theory contains context-sensitive rules (and transformations), we have seen that it is equivalent to a context-free grammar. In the following paragraph we shall follow — in somewhat less detail — the argumentation advanced by Peters and Ritchie that in that case, and even when the base is much more elementary, the generative overcapacity remains. For the latter it will be necessary to use the filtering function of transformations. That property, however, will play no role in the present paragraph.

THEOREM 5.1. Every type-0 language can be generated by a transformational grammar with a context-sensitive base and *Aspects* type transformations.

PROOF. Let $G = (V_N, V_T, P, S)$ be a type-0 grammar. We can suppose, without loss of generality, that all the production rules in P have the form $\chi\alpha\omega \to \chi\beta\omega$ or $A \to a$, where χ and ω are strings in V^* (possibly λ), α and β are strings in V_N^* (strings of variables, indefinite in length), $A \in V_N$ and $a \in V_T$. The obvious reason for this is the same as that discussed in the proof of Theorem 2.10 in Volume I.

We first construct a context-sensitive grammar $G' = (V_N', V_T', P', S)$ which has the following relations with G:

(i) $V_T' = V_T \cup b$ (there is one new terminal element b)

(ii) $V_N' = V_N \cup B$ (there is one new nonterminal element B)

(iii) P' is composed as follows:

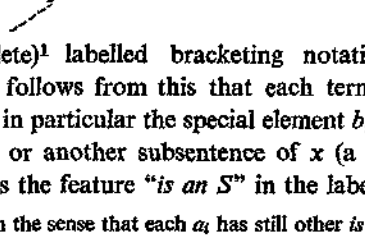
If $\alpha \to \beta$ is a production in P and $|\alpha| \leqslant |\beta|$, then $\alpha \to \beta$ is a production in P' (non-abbreviating productions are taken over unchanged).

If $\alpha \to \beta$ is a production in P and $|\alpha| > |\beta|$, then $\alpha \to \beta\, B^n$ is a production in P', where B^n is a string of n successive B's, and $n = |\alpha| - |\beta|$ (β is thus supplemented with B's until the length of α is attained).

For every A in V_N, P' contains a production $BA \rightarrow AB$, and finally P' contains the production $B \rightarrow b$.

G′ is thus constructed in such a way that it contains no abbreviating productions; it is therefore a type-1 grammar. The language $L(G')$ generated by it has the following relation to the type-0 language $L(G)$: $L(G)$ contains all and only the strings obtained by the deletion of the terminal element b from the sentences of $L(G')$.

The next step is to construct a context-sensitive base grammar in Kuroda normal-form, and equivalent to G' (cf. Volume I, Chapter 2, paragraph 4.2). Such a grammar B exists (cf. Theorem 2.11 in Volume I) and the reader may remember that in Kuroda normal-form the only productions in which S occurs to the right of the arrow are of the form $S \rightarrow SF$; these are also the only *expanding* productions in the grammar. Because S can never exchange places with another element (in the normal-form a production of the form $AB \rightarrow CD$ never contains an S), the tree-diagram for every sentence $x = a_1a_2...a_n$ in $L(B)$ is of the form:



or in (incomplete)[1] labelled bracketing notation, $(_S(_S...(_Sa_1)_S a_2)_S...)_Sa_n)_S$. It follows from this that each terminal element of sentence x, and in particular the special element b, is the rightmost element of one or another subsentence of x (a subsentence is a string which has the feature "*is an S*" in the labelled bracketing).

[1] "Incomplete" in the sense that each a_i has still other *is a* relations to other nonterminal elements. At least one sequence of direct derivations $S \rightarrow A$, $A \rightarrow a$ must be carried out for the generation of a terminal element. For each a there is therefore at least one extra pair of brackets $(_A,)_A$ in a complete labelled bracketing. This, however, is not important to the argument.

The transformational component T of the transformational grammar TG is composed as follows. There is one and only one transformation. This deletes the rightmost element of a subsentence, if that element is b. This means that the transformation is applicable if the bracketing of the subsentence can be divided into two factors, the second of which has the debracketization b. In terms of Definition 3.17, $T = (C, M)$, where C, the structural condition, is $d(\psi^2_{2\rightarrow 2}) = b$, and $M = T_d(\psi_{2\rightarrow 2})$. The result is the deletion of the interior of the second factor, b with corresponding brackets. Further, we allow the transformational component to work cyclically, according to the *Aspects* theory, first on the most embedded sentence, and thence according to Definition 3.18. Thus each b in the labelled bracketing is successively deleted. This transformation satisfies the principle of recoverability (Definition 3.16), because the structural condition states that $d(\psi^2_{2\rightarrow 2})$ is one of a finite number of terminal strings, namely b (see the definition under (ii)). Since the transformation eliminates all the b's from the sentences of $L(B) = L(G')$, it holds that $L(TG) = L(G)$.

The inverse of the proposition also holds, as we see in the following theorem.

THEOREM 5.2. Every transformational grammar with *Aspects* type transformations generates a type-0 language.

PROOF (outline). It follows from the equivalence of type-0 grammars and Turing machines (Theorems 7.1 and 7.2 in Volume I) that it is sufficient to prove that for every transformational grammar with a context-sensitive base there is a Turing machine which accepts $L(TG)$ and only $L(TG)$. In other words, there must be a procedure for the enumeration of the sentences of $L(TG)$ and only the sentences of $L(TG)$. That procedure exists; in its general lines, it is as follows.

Let V_T be the terminal vocabulary of $TG = (B, T)$. Number the strings in V_T^* in the way indicated in Volume I, Chapter 7, paragraph 4. Enumerate the pairs (n, m), where n and m are natural numbers, in the "diagonal" way given in Table 7.1 of Volume I. For every pair (n, m) there is a procedure to determine whether the string in V_T^* with number m has a deep structure in the transformational grammar TG with no more than n subsentences. Such a procedure exists, for the number of sentences in $L(B)$ with no more than n subsentences is finite (as the rules which introduce S are the only recursive rules in the base grammar (cf. Chapter 3, paragraph 1), and if another context-sensitive base grammar is chosen, there is always an equivalent grammar in Kuroda normal-form which does have this property). That finite number of sentences can be enumerated. In the procedure, T is then applied cyclically to each of those sentences. If the result of this contains the string with number m, the string is accepted and "enumerated". If the procedure yields the string with number m for none of those sentences, it goes on to the following pair (n', m'). In this way the Turing machine generates the sentences of $L(TG)$ and only the sentences of $L(TG)$. Thus, $L(TG) = L(TM)$, and $L(TG)$ is of type-0.

The two theorems in this paragraph show the equivalence of the class of type-0 grammars and the class of transformational grammars with a context-sensitive base and *Aspects* type transformations. We can therefore conclude that such transformational grammars offer no guarantee that the language generated is recursive.

5.2. THE GENERATIVE POWER OF TRANSFORMATIONAL GRAMMARS WITH A SIMPLER BASE

At first glance one might be inclined to attribute the overcapacity of transformational grammars pointed out in the preceding paragraph to the rather strong base, a context-sensitive grammar. But this is not where the difficulty lies. It can be shown, in effect, that the equivalence of transformational grammars of the *Aspects* type and Turing machines also holds when the base is of a simpler form. Proof of this was presented more or less simultaneously and more or less independently by Ginsburg and Hall (1969) for a context-free base, by Kimball (1967) and Salomaa (1972) for a regular base, and by Peters and Ritchie (1971) for both. We shall follow the formulations given by the last, because it comes closest to that of *Aspects*. Using the filtering function of transformations, they were able to prove a number of theorems, the most important of which we state (without proof):

THEOREM 5.3. Every type-0 language can be generated by a transformational grammar $TG = (B, T)$, where $B = (V_N, V_T, P, S)$, with $V_N = \{S\}$ $V_T = \{a_1, a_2, \ldots, a_n, b, \#\}$, and the following two productions in P:
(i) $S \rightarrow S\ \#$
(ii) $S \rightarrow a_1a_2\ldots a_nb\#$, and where T only contains *Aspects*-type transformations.

Notice here that B is a right-linear grammar, by which a regular language is generated (Theorem 2.1 in Volume I). The language generated is, moreover, of an extremely elementary kind, i.e. $\{a_1a_2\ldots a_nb\#^m | m > 0\}$. Every base sentence consists of the concatenation of the vocabulary, ending with a b followed by a string of boundary symbols of indefinite length. The labelled bracketing for such a sentence in $L(B)$ has the form:

$$(_S(_S\ldots(_S(_Sa_1a_2\ldots a_nb\#)_S\#)_S\ldots\#)_S\#)_S.$$

Peters and Ritchie show that for every type-0 language L there is a series of transformations, as defined in Definition 2.17, by which this trivial regular set can be transformed into L. Every transformation, moreover, satisfies the principle of recoverability.

Even transformational grammars with such degenerate bases generate undecidable sets, if they contain *Aspects*-type transformations. The main reason for that undecidability is that for a given sentence in such a language there is no upper limit to the number of subsentences in the deep structure. A Turing machine for deciding if a given string *does not* belong to the language would be faced with the hopeless task of seeing whether x could be derived transformationally from each of an infinite set of underlying structures. There is therefore no procedure to determine ungrammaticality;

the complement of language L is not recursively enumerable, and L is therefore not recursive.

