W.J.M. LEVELT (*)

FORMAL GRAMMARS AND THE NATURAL LANGUAGE USER : A REVIEW

(*) Psychologisch Laboratorium Katholieke Universiteit Nijemegen

1. INTRODUCTION:

General review of relations between formal grammar theory, natural linguistics and psycholinguistics.

1.1 Origin and basic problems of formal grammar theory

This chapter is introductory to the following three. Its aim is to give an historical outline of the mutual inspiration that we have seen in the last fifteen or so years between formal grammar theory, natural language theory and psycholinguistics. In the following three chapters we will discuss some recent characteristic examples of this interaction.

Fifteen to twenty years are long enough to have almost forgotten how formal grammar theory came into existence. The origin of this theory comes from the study of natural language. A description of natural language is traditionally called a grammar. It specifies construction of sentences, relations between linguistic units, etc.. Formal grammar theory started from the need to give a formal mathematical basis for such descriptions. Initially the creation of these new formal systems was largely the work of Noam Chomsky. His aim was not so much to refine linguistic descriptions, but to construct a formal basis for the discussion of the foundations of linguistics. "What should be the form of a linguistic theory?", "What sort of problems can be expressed by way of different formal means, and what do we take to be a solution?": these were the main issues to be tackled. In short, formal grammars were developed as mathematical models for linguistic structure.

The first developments only concerned the syntax of natural languages, not their semantics. The most successful application of formal grammar theory have been up to now in the area of syntax. All our discussion will therefore be largely limited to syntactic issues.

The first and most obvious use of formal grammar theory in linguistics was to create a variety of more or less restrictive grammars, and to compare their generative power to the empirical requirements of linguistic data. Let us

call this the problem of "generative power". In this chapter we will discuss and criticize some historical highlights in the approach to this problem. In the next chapter some important recent results will be discussed.

The explicit use of formal grammars in linguistics also created a more general and more philosophical problem. It is one thing to formalize a linguistic theory, it is quite another thing to formulate the relation between such a theory on the one hand, and empirical linguistic data on the other hand. The problem here consists in clarifying what, exactly, is the empirical domain of the linguistic theory, and what is the empirical interpretation of the elements and relations that figure in the theory. We will call this problem the interpretation problem, after Bar-Hillel. In this chapter we will only make some general points relating to this issue. The third chapter, however, will be devoted to a formal psycholinguistic analysis of the interpretation problem.

The linguistic origin of formal grammar theory, finally, also led to the early development of theories of grammatical inference. There were two reasons for this. Firstly, a main theme in structural linguistics had for a long time been the development of so-called "discovery procedures", i.e. methods to detect structures in linguistic data. Secondly, probably under the influence of the psychologist George Miller, Chomsky had realized the fundamental problem of language acquisition. The description of a language is one thing, but the causation of linguistic structures is another more fundamental issue. Only a solution of this latter problem will give linguistic theory an explanatory dimension. Efforts to write formal systems which are able to infer a grammar from a data corpus can be found as early as 1957. Since then, inference theory has had a considerable development. In the last chapter we will be concerned with some relations between recent inference theory and psycholinguistic models of language acquisition.

1.2 Observational adequacy of regular and context-free grammars.

Let us now return to the early developments of formal

grammar theory. We will first very quickly review the variety of grammars that Chomsky developed in the second half of the fifties. Then we will discuss some problems relating to the linguistic adequacy of regular and contex-free grammars.

According to Chomsky, a grammar is defined as a system

$$G = \langle V_N, V_T, P, S \rangle ,$$

where V_N is a finite nonempty set, the nonterminal vocabulary, whose elements are called category symbols or auxiliary variables;

V_T is a finite nonempty set, the terminal vocabulary whose elements are usually called "words" or "morphemes";

S is an element of V_N (the start symbol).

Given a set E of symbols, we denote by E^* the set of all strings of finite length which can be obtained by concatenation of symbols in E; by E^+ we shall denote the set $E^* - \{\lambda\}$, where λ is the null string (of zero length).

Now P (the set of production rules of the grammar) is a finite set of rules of the form $\alpha \longrightarrow \beta$, with $\alpha \in V^+$ and $\beta \in V^*$, where $V = V_T \cup V_N$.

We shall say that a string $\gamma \in V^+$ directly produces a string $\delta \in V^*$ (in symbols $\gamma \Rightarrow \delta$) if $\gamma = \varphi\eta\psi$, $\delta = \varphi\theta\psi$, for some $\eta, \theta, \varphi, \psi \in V^*$, and $\eta \rightarrow \theta$ is in P. Finally, we say that $\gamma \in V^+$ derives (directly or not) a string $\delta \in V^*$ (in symbols $\gamma \overset{*}{\Rightarrow} \delta$) if either $\gamma = \delta$, or there exist strings $\gamma_0, \gamma_1, \ldots, \gamma_n$, for some finite, n, such that

$$\gamma_0 = \gamma, \quad \gamma_i \Rightarrow \gamma_{i+1}, \text{ for } i = 0, \ldots, n-1,$$

and $\gamma_n = \delta$.

Now the language L_G generated by a grammar G as above is defined as the set

$$L_G = \{\alpha \mid \alpha \in V_T^*, S \overset{*}{\Rightarrow} \alpha\}$$

The variety of grammars that Chomsky defined came about by putting more and more restrictive conditions on the format of production rules.

These are:

(0) no restriction: type 0 grammars.

(1) for all rules $\alpha \longrightarrow \beta$ of P, the length of β should be not less than the length of α : context-sensitive grammars (type 1)

(2) for all rules $\alpha \longrightarrow \beta$ of P, we must have $\alpha \in V_N, \beta \neq \Lambda$: context-free grammars (type 2).

(3) for all rules $\alpha \longrightarrow \beta$ of P, we must have $\alpha \in V_N$, and either $\beta \in V_T$, or β equal to the concatenation of an element of V_T and one of V_N, in that order: regular grammars (type 3).

A language is called type-i if it can be generated by a type-i grammar.

There is a strict inclusion relation among the classes of languages defined above: if C_i is the class of languages of type i, then $C_{i+1} \subsetneq C_i$. In particular there are not regular (i.e. type 3). These are exactly the languages that are called "self-embedding". A context-free language is self-embedding if all grammars generating it are self-embedding. A context-free grammar is self-embedding if there is a $B \in V_N$ such that $B \Longrightarrow \alpha B \gamma$, where α and γ are non-empty strings.

Chomsky (1956, 1957) rejected regular languages as adequate models for natural languages. The argument used by Chomsky to conclude that natural languages are at least non