5.3. LINGUISTIC CONSEQUENCES

The linguistic consequences of the overcapacity of transformational grammars are great. In the first place, the three conclusions of paragraph 5 in Chapter 2 follow directly: (1) the grammar cannot account for intuitions on ungrammaticality, (2) the language is unlearnable, (3) the chance for descriptive adequacy in the grammar is practically lost, and with it, the possibility of an explanatory linguistic theory. We shall illustrate the last point with the following theorem.

THEOREM 5.4. (Universal base). There is a universal base grammar U, from which all natural languages can be derived transformationally.

PROOF. A trivial example of such a grammar is $U = (V_N, V_T, P, S)$, with $V_N = \{S\}$, $V_T = V_{L_1} \cup V_{L_2} \cup \ldots \cup V_{L_n} \cup \{b\} \cup \{\#\}$, where V_{L_i} is the vocabulary of natural language L_i, and P consists of the productions mentioned in Theorem 5.3. With this base, there is a transformational grammar TG for every language L_i.

An important question in general linguistics is whether a universal base can be found for all natural languages (cf. Chapter 1, paragraph 2). A pet idea among transformational linguists is that transformations tend to be peculiar to specific languages, while the base grammars of various languages coincide to a considerable extent. The theorem on the universal base states that such a base exists on purely formal grounds; the statement, in other words, is not an empirical issue, but only a formal triviality.

The base grammar U is indeed universal, but it is clear that it will generate linguistically absurd deep structures. The strong generative power of U is therefore insufficient. The universal base must also be descriptively adequate, and linguists could maintain that it is very much an empirical question whether a descriptively adequate universal base can be found. But this appears not to be the case. Peters and Ritchie (1969) show that if the class of transformational grammars is limited to those grammars which have an upper limit to the number of subsentences in the deep structure of a sentence (for example, a limit which is a function of the length of the sentence), universal bases exist which have a strong generative power sufficient for linguistic purposes. More specifically, Peters and Ritchie define "sufficient strong generative capacity" as follows. Such transformational grammars can account for intuitions on:

(i) the grammaticality and ungrammaticality of sentences,
(ii) the number of different structural descriptions which an ambiguous sentence should have,
(iii) which sentences are paraphrases of each other in at least one respect.[1]

The introduction of the upper limit means, of course, that we only consider the transformational grammars which generate recursive languages. The Turing machine in the preceding paragraph which was to decide on ungrammaticality is no longer confronted with what we have called a "hopeless task". The number of deep structures it must examine is now limited to a certain number of subsentences, and a decision can be made after a finite number of steps in the procedure. If we make the transformational grammar decidable by building an upper limit into it, then (i) will follow automatically. The number of different deep structures at the base of a given sentence x also becomes decidable, and for the same reason only a finite number of deep structures need be examined in order to determine how many of them lead to a transformational derivation of x. From this, (ii) follows. A similar argument holds for (iii): given sentence x, there is only a finite number of deep structures for x, each of which leads to only a finite number of transformational derivations, one of which is x.

[1] One might wonder if no requirements should be stated on the parsing of the sentences generated.

The other transformational derivations (non-x) of these deep structures are precisely the sentences which are paraphrases of x in at least one respect.

Provided that we suppose that there are transformational grammars for natural languages, with the extra, but extremely reasonable restriction that for every sentence there is a certain upper limit on the size of the deep structure, we can state that there must be a universal base by which such transformational grammars possess the descriptive adequacy specified in (i), (ii) and (iii). Thus also from the point of view of descriptive adequacy, the question as to whether or not a universal base grammar exists is no empirical question. For purely formal reasons, there is a class of bases which satisfy all three requirements, and we shall, thus, never be able to tell which of two universal bases is correct, if both belong to that class.

More serious still, we cannot even decide if a base for a particular natural language is the correct one. If that were possible, we would in principle be able to decide that two natural languages have different bases, which would conflict with both the strong and the weak versions of the theorem on the universal base.

The importance of this impasse for linguistics should not be underestimated. The whole controversy between generative and interpretative semanticists, for example (cf. Chapter 3, paragraph 3), is carried on in transformational terms which do not differ essentially from the formulation used by Peters and Ritchie. Where deviations do occur, namely, by the addition of *derivational constraints* on transformational derivations, this only leads to increases in the generative power of the grammar, and not to reductions of it. As long as both parties work inside the class of universal bases, it will remain impossible to tell who is right. The controversy does not deal with an empirical question, and it is not unlikely that this is, entirely or to a great extent, the case. This appears in the practice of generative semanticists of freely using that property which precisely leads to undecidability, namely, extremely extensive deep structures. One can often see a veritable morbid growth of recursive embedding for the description of a three or four word sentence; this is the case, for example in Figure 3.9a.

To summarize, we can state that the main problems of linguistics are insolvable by the formal means of the *Aspects* type. Such a linguistic theory can make no judgment on observational adequacy, if this is defined as the possibility of deciding whether or not sentence x belongs to language L. Nor can it satisfy descriptive adequacy (no decision can be made as to which grammar offers the "correct" structural description), or explanatory adequacy (no decision can be made as to which is the universal base, and the theory can give no account of the learnability of natural languages).

5.4. SOLUTIONS FOR THE PROBLEM OF TRANSFORMATIONAL OVERCAPACITY

The principal cause of the undecidability of *Aspects* type transformational grammars is the fact that there is no upper limit to the size of the deep structure of a given sentence. As a consequence of this, an infinite number of underlying structures must be examined in order to make the decision "x is not in L". It was precisely the purpose of the principle of recoverability to avoid this. The principle should have guaranteed that a Turing machine (and, in the abstract, the native speaker) be able, for every string of words, either to reconstruct the deep structures, or to state that there are no deep structures for the string. But on reflection, it is striking to notice how poorly the *Aspects* definition (faithfully formalized in Definition 3.16) fulfills that original purpose. The principle guarantees only that if a labelled bracketing and the transformations by which it was derived are given, it will be possible to reconstruct the original labelled bracketing. For the reconstruction of the deep structure, therefore, the Turing machine would also have to dispose of a list of the transformations used, or at least of the maximum number of transformations which can be performed in a derivation. This would guarantee that no more than a finite number of transformational derivations need be reconstructed.

For each of these it could be determined if the first labelled bracketing in that derivation is generated by the base grammar. The principle, however, does not provide such a guarantee. It allows that for every labelled bracketing there is an earlier labelled bracketing, since a cycle of transformations can always have come before. Suppose, for example, that the word "*a*" is a sentence in *L*(*TG*). It follows from the construction in the proof that the sentence can be derived from each of the infinite number of deep structures $(_S a)_S$, $(_S(_S a)_S b)_S$, $(_S(_S(_S a)_S b)_S b)_S$, ...

A step in the direction of a solution would therefore be to set a limit on this unrestricted cyclic capacity of the transformational component. There are two ways in which this might be done. The first is empirically to establish whether or not in current linguistic practice any upper limit to the number of subsentences in the deep structure is implicitly taken into consideration. Peters and Ritchie (1973) suppose that this is indeed the case. They state that a number k can certainly be found for which a sentence x of length $|x|$ has fewer than $k|x|$ subsentences in its deep structure. They show that this is sufficient to guarantee the recursiveness of such transformational grammars. But this is a very non-committed method. What is needed is an argument for that upper limit. The second way is therefore more interesting: is it possible to change the definition of transformation (including the principle of recoverability) in such a way that the upper limit will automatically follow from it? This has not yet been done for *Aspects* transformations. The only mixed model for which it has been done is the Joshi adjunct grammar. In Chapter 4, paragraph 4 we discussed the trace condition in that grammar. The trace condition requires that each transformation leave one or more elements behind, and that those elements (or that element) may no longer be deleted by further transformations. It is obvious that for a sentence of a given length there is an upper limit to the number of transformations which are applied to that sentence in derivation, and Joshi (1972) shows that this does indeed guarantee the recursiveness of his grammar. From an empirical point of view, however, it remains an open question whether the trace condition holds in all cases. If it holds for the transformations of an adjunct grammar, it need not necessarily hold for the transformations of grammars of the *Aspects* type. Moreover, the trace condition is applied to the transformational component as a whole, and not to individual transformations: the trace of a transformation must remain in every *possible* transformational derivation. It would be a rather heavy empirical task to account for the plausibility of such a condition.

However, more has to be solved than only the problem of decidability. As we have seen, a strong form of the theorem on the universal base is maintained, even if only decidable transformational grammars are taken into consideration. This may be attributed to the filtering function of transformations. Every type-0 language can be derived from a trivial base, by the intensive use of the filtering function of the transformational component.

The filtering possibility should either be eliminated from the model, or at least limited. It would be interesting here to find linguistic arguments for one or another solution, but until now little effort has been made in that direction. An empirically interesting question, for example, is whether a $\#$ which occurs within the domain of a particular transformational cycle and is not removed during that cycle can still be eliminated in a later cycle. The *Aspects* theory allows this, but the need of it, from a linguistic point of view, is doubtful. If, for example, a relative clause transformation in a particular cycle fails because the structural condition $NP_1 = NP_2$ is not fulfilled (cf. Chapter 3 paragraph 1.2), it is unlikely that this might be "repaired" in a later cycle. On such ground the filtering function of transformations might be sufficiently limited to give the question of the universal base empirical content.

This all should encourage great reserve concerning the grammatical means used and the range of the results attained. Since the publication of *Aspects*, however, interest in the formal structure of grammars has rather decreased than increased. Very many interesting linguistic phenomena have been discovered and discussed, but their formulations are only details of a theory which as yet does not exist. Such formulations are always based on implicit or explicit assumptions concerning the theory as a whole, justification is lacking precisely on essential points. The assumption on the universal base, for example, is incorrectly considered empirically verifiable in the present state of theory. History is obviously repeating itself; in 1965 Chomsky wrote:

> The critical problem for a grammatical theory is not a paucity of evidence, but rather the inadequacy of present theories of language to account for masses of evidence that are hardly open to serious question (*Aspects,* 20).

On a different level, this applies as well to the present situation.

6

STATISTICAL INFERENCE IN LINGUISTICS

We have so far been concerned in this volume with linguistic theory, and have not yet treated the interpretation problem (Chapter 1, paragraph 2) from the point of view of formal grammars. Of the three cases in which that problem appears most strikingly, two, the investigation of linguistic intuitions, and the investigation of language acquisition, will be treated in Volume III. In the present chapter we shall deal with a few applications of formal language theory to the third case, statistical inference with respect to the analysis of a corpus. This chapter will not offer a survey of statistical linguistics; the discussion will be limited to two examples which are relevant to psycholinguistics in particular. The aim of the chapter is principally to show that the interpretation problem calls for linguistic methods other than the "usual" ones, and that the widespread opinion that statistical methods are inappropriate in linguistics is not only unfounded, but it is also a hindrance to linguistic research on interpretation. In the first paragraph (6.1) we shall discuss a few aspects of communication theory from the point of view of inference theory; some linguistic applications of communication theory can be considered as statistical inference with respect to regular grammars. In the second paragraph (6.2) we shall show a linguistic application of the material treated in Volume I, Chapter 8, paragraph 2; this will consist of an estimate of parameters for a probabilistic context-free grammar.

6.1. MARKOV-SOURCES AND NATURAL LANGUAGE

In Chapter 2, paragraph 2 we showed that regular grammars are decidedly unsuited for describing natural languages. But there is a class of probabilistic finite automata which has long served as a model in the analysis of natural languages; the class in question is that of *Markov-sources*. Although there is no doubt that such models are inadequate as linguistic theory, it is nevertheless a practical fact that they are often suitable means for the description of rough parameters of verbal communication processes. They are still used as such in applied communication theory. These rough parameters refer to that which is called the *information value* of the verbal message. "Information" in this sense of the word is a quantitative concept, distinct from the content or meaning of the message. Moreover, it is not an absolute, but rather a relative concept. In information theory it is impossible to say how much information an isolated message contains. Information is defined precisely on the basis of the number of alternative messages which the same source could have produced in the same length of time. The information value of a message should indicate, given the source, the probability of that message. The idea is that a message with a probability of 1 contains no information, for the receiver can predict exactly what the source will produce in that length of time. Only when some uncertainty exists, will the message contain information. Information is equal to the amount of uncertainty which the message eliminates.

The nature of the source is determinant for the probabilities of the various messages, and consequently for their information value. If the source is discrete, that is, if it has a finite vocabulary, we can consider it as a probabilistic grammar, a system which generates sentences with particular probabilities (cf. Volume I, Chapter 3). The most important generalizations in communication theory concern right-linear sources (cf. Volume I, Chapter 2, paragraph 3.5), thus sources which generate regular languages. This is not an essential restriction; a context-free probabilistic grammar might also be taken as source, and the definitions of information, redundancy, etc. would not have to be altered (an example of this has been treated in the discussion of grammar-grammars in Volume I, Chapter 8, paragraph 4). For historical reasons, however, the restriction does exist, and it is carried over into the applications of communication theory to natural languages.

In the simplest case the source of messages is considered as a finite automaton with as many states as vocabulary elements. Each vocabulary element ($a_1, a_2, ..., a_n$) serves as the label of one state ($s_{a_1}, ..., s_{a_n}$), and the transition rules are such that the automaton always passes to the state labelled after the element it has just accepted. The state transition function δ thus contains all and only the rules $\delta(s_{a_i}, a_j) = s_{a_j}$) for all s_{a_i}, s_{a_j} in S, and every a in I. It is clear, then, that all the states are connected with each other, and that the automaton is 1-limited. Finally, every state is assigned a probability, normalized on the basis of the *state* (cf. Volume I, Chapter 4, paragraph 4), that is, the total probability of a transition from a given state, over all possible input symbols (the entire vocabulary), is equal to 1. Such an automaton is called a *Markov-source*. Before defining this more formally, we offer the transition diagram for an elementary Markov-source in Figure 6.1. The source has two elements in the

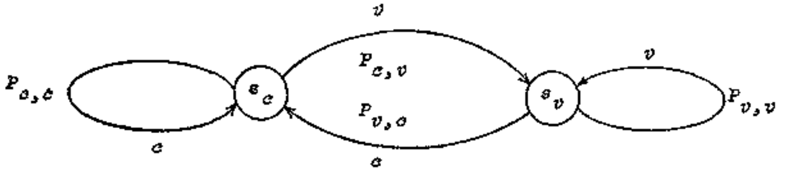


Fig. 6.1. An elementary Markov-source.

input vocabulary, c and v, and therefore two states, s_c and s_v. When the input is v, the automaton passes to state s_v, and when the input is c, to state s_c. The chance that the automaton will go from state s_c to state s_v is indicated in the diagram as $p_{c,v}$ for that transition, of course, the input must be v. The chance for the opposite transition, from s_v to s_c, is $p_{v,c}$, and the chances for transition to the same state are respectively $p_{c,c}$ and $p_{v,v}$. Normalization means that the total chance of transition from a given state is equal to 1. Thus the total chance of transition form s_c is $p_{c,c} + p_{c,v} = 1$, and the total chance of transition from s_v is $p_{v,c} + p_{v,v} = 1$. We must now determine the state in which the automaton starts, and that in which it stops. In complete analogy with the definition of probabilistic finite automata (Volume I, Chapter 4, paragraph 4), it is customary to consider all the states of a Markov-source as possible initial states, and for each of these states s_i an initial probability $p(s_i)$ is defined. The sum of the initial probabilities is equal to 1. The vector of initial probabilities $(p(s_1), \ldots, p(s_n))$ is called the INITIAL DISTRIBUTION of the Markov-source. It also holds for a Markov-source that every state is a final state. For a certain class of Markov-sources the initial distribution is of little importance; the statistical properties of a long, generated string are, in the limit, independent of the initial state. Markov-sources with this characteristic are called ERGODIC. For them, we can simply suppose that they are generating from infinity, rather than defining an initial distribution. Because every state can be a final state, we can likewise suppose that the source never stops, and generates a string which is infinite both to the left and to the right. Each finite segment of that infinite string is then a sentence (message), generated (accepted) by the Markov-source. As linguistic applications of communication theory always suppose ergodic sources, we shall limit further discussion to this subclass, and omit definition of an initial distribution.

A MARKOV-SOURCE, then, is completely characterized by its finite INPUT VOCABULARY I $\{a_1, \ldots, a_n\}$, and its TRANSITION PROBABILITIES, $p_{i,j}$, where $p_{i,j}$ is defined for all pairs a_i, a_j (with a_i, a_j in I) and stands for the chance that element a_i is followed by element a_j, and in which the probabilities are normalized as follows: $\sum_{j=1}^{n} p_{i,j} = 1$. Because of the one-to-one relation between input vocabulary and state set, it would be redundant to include this last in the characterization of a Markov-source.

Such a Markov-source can quite as well be written as a regular grammar, with rules of the form $A_i \overset{p_{i,j}}{\rightarrow} a_j A_j$ for every pair a_i, a_j in the terminal vocabulary. Such a grammar is thus considered to generate a string infinite to the left and to the right.

The input vocabulary and the transition probabilities for the Markov-source in Figure 6.1 are given in the following transition matrix, which gives a complete characterization of the source:

P	v	c
v	$p_{v,v}$	$p_{v,c}$
c	$p_{c,v}$	$p_{c,c}$

This source is a linguistic example *par excellence*. It is the model which Markov (1913) constructed for the description of the sequence of vowels (v) and consonants (c) in Pushkin's *Eugene Onegin*, and the origin of the Markov theory. It is a clear example of the problem of inference: given a corpus (Pushkin's text) and a grammar (the finite automaton in Figure 6.1), can the transition probabilities be estimated? Markov found estimates by determining, for 20,000 pairs of consecutive letters (*digrams*), to which of the four categories, vv, vc, cv, cc, they belonged. He found the frequencies given in Table 6.1, with the corresponding transition

TABLE 6.1. Digram frequencies and transition matrix for *Eugene Onegin*.

	vv	vc	cv	cc	*total*
Digram frequencies	1104	7534	7534	3827	19999

Transition probabilities:

P	v	c
v	0.128	0.872
c	0.663	0.337

matrix. (The number of digrams is, of course, one less than the number of letters.) The value $p_{v,v} = 0.128$ means that of the 1000 vowels, an average of 128 were followed by another vowel, etc. There appears to be a preference for the alternation of vowels and consonants, since the chance for two consecutive consonants or two consecutive vowels is relatively small. How good is this model? Does it, for example, give correct predictions on the chances of *trigrams* such as *vvv*, *vcv*, etc.? If not, is there a better source for the description of vowel/consonant sequences? Before going into these questions, we give a somewhat more ambitious example of a Markov-source in linguistics, not for letter orders, but for word sequences.

Suppose that English has 100,000 words. We can imagine a Markov-source with 100,000 states, corresponding to the 100,000 words. The source can again be characterized completely by its transition matrix P. The matrix element $p_{i,j}$ stands for the chance that word i is followed by word j. The matrix will thus contain $100{,}000^2 = 10^{10}$ probabilities. Since the source is normalized, the rows of the matrix add up to 1. In each row, therefore, there are $100{,}000-1$ *independent* p-values, and the model contains a total of $10^{10}-10^5$ independent parameters. It holds in general that a Markov-source with n elements in the input vocabulary has n^2-n independent parameters. Obviously no one has undertaken the impossible task of determining these parameters for English, and it seems excluded that we might make a judgment on the quality of the English generated by this Markov-source. However, means have been found for arriving at some impression of that which is generated by the source. One way is to present speaker A with a word, for example *the*, and to ask him to compose a sentence in which that word occurs. Let us suppose that A forms the sentence *that is the head.* We then go on the the word which follows *the*, namely *head,* and ask speaker B to compose a sentence in which that word occurs. If B in turn produces the sentences *head and feet are parts of the body*, we take the word following *head*, namely *and*, and go on to speaker C, and so forth. The sequence of words obtained in this way, *the, head, and,* etc., may be considered to be generated by the Markov-source. We call such a sequence a SECOND ORDER APPROXIMATION of English. A FIRST ORDER APPROXIMATION can be imagined by analogy; it is based on the probability of occurrence of the various individual words (and not pairs) in English. It could be composed, for example, by taking the twenty-fifth word of every column in a newspaper, and forming a list of them in sequence. The more probable a word is in English, the greater the chance of meeting it in the local newspaper. For this we would imagine a probabilistic automaton with only one state, where the input of any word will bring the automaton to that state. The chance that a given loop be chosen is equal to the probability of the word in question in the language. This is shown in Figure 6.2.

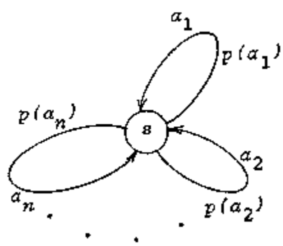


Fig. 6.2. A probabilistic automaton for first and zero order approximations.

A ZERO ORDER APPROXIMATION is a string of words chosen at random and without regard for their frequency of occurrence in the language. Such an approximation could be made, for example, by taking every thirteenth word which occurs on a page, chosen at random, of an English dictionary. If all the loops in the automaton in Figure 6.2 have equal probabilities, the automaton generates a zero order approximation.

The following are examples of zero, first and second order approximations (the first and second order approximations are taken from Miller and Chomsky (1963)); only the second order approximation can be taken as generated by a Markov-source.

ZERO ORDER: *splinter shadow dilapidate turtle pass stress grouse appropriate radio whereof also appropriate gourd keeper clarion wealth possession press blunt canter chancy vindicable corpus*

FIRST ORDER: *representing and speedily is an good apt or came can different natural here he the a in came the to of to expert gray come to furnish the line message had be these*

SECOND ORDER: *the head and in frontal attack on an English writer that the character of this point is therefore another method for the letter that the tired of who even told the problem for an unexpected*

Although the Markov-source clearly produces "better English" than the zero and first order approximations, the result is rather disappointing if we consider any other characteristic than immediate succession of words. Every sequence of five or more words looks strange. This is disappointing because the very limited result is attained by a model with an astronomically high number of parameters. It is more difficult to evaluate sequences of three or four words, and there is no method to determine how good the source is in this respect. This brings us back to Markov. For his vowel/consonant model it is possible to determine how well trigrams (and longer strings) are predicted, for we know the precise values of the transition probabilities in the model. The chance for a trigram *ccc* is equal to the chance for a digram *cc*, $p(cc)$, multiplied by the chance that the second *c* be followed by another *c*, $p_{c,c}$. The best estimate of $p(cc)$ is the relative digram frequency (cf. Table 6.1), $\frac{3827}{19999} = 0.191$.[1] The expected relative frequency for the trigram *ccc* is thus $0.191 \times 0.337 = 0.065$. Table 6.2 shows the

TABLE 6.2. Expected and Observed Relative Frequencies of Trigrams in *Eugene Onegin*.

	vvv	*vvc*	*vcv*	*vcc*	*cvv*	*cvc*	*ccv*	*ccc*
Expected	0.007	0.048	0.250	0.127	0.048	0.329	0.126	0.065
Observed	0.006	0.049	0.211	0.166	0.049	0.327	0.166	0.025

[1] One might wonder if $p(cc)$, which is determined on the basis of the digram frequency, is indeed predicted by the model, i.e. on the basis of the transition matrix. It may be argued as follows that this is in fact the case. For an ergodic Markov process, the probabilities of the various elements ($p(v)$ and $p(c)$ in the example) are given in the stochastic eigenvector of P, i.e. the vector α for which $\alpha P = \alpha$. In the example, $\alpha = (p(v), p(c))$, and the vector is stochastic because $p(v) + p(c) = 1$. The value of α can be found here by solving the equation $p(v)p_{v,v} + p(c)p_{c,v} = p(v)$. Substitution of $p_{v,v} = 0.128$ and $p_{c,v} = 0.663$ (cf. Table 6.1.) yields $p(v) = 0.432$ and $p(c) = 0.568$. The chance for the digram *cc* is then $p(c).p_{c,c} = 0.568 \times 0.337 = 0.191$, which corresponds to the actual value.

expected and observed frequencies for the eight possible trigrams. The Markov model is evidently quite accurate in this respect, but this is further left to the judgment of the reader.

If the predictions for trigrams (or n-grams of a higher order) are not considered satisfactory, a more complicated model can be chosen. One could select a model based on the probability of transition, not from letter to letter (or from word to word), but from *pair* of letters to letter; thus, for example, the probability of v after the sequence vc is $p_{vc,v}$. A finite automaton can also be constructed for this end. Such an automaton, unlike the Markov-source, will have a state for every *pair* of letters; for n vocabulary elements, then, there will be n^2 states in the automaton. Therefore for the vowel/consonant example, four states will be needed, s_{cv}, s_{cc}, s_{vv}, s_{vc}. The automaton is shown in Figure 6.3, and is called a PROJECTED MARKOV-SOURCE. It is a 2-limited automaton,

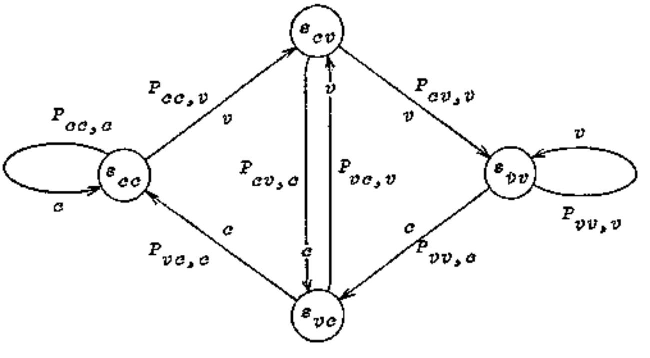


Fig. 6.3. A Projected Markov-source.

because every sequence of two input elements unambiguously determines the state of the automaton. Each state is labelled according to a sequence of two input elements which necessarily lead to it, and each state has as many inputs and outputs as there are vocabulary elements. In Figure 6.3 there are two vocabulary elements, and eight transition probabilities of the form $p_{ij,k}$, which logically find their places in the model. In the projected

Markov-source, the transition probabilities are also normalized according to state. Just as the model in Figure 6.1 represents the exact digram frequencies, the projected Markov-source represents the exact trigram frequencies. The characteristics of the automaton can once again be summarized completely in a transition matrix; this, however, will not be square (2×2), but rather rectangular ($2^2 \times 2$), with the rows labelled according to the pairs, and the columns according to the vocabulary elements.

P	v	c
vv	$p_{vv,v}$	$p_{vv,c}$
vc	$p_{vc,v}$	$p_{vc,c}$
cv	$p_{cv,v}$	$p_{cv,c}$
cc	$p_{cc,v}$	$p_{cc,c}$

Like the ordinary Markov-source, the projected Markov-source can be represented as a grammar. If the terminal vocabulary contains n elements, $\{a_1, \ldots, a_n\}$ the nonterminal vocabulary will contain n^2 elements $\{A_{11}, A_{12}, \ldots A_{21}, A_{22}, \ldots A_{nn}\}$ and n^3 productions of the form $A_{ij} \xrightarrow{p_{ij,k}} a_k A_{jk}$.

We can now attempt to generate English by means of the 2-limited projected Markov-source. This will give English to a *third order approximation*. With $n = 100{,}000$ words, there are $n^2 = 10^{10}$ states, and $n^3 = 10^{15}$ transition probabilities $p_{ij,k}$. Of these, there are $n^3 - n^2$ independent parameters (by normalization, one column of the transition matrix becomes redundant). Calculation of all these parameters is excluded, but to have an impression of how well the projected Markov-source can generate English, we can once again play the above game with speakers. We present speaker A with a *pair* of words (chosen at random from a newspaper or from a sentence composed by another speaker), for example, *family was*, and ask him to form a sentence in which the pair occurs. Suppose that the sentence which he produces is *the family was large*. We then present *was large* to speaker B, and request that he in turn form a sentence in which this pair occurs. If his sentence is *the forest was large, dark and dangerous* we present *large* *dark* to speaker *C*, and so forth. The following string (Miller and Chomsky 1963) was obtained in this way.

THIRD ORDER: *family was large dark animal came roaring down the middle of my friends love books passionately every kiss is fine.*

Obviously we can go on to construct still higher order projected Markov-sources. A 3-limited source, the vocabulary of which has n elements, will have n^3 states, and each state will have n inputs and outputs. Every output of a state has a probability, and the transition probabilities are normalized for each state. An example of an approximation of English, generated by a 3-limited source, is the following (Miller and Selfridge 1950);

FOURTH ORDER: *went to movies with a man I used to go toward Harvard Square in Cambridge is mad fun for*

In general, a k-limited projected Markov-source has n^k states, and therefore a $n^k \times n$ matrix of transition probabilities. The number of independent parameters in such a model is thus $n^{k+1} - n^k$. The following is an example of a fifth order approximation of English (Miller and Chomsky 1963).

FIFTH ORDER: *road in the country was insane especially in dreary rooms where they have some books to buy for studying Greek*

All Markov-sources are k-limited, but as we have seen (in Volume I, Figure 4.5), not all finite automata are k-limited. Consequently it is not the case that all regular languages can be generated by Markov-sources.

The five approximations of English given in the course of this paragraph were progressively "better". The higher the order, the more predictable the text, and therefore the less "informative" (according to the definition given in communication theory). A zero order approximation is a string in which all elements have an equal chance to occur. Suppose we take a random segment of m elements from the infinite string produced by a zero order automaton. How great is the chance that a second random segment of m elements will contain the same elements as the first segment, and in the same order? The probability that the second segment begins with the same element as the first is $\frac{1}{n}$, if the vocabulary contains n elements. The chance that the second element is the same is also $\frac{1}{n}$, and so forth. The chance that the entire segment is the same is therefore $p = \left(\frac{1}{n}\right)^m$. Suppose that the vocabulary contains only one element, $n = 1$; in that case any two segments of m elements will be identical, for $p = \left(\frac{1}{1}\right)^m = 1$. Predictability is then complete, the message does not reduce uncertainty, and there is no information. The uncertainty of a message is defined as the logarithm (base 2) of the probability p of that message. In the example, $p = 1$, and the uncertainty is therefore $\log p = 0$. Uncertainty increases with the number of vocabulary elements. For this source, the uncertainty relative to a segment of m elements is $\log\left(\frac{1}{n}\right)^m = m \log\frac{1}{n}$. The information H, the amount of uncertainty reduced, is defined as the complement of the uncertainty. With a zero order approximation, we therefore have $H(0) = -m \log\frac{1}{n}$ for a string of m elements.

For a first order approximation, the probabilities p_i of the various vocabulary elements a_i are not necessarily equal. How great is the information $H(1)$ of a random segment with m elements? If m is large, a string of m elements should contain the word a_1 about mp_1 times, the word a_2 about mp_2 times, and in general, the word a_i about mp_i times. The chance for the entire string is once again the product of the probabilities of the individual elements. Since the element a_i occurs approximately mp_i times, and a_i occurs with probability p_i, the probability of this string of m elements is $p = p_1{}^{mp_1}.p_2{}^{mp_2} \ldots\ldots p_2{}^{mp_n}$, and $H(1)$ is therefore approximately $-\log p = -(mp_1 \log p_1 + mp_2 \log p_2 + \ldots + mp_n \log p_n) = -m \sum_i p_i \log p_i$. If all p_i are equal, thus $p_i = \frac{1}{n}$, then the information will, of course, be equal to that of the zero order approximation, $-m \log \frac{1}{n}$. If the probabilities are not equal, H is smaller. Therefore $H(0) \geqslant H(1)$.

One could go on to prove that $H(0) \geqslant H(1) \geqslant H(2), \ldots$, and in general that $H(i) \geqslant H(i+1)$. The information will be equal only when probabilities or transition probabilities are equal. For English, these are obviously unequal, and it holds, therefore, that $H(i) > H(i+1)$: the higher the order, the less informative (or more redundant) the text. In Volume III, Chapter 2, paragraph 2 we shall examine psychological applications of this. General introductions to communication theory may be found in other literature; a few sources are mentioned in the bibliographic survey at the end of this volume.

6.2. A PROBABILISTIC GRAMMAR FOR A CHILD'S LANGUAGE

The simplest case of statistical inference occurs when grammar and corpus are given, and production probabilities must be deduced. The procedure necessary for this is treated in detail in Voume I, Chapter 8, paragraph 5; the grammar in question was context-free.

An interesting linguistic example of this method is Suppes' analysis of the language of Adam, one of Brown's young subjects. The corpus analyzed by Suppes was recorded when Adam was two years and two months old, and consists of eight hours of tape recordings. After the elimination of immediate repetitions, the corpus contains 6109 words, over a vocabulary of 673 different words. It was segmented into 3497 utterances ("sentences"). Suppes analyzed this material in various ways. He attempted to give a complete grammar for it (which we shall discuss later), and he made an analysis of only the noun phrases in the material.

There is a certain amount of freedom for the definition of noun phrase, but as soon as a grammar is written, the sequences of categories which may be called "*NP*" are clearly determined.

Uncertainties concerning the frequencies in the corpus can only come about then through uncertainty in the categorization of individual words. Is *fly*, for example, a noun or a verb? This is the sort of problem of interpretation which typically occurs in applied linguistics. Suppes (1970, 1971) gives the following context-free *NP* grammar:

1. $NP \overset{a_1}{\rightarrow} N$
2. $NP \overset{a_2}{\rightarrow} AP$
3. $NP \overset{a_3}{\rightarrow} AP + N$
4. $NP \overset{a_4}{\rightarrow} P$
5. $NP \overset{a_5}{\rightarrow} NP + NP$
6. $AP \overset{b_1}{\rightarrow} AP + A$
7. $AP \overset{b_2}{\rightarrow} A$

The symbol *P* stands for *pronoun*, and *AP* for *adjective phrase*. The production probabilities are denoted by $a_1, \ldots a_5, b_1, b_2$. The grammar is normalized, and therefore $\Sigma a_i = 1$ and $\Sigma b_i = 1$. Consequently there are five independent parameters in the model.

A typical sentence in the corpus is *take off Adam paper*, with the *NP Adam paper*. The noun phrase is of the form $N + N$; a leftmost derivation of it is $NP \overset{a_5}{\Rightarrow} NP+NP \overset{a_1}{\Rightarrow} N+NP \overset{a_1}{\Rightarrow} N+N$. If it is supposed that the productions are applied independently, then $p(NN) = a_5.a_1.a_1$. The chances for all other observed *NP* forms are also determined in this way on the basis of the grammar. This led, according to the procedure presented in Volume I, Chapter 3, paragraph 5, to estimates of the seven production probabilities (cf. Table 6.2), and the calculation of expected frequencies for the various types of *NP*. The latter are given, together with the observed frequencies in Table 6.3. The difference in total between observed

TABLE 6.2. Estimated Production Probabilities for the *NP*-Grammar for Adam's Language

a_1	a_2	a_3	a_4	a_5	b_1	b_2
0.6391	0.0529	0.0497	0.1439	0.1144	0.0581	0.9419

TABLE 6.3. Observed and Expected Frequencies of Various Noun Phrase Types in Adam's Language

NP-Type	*Observed*	*Expected*	*NP-Type*	*Observed*	*Expected*
N	1445	1555.6	*PPN*	6	0.4
P	388	350.1	*ANN*	5	8.3
NN	231	113.7	*AAN*	4	6.6
AN	135	114.0	*PA*	4	2.0
A	114	121.3	*ANA*	3	0.7
PN	31	25.6	*APN*	3	0.1
NA	19	8.9	*AAA*	2	0.4
NNN	12	8.3	*APA*	2	0.0
AA	10	7.1	*NPP*	2	0.4
NAN	8	8.3	*PAA*	2	0.1
AP	6	2.0	*PAN*	2	1.9
			Total	2434	2335.8

and expected frequencies (98.2) is due to the fact that the grammar generates other (longer) noun phrases which do not occur in this corpus. There are also very noticeable differences in detail among the various types of noun phrase. Thus the actual frequency of the sequence $N+N$ is considerably underestimated in the theoretical expectations. Many *NN* sequences prove to be possessive from the context, such as *Adam bike* and *Daddy suitcase*. Others, however, are "is a" relations, like *toy train* and *lady Ursula*, and still other *N*/*N* relations have been distinguished in the material. One might consider introducing a separate possessive production rule to obtain better theoretical predictions. Or one could introduce statistical dependencies between productions. But linguists will be more inclined to treat these differences transformationally. Although there is no theoretical difficulty in writing a probabilistic transformational grammar, important practical problems are involved. It would be necessary (a) to assign production probabilities to the base grammar and (b) to assign probabilities to optional transformations. But then special provisions would have to be made for ambiguities, and, in grammars of some complication, for the treatment of transformational filtering. Suppes attempted to refine the grammar with regard to the *NN* sequences by means of a semantic analysis, which we shall not discuss further here.

The point here is to show the strength of such a probabilistic analysis. Direct information is obtained on which production rules do the actual work in the grammar, and which are used only occasionally. But above all there is a direct feed-back on which rules fail, and thus on the direction in which further improvement of the grammar must be sought. We shall return to this subject at the end of this paragraph.

Suppes attempted to write a complete grammar for the corpus; the form he chose was that of a probabilistic categorial grammar. The very limited number of rules in a categorial grammar (cf. Chapter 4, paragraph 2) restricts the freedom of movement to such a small number of parameters, that the undertaking — however interesting — is bound to fail, as was indeed the case for Suppes' grammar. Success would have meant a deep insight into the structure of the child's language; it would have meant that the child's syntax develops exclusively by the differentiation of categories, and changes in rule parameters. The number of parameters (and rules) would, however, be small and constant throughout the development. For details on this, we refer to Suppes (1970).

A probabilistic grammar also gives various additional information which can be of great use in applied linguistics. On the basis of such a grammar characteristics of the corpus can be treated, which might lie beyond the range of theoretical linguistics, but which are sometimes the *pièce de résistance* in practical applications. Thus *sentence length* is an essential variable in the analysis of style, in the analysis of children's languages, in the investigation of speech intelligibility, etc. An accurate probabilistic grammar also provides a description of sentence length in the corpus, as well as the distributions of length of other constituents.

An example can again be taken in Suppes' analysis of Adam's noun phrases. With the given grammar, a noun phrase of length 1 can be derived in three ways: (i) by applying production 1: NP will then be rewritten as N, with probability a_1; (ii) by first applying production 2, then production 7: $NP \Rightarrow AP \Rightarrow A$ with $p(A) = a_2 . b_2$; (iii) by applying production 4: NP will be rewritten as P, with probability $p(P) = a_4$. The total probability of a noun phrase of length 1 is thus $p(1) = a_1 + a_2.b_2 + a_4$. For the production probabilities in Table 6.2, this is $p(1) = 0.8329$. Of the 2434 noun phrases in the corpus, there should therefore be $2434 \times 0.8329 = 2027$ of length 1. The observed value is 1947. The expected value for length 2 can be calculated in the same way; for Adam's noun phrases the expected value is 314, and the observed value is 463. Likewise for length 3, the expected value is 67, the observed, 51, and 26 noun phrases are expected of length greater than 3, but none occur in Adam's speech.

One of the most noteworthy advances in the modern investigation of children's languages is that which one could call the *linguistic* method. In the 1960's explicit grammars were written for the first time for the languages of two and three-year-olds. Language development was studied for the first time from the point of view of grammar, and such matters as the differentiation of categories and rewrite rules and the growth of transformational skills such as in negation and question were investigated. In the meantime this research has begun to be integrated into a much wider framework, that of the cognitive-conceptual development of the child; we shall return to this subject in Volume III, Chapter 4. But in the beginning, the opinion of the transformational linguists of the time was the touchstone for this renewal, and consciously or unconsciously many accidental attitudes were taken over from them into the practice of research. One of these attitudes was an aversion to statistical concepts. In 1969 Chomsky wrote "It must be recognized that the notion 'probability of a sentence' is an entirely useless one, under any known interpretation of this term". Traditionally however, research on children's languages was very much interested in the development of the statistical aspects of the language, the development of sentence length, frequencies of the various types of sentences and classes of words, etc. There can be no doubt but that a complete theory of the development of children's languages must also be able to explain those phenomena. A(non-probabilistic) grammar is perhaps half the work in this, but it is still no more than a good beginning. Probabilistic grammars, however, make it possible to establish the relations between modern structural linguistic insights and the abundance of traditional statistic data on the development of children's languages. The reason for such an approach is not simply the need to reconcile (apparent) contradictions, but rather the desire to find a structural explanation for the *patterns* which appear in those statistical phenomena. The change of one parameter in a probabilistic grammar can lead to statistical changes in very divergent aspects of the corpus generated, for example, simultaneous changes in the frequency of words of a certain class and in the distribution of sentence lengths. If the relationship were known, it would be possible to find an economical explanation for the development of phenomena which appear on the surface to be independent. This is precisely what is needed, but the traditional approach did not provide the means to accomplish this. Every statistical phenomenon was given a separate psychological "explanation": sentence length was said to grow with memory, verb/noun ratios with "functioning pleasure" *(Funktionslust)*, etc. Probabilistic grammars, applied with insight, can show how such apparently independent phenomena are in fact based on the same structural variable. Developmental language theory should therefore be oriented in this direction. Such an approach would not only be useful for developmental psychology, but also would help to attain explanatory adequacy in linguistic theory (cf. Chapter 1, paragraph 2). The question as to the cause of a universal systematics in natural languages should be traced back partially to the fundamental characteristics of human cognitive structure, and their development in the child. A probabilistic grammar is one of the means by which such fundamental characteristics can be localized, on the basis of the speech of the growing child. Thus statistical methods must not be excluded from theoretical linguistics. Theory and interpretation are interdependent, and interpretation often demands the use of statistical inference.

But what is the basis of the former aversion to statistics in linguistics? Chomsky, repeatedly and with great eloquence, emphasized the unpredictability of human language. Speech is creative, for every utterance is new (with the exception of a few clichés); it shows no simple dependence on the situation in which it is generated. This virtual unlimitedness and freedom of human language is rightly used as an argument against over-simplistic theories of verbal conditioning, such as that of Skinner. But no argument against the statistical investigation of language can be based on these uncontrovertible facts. However this is precisely what Chomsky does. The newness of nearly every linguistic utterance means in statistical terms that every sentence has a probability of occurrence which is indistinguishable from zero. It is on this ground that Chomsky, and with him many other linguists, bans the concept of "probability" from linguistics. It is Suppes' merit to have refuted this argument. He points out that construction of statistical theory is necessary in science precisely where deterministic models are excluded in principle or in fact, and mentions quantum mechanics as the classic example of the impossibility of using a deterministic model. A sentence is precisely as unpredictable as the trajectory of an electron: in both cases the phenomena have a probability which is practically equal to zero. This is the situation in which statistics is applicable *par excellence*. A model is then tested by investigating various statistical parameters in their mutual relations. This holds as much for quantum mechanics as for linguistics. The fact that a sentence has a probability of zero does not mean that the sentence length involved does not occur in the corpus, nor does it mean that words or categories of words in the sentence have a probability of zero. It is on the basis of such data that a model can in fact be tested.

It should be pointed out that the situation here is essentially different from the usual empirical situation in linguistics, which involves the testing of linguistic intuitions. The linguist can question informants at will on their intuitions regarding a linguistic object, and if the phenomenon under study is of any importance, the answers will agree in that regard. But it is not possible, except in trivial circumstances, to make informants spontaneously produce a particular sentence. A sentence cannot be "repeated" like an intuition. It is this circumstance which makes the analysis of a corpus more difficult than work with informants.

As we have seen in Chapter 1, paragraph 2, the analysis of a corpus is one of the forms in which the problem of interpretation occurs in linguistics. Linguists who are concerned with such questions of interpretation must use other methods and types of analysis than those used by theoreticians. But theoretical linguistics is pointless, and ultimately impossible, without interpretation; both aspects of linguistics must develop in interaction. Methodological absolutism in linguistics would be entirely out of place.

HISTORICAL AND BIBLIOGRAPHICAL REMARKS

The distinction between theory and interpretation mentioned in Chapter 1 goes back directly to the work of Bar-Hillel, and indirectly to that of Carnap (cf. Bar-Hillel 1970, 364ff.). The notions of *language* and *observable linguistic phenomena* may be found in de Saussure (1916) as *langue* and *parole*, in Chomsky (in many places, especially Chomsky (1965)) as *competence* and *performance*. These distinctions, however, do not coincide precisely; the distinction between competence and performance in particular has not only the theoretical function emphasized in this volume, but also a psychological function which will be analyzed in Volume III. Literature on the metalinguistic character of linguistic data may be found in Bever (1970 a, b), Levelt and Schils (1971) Levelt (1972), and Watts (1970). The various forms of grammatical adequacy are treated extensively in Chomsky (1965) and in other places by the same author. A detailed treatment of concepts such as "utterance", "word", and "morpheme" may be found in Lyons (1968), to which we refer for further literature on the subject.

Nearly all the essential questions touched upon in Chapter 2 were dealt with by Chomsky before the publication of *Syntactic structures* (1957), in particular in *The Logical Structure of Linguistic Theory* (mimeo, 1955) and in *Three Models for the Description of Language* (1956). The last publications by Chomsky on this subject are those in the *Handbook of Mathematical Psychology* (1963). Our section on context-free grammars borrows some material from Postal (1964b). That article contains some errors, as well as a one-sided treatment of the work of a number of linguists such as Harris and Halliday. Among others, the criticisms by Thorne (1965) and Robinson (1970) are interesting in this connection. Interest in finite automata has received a new impetus in the theory of formal languages, in two forms, (a) natural language parsing programs, based on *augmented transition networks* which are "expanded" finite automata, to be discussed further in Volume III, Chapter 3, paragraph 6.4 (cf. Woods 1970, and Kaplan 1972), and (b) in *tree automata*, which have tree-diagrams for their input and output, instead of terminal strings. These are finite automata, which can nevertheless recognize context-free languages (cf. Thatcher 1967, and Levy and Joshi 1971). There is an interesting future for language parsing programs in both.

The sources for Chapter 3 are Chomsky's *Aspects of the Theory of Syntax* (1965), and a few articles by Peters and Ritchie (1969 a, b, 1971, 1972). *Aspects* gives two different formulations for lexical insertion rules, and we follow the second. The general definition of transformations in Chapter 3, paragraph 2.2 follows Brainerd (1971), who also treats other grammatical systems formally. Chapter 3, paragraph 2.4 follows Peters and Ritchie (1973), a fundamental but extremely laborious formulation. We have tried to extend its readability by introducing the concept of "elementary factorization", and by omitting a few technical details of secondary importance, in particular with regard to definitions of transformational cycle and derivation. There is still no other summary of Peters and Ritchie's formalization of the *Aspects* theory. Later developments (Chapter 3, paragraph 3) originated in work by McCawley (1968 a, b) and by G. Lakoff (1970). The most important sources for the work of interpretative semanticists are Chomsky (1970a, 1971), Jackendoff (1969, 1971). A theoretical survey of generative semantics may be found in Lakoff (1971). This point of view may also be found in Postal (1970, 1971), articles in Bach and Harms (1968), Jacobs and Rosenbaum (1970), Steinberg and Jakobovits (1971), and others. A third trend originating in the *Aspects* theory is the work of Montague (1970, to be published), which was not discussed here. Before his sudden death, Montague had elaborated the formal aspects of his theory in detail. Chapter 3 was written from a formal point of view. There are many introductions to transformational grammar which place more emphasis on content, such as Bach (1964), Lyons (1968), Liles (1971). Two articles by Hall-Partee (1971 a, b) give a good survey of later developments.

The four grammars treated in Chapter 4 come from the following literature. Categorial grammars are found in Leśniewski (1929) and Ajdukiewicz (1935), and related formal systems are treated by Curry (1961) and Lambek (1961). The work of Bar-Hillel, recapitulated in Bar-Hillel (1964), contains the principal background of Chapter 4, paragraph 2; it gave explicit linguistic motivation to the use of categorial grammars. A categorial variant of the base rules in *Aspects*, not discussed here, may be found in Miller (1968). Lewis (1970) treats the semantic component of a categorial grammar. The literature concerning operator grammars is sufficiently indicated in Chapter 4, paragraph 3. The most important source for Harris' work in the field of adjunct grammars, is Harris (1968), where an automaton is also developed to accept such *string languages*. The formal development of transformational adjunct grammars is the result of work by Joshi (1972) and Joshi, Kosaraju and Yamada (1972 a, b). Dependency grammars may be found in Tesnière (1959), Hays (1964), Robinson (1970), Anderson (1971). Articles by the last two authors as well as other important texts on case grammars are found in Abraham (1971). Gaifman (1965) provides a mathematical foundation for dependency grammars. Some material for Chapter 5 was also borrowed from an unpublished survey by Hirschman (1971).

The main point of Chapter 5, the undecidability of an *Aspects* type transformational grammar, was proved at almost the same time by Kimball (1967), Ginsburg and Hall (1969), Salomaa (1971) and Peters and Ritchie (1973). The dates here are misleading. The present writer remembers following a lecture by Ritchie at Harvard University in 1966; notes taken at that lecture show that proof was already given for transformational grammars with a context-sensitive base. Could not more rapid publication of that proof have been of great service to transformational linguistics?

Kimball was decidedly the first to give the proof for transformational grammars with a regular base.

Chapter 6, paragraph 1 is not intended as an introduction to information or communication theory. The most important mathematical source for this is Shannon and Weaver (1949). An excellent introduction is Cherry (1957). Miller (1951) gives more exclusively psycholinguistic applications. Miller and Chomsky (1963) place information theory in the framework of formal languages; the work offers the derivation of the information value of the various approximations of natural language. Adam's language (Chapter 6, paragraph 2) is described in Brown, Cazden and Bellugi (1968); other analyses of Adam's language can be found in McNeill (1970). Suppes' analysis is the only probabilistic approach to the grammar of children's languages available at the moment.

BIBLIOGRAPHY

Abraham, W. (ed.)
1971 *Kasustheorie* (Frankfurt: Athenäum).

Ajdukiewics, K.
1935 "Die sytaktische Konnexität", *Studia Philosophica* 1, 1-27.

Anderson, J. M.
1971 *The Grammar of Case. Toward a Localistic Theory* (Cambridge: Cambridge University Press).

Bach, E.
1964 *An Introduction to Transformational Grammars* (New York: Holt, Rinehart and Winston).

Bach, E. and R. T. Harms (eds.)
1968 *Universals in Linguistic Theory* (New York: Holt, Rinehart and Winston).

Bar-Hillel, Y.
1964 *Language and Information. Selected Essays on Theory and Application* (Reading, Mass.: Addison-Wesley).
1970 *Aspects of Language* (Amsterdam: North Holland).

Bar-Hillel, Y., C. Gaifman, and E. Shamir
1960 "On Categorial and Phrase Structure Grammars", *Bull. Res. Council of Israel* 9F, 1-16.

Bever, T. G.
1970a "The Influence of Speech Performance on Linguistic Structure", in: *Advances in Psycholinguistics*, G. B. Flores d'Arcais and W. J. M. Levelt (eds.) (Amsterdam: North Holland).
1970b "The Cognitive Basis for Linguistic Structures", in: *Cognition and the Development of Language*, J. R. Hayes (ed.) (New York: Wiley).

Brainerd, B.
1971 *Introduction to the Mathematics of Language Study* (New York: Elsevier).

Brown, R., C. Cazden and U. Bellugi
1968 "The Child's Grammar from I to III", in: *Minnesota Symposium on Child Psychology*, J. P. Hill (ed.) (Minneapolis: University of Minnesota Press).

Cherry, C.
1957 *On Human Communication* (Cambridge, Mass.: MIT Press).

Chomsky, N.
1955 *The Logical Structure of Linguistic Theory*. Microfilm, MIT-Library.
1956 "Three Models for the Description of Language", in: *I.R.E. Transactions on Information Theory*, vol. IT-2, Proceedings of the Symposium on Information Theory, Sept., 113-23.
1957 *Syntactic Structures* (The Hague: Mouton).
1963 "Formal Properties of Grammar", in: *Handbook of Mathematical Psychology*, R. D. Luce, R. R. Bush and E. Galanter (eds.) (New York: Wiley).
1965 *Aspects of the Theory of Syntax* (Cambridge, Mass.: MIT Press).
1969 "Quine's Empirical Assumptions", in: *Words and Objections. Essays on the Work of W. V. Quine*, D. Davidson and J. Hintikka (eds.) (Dordrecht: Reidel).
1970a "Remarks on Nominalization", in: *Readings in Transformational Grammar*, R. A. Jacobs and P. S. Rosenbaum (eds.) (Waltham: Ginn).
1971 "Deep Structure, Surface Structure, and Semantic Interpretation", in: *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, D. D. Steinberg and L. A. Jakobovits (eds.) (Cambridge: Cambridge University Press).
1972 "Some Empirical Issues in the Theory of Transformational Grammar". In N. Chomsky, *Studies on Semantics in Generative Grammar*. The Hague: Mouton

Chomsky, N. and M. Halle
1968 *The Sound Pattern of English* (New York: Harper and Row).

Chomsky, N. and G. A. Miller
1963 "Introduction to the Formal Analysis of Natural Languages", in: *Handbook of Mathematical Psychology*, R. D. Luce, R. R. Bush and E. Galanter (eds.) (New York: Wiley).

Curry, H. B.
1961 "Some Logical Aspects of Grammatical Structure", in: *Structure of Language and its Mathematical Aspects. Proc. Twelfth Symp. Appl. Math.*, R. Jakobson (ed) (Providence, R. I.: American Mathematical Society).

Dik, S. C.
1968 *Coordination: Its Implications for the Theory of General Linguistics* (Amsterdam: North Holland).

Fillmore, C. J.
1968 "The Case for Case", in: *Universals in Linguistic Theory*, E. Bach and R. T. Harms (eds.) (New York: Holt, Rinehart and Winston).
1969 "Verbs of Judging: An Exercise in Semantic Description", *Papers in Linguistics* 1, 91-117.

Fodor, J. A.
1970 Three Reasons for Not Deriving "kill" from "cause to die", *Linguistic Inquiry* 1, 429-38.

Gaifman, C.
1965 "Dependency Systems and Phrase Structure Systems", *Information and Control* 8, 304-337.

Geach, P. T.
1970 "A Program for Syntax", *Synthese* 22, 3-18.

Ginsburg, S. and B. Hall-Partee
1969 "A Mathematical Model of Transformational Grammar", *Information and Control* 15, 297-334.

Ginsburg, S. and H. G. Rice
1962 "Two Families of Languages Related to ALGOL", *J. Assoc. Comp. Mach.* 10, 350-71.

Hall-Partee, B.
1971a "Linguistic Metatheory", in: *A Survey of Linguistic Science*, W. O. Dingwall (ed.) (College Park: Linguistics Program/University of Maryland).
1971b "On the Requirement that Transformations Preserve Meaning", in: *Studies in Linguistic Semantics*, C. J. Fillmore and D. T. Langendoen (eds.) (New York: Holt, Rinehart and Winston).

Harman, G.
1970 Deep Structure as Logical Form", *Synthese* 21, 275-97.

Harris, Z. S.
1968 *Mathematical Structures in Language* (New York: Wiley).
1970a "The Two Systems of Grammar: Report and Paraphrase", in *Papers in Structural and Transformational Linguistics*, Z. S. Harris (Dordrecht: Reidel).
1970b *Papers in Structural and Transformational Linguistics* (Dordrecht: Reidel).

Hays, D. G.
1964 "Dependency Theory: A Formalism and Some Observations", *Language* 33, 283-340.

Hirschman, L.
1971 "A Comparison of Formalisms for Transformational Grammar", *Transformations and Discourse Analysis Papers* 87 (University of Pennsylvania).

Hopcroft, J. E. and J. D. Ullman
1969 *Formal Languages and Their Relations to Automata* (Reading, Mass.: Addison-Wesley).

Huddleston, R. D.
1971 *The Sentence in Written English: A Syntactic Study Based on an Analysis of Scientific Texts* (Cambridge: Cambridge University Press).

Hudson, R. A.
1971 *English Complex Sentences: An Introduction to Systemic Grammar* (Amsterdam: North Holland).

Jackendoff, R.
1969 "Some Rules for English Semantic Interpretation", MIT Dissertation.
1971 "Modal Structure in Semantic Representation", *Linguistic Inquiry* 2, 479-514.

Jacobs, A. K. and P. S. Rosenbaum, (eds.)
1970 *Readings in Transformational Grammar* (Waltham: Ginn).

Joshi, A. K.
1971 "How Much Hierarchical Structure is Necessary for Sentence Description?" *Proc. Intern. Symp. Comp. Ling.* (Debrecen, Hungary).
1972 "A Class of Transformational Grammars", in: *Formal Language Analysis*, M. Gross, M. Halle and M. P. Schützenberger (eds.) (The Hague, Mouton).

Joshi, A. K., S. Kosaraju and H. M. Yamada
1972a "String Adjunct Grammars I: Local and Distributed adjunction", *Information and Control* 21, 93-116.
1972b "String Adjunct Grammars II: Equational Representation, Null Symbols, and Linguistic Relevance", *Information and Control* 21, 235-60.

Kaplan, R. M.
1972 "Augmented Transition Networks as Psychological Models of Sentence Comprehension", *Artif. Intell.* 3, 77-100.

Katz, J. J.
1971 "Generative Semantics Is Interpretative Semantics", *Linguistic Inquiry* 2, 313-32.

Katz, J. J. and P. Postal
1964 *An Integrated Theory of Linguistic Descriptions* (Cambridge, Mass.: MIT Press).

Kimball, J. P.
1967 "Predicates Definable over Transformational Derivations by Intersection with Regular Languages", *Information and Control* 11, 177-95.

Kraak, A. and W. G. Klooster
1968 *Syntaxis* (Culemborg: Stam-Kemperman).

Kuroda, S. Y.
1972 "Généralisation de la notion d'équivalence de grammaires — une méthode topologique", in: *Formal Language Analysis*, M. Gross, M. Halle and M. P. Schützenberger (eds.) (The Hague: Mouton).

Lakoff, G.
1970 *Irregularity in Syntax* (New York: Holt, Rinehart and Winston).
1971 "On generative Semantics", in *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics and Psychology*, D. D. Steinberg and L. A. Jakobovits (eds.) (Cambridge: Cambridge University Press).

Lambek, J.
1961 "On the Calculus of Syntactic Types", in: *Structure of Language and Its Mathematical Aspects. Proc. Twelfth Symp. Appl. Math.*, R. Jakobson (ed.) (Providence R. I.: American Mathematical Society).

Leśniewski, S.
1929 "Grundzüge eines neuen Systems der Grundlagen der Mathematik", *Fundamenta Mathematicae* 14, 1-81.

Levelt, W. J. M.

1966 "Generative Grammatika en Psycholinguistiek I. Inleiding in de Generatieve Grammatika", *Nederlands Tijdschrift voor Psychologie* 21, 317-37.

1972 "Some Psychological Aspects of Linguistic Data", *Linguistische Berichte* 17, 18-30.

Levelt, W. J. M. and E. Schils

1971 "Relaties tussen psychologie en linguistiek", in: *De plaats van de psychologie temidden van de wetenschap*, Annalen van het Thijmgenootschap (Bussum, Holland: Paul Brand).

Levy, L. S. and A. K. Joshi

1971 "Some Results in Three Automata", *Proc. ACM Symposium on Theory of Computing* (Cleveland). *Mathematical Systems Theory* 6 (1973), 334-42.

Lewis, D.

1970 "General Semantics", *Synthese* 22, 18-67.

Liles, B. L.

1971 *An Introductory Transformational Grammar* (Englewood Cliffs, N. J.: Prentice-Hall).

Lyons, J.

1968 *Introduction to Theoretical Linguistics* (Cambridge: Cambridge University Press).

Markov, A. A.

1913 *Essai d'une recherche statistique sur le texte du roman "Eugène Onégin"* (= *Bulletin de l'Académie Impériale des Sciences*, St. Petersbourg, 7).

McCawley, J. D.

1968a "The Role of Semantics in Grammar", in: *Universals in Linguistic Theory*, E. Bach and R. T. Harms (eds.) (New York: Holt, Rinehart and Winston).

1968b "Concerning the Base Component of a Transformational Grammar", *Foundations of Language* 4, 55-81.

McNeill, D.

1970 *The Acquisition of Language. The Study of Developmental Psycholinguistics* (New York: Harper and Row).

Miller, G. A.

1951 *Language and Communication* (New York: McGraw-Hill).

1968 "Algebraic Models in Psycholinguistics", in: *Algebraic Models in Psychology*, C. A. J. Vlek (ed.) (The Hague, NUFFIC).

Miller, G. A. and N. Chomsky

1963 "Finitary Models of Language Users", in *Handbook of Mathematical Psychology*, R. D. Luce, R. R. Bush and E. Galanter (eds.) (New York: Wiley).

Miller, G. A. and J. A. Selfridge

1950 "Verbal Context and the Recall of Meaningful Material", *American Journal of Psychology* 63, 176-85.

Montague, R.

1970 "English as a Formal Language", in: *Linguaggi nella società e nella tecnica* (Milan: Edizioni di Comunità), 189-224.

to be published.

"The Proper Treatment of Quantification in English", *Proceedings of the* 1970 *Stanford Workshop on Grammar and Semantics*, J. Moravscik and P. Suppes (eds.) (Dordrecht: Reidel).

Peters, P. S.

1966 "A Note on the Equivalence of Ordered and Unordered Grammars", *Harvard Computation Laboratory Report of NSF* 17.

Peters, P. S. and R. W. Ritchie

1969a "A Note on the Universal Base Hypothesis", *Journal of Linguistics* 5, 150-52.

1969b "Context-sensitive Immediate Constituent Analysis — Context-free Languages Revisited", *Proceedings of the Ninth ACM Symposium on Theory of Computing*, 1-8. Also in *Mathematical Systems Theory* 6, 1973, 324-333.

1971 "On Restricting the Base Component of Transformational Grammars", *Information and Control* 18, 483-501.

1973 "On the Generative Power of Transformational Grammars", *Information Sciences* 6, 49-83.

Postal, P.

1964a "Limitations of Phrase Structure Grammars", in: *The Structure of Language*, J. A. Fodor and J. J. Katz (eds.) (Englewood Cliffs, N. J.: Prentice-Hall).

1964b *Constituent Structure: A Study of Contemporary Models of Syntactic Description* (Bloomington, Ind.: Indiana University Press).

1970 "On the Surface Verb 'remind'", *Linguistic Inquiry* 1, 37-120.

1972 "A Global Constraint on Pronominalization", *Linguistic Inquiry* 3, 35-60.

Reich, P. A.

1969 "The Finiteness of Natural Language", *Language* 45, 831-43.

Robinson, J. J.

1969 "Case, Category, and Configuration", *Journal of Linguistics* 6, 57-80.

1970 "Dependency Structures and Transformational Rules", *Language* 46, 259-85.

Ross, J. R.

1967 "Constraints on Variables in Syntax" (Cambridge, Mass.: MIT doctoral dissertation).

Sager, N.

1967 "Syntactic Analysis of Natural Language", in: *Advances in Computers* 8, F. Alt and M. Rubinoff (eds.) (New York: Academic Press).

Salomaa, A.

1971 "The Generative Capacity of Transformational Grammars of Ginsburg and Partee", *Information and Control* 18, 227-32.

Šaumjan, S. K. and Soboleva, P. A.

1963 *Applikativnaja poroždajuščaja model' i isčislenie transformacij*

v russkom jazyke (Moscow, Izdatel'stvo Akademii Nauk SSSR).
Saussure, F. de
1916 *Cours de linguistique générale* (Paris).
Schultink, H.
1967 "Transformationeel-generatieve taalbeschrijving", *De Nieuwe Taalgids* 60, 238-57.
Seuren, P. A. M.
1969 *Operators and Nucleus: A Contribution to the Theory of Grammar* (Cambridge: Cambridge University Press).
Shannon, C. and W. Weaver
1949 *The Mathematical Theory of Communication* (Urbana, Ill.: University of Illinois Press).
Steinberg, D. D. and L. A. Jakobovits
1971 *Semantics. An Interdisciplinary Reader in Philosophy, Linguistics and Psychology* (Cambridge: Cambridge University Press).
Suppes, P.
1970 "Probabilistic Grammars for Natural Languages", *Synthese* 22, 95-116.
1971 "Semantics of Context-free Fragments of Natural Languages", *Technical Report* 171, *Inst. Math. Stud. Soc. Sc.* (Stanford University).
Tesnière, L.
1959 *Eléments de syntaxe structurale* (Paris: Klincksieck).
Thatcher, J. W.
1967 "Characterizing Derivation Trees of Context-free Grammars through a Generalization of Finite Automata Theory", *Journal of Computer and Systems Sciences*, 317-22.
Thorne, J. P.
Review of P. Postal's *Constituent Structure: A Study of Contemporary Models of Syntactic Description.*
Watts, W. C.
1970 "On Two Hypotheses Concerning Psycholinguistics", in *Cognition and the Development of Language*, J. R. Hayes (ed.) (New York: Wiley).
Woods, W. A.
1970 "Transition Network Grammars for Natural Language Analysis", *Communications ACM* 13, 591-606.
Wundt, W.
1900 "Volkerpsychologie I und II", *Die Sprache* (Leipzig).

AUTHOR INDEX

SUBJECT INDEX

(Italicized numbers refer to definitions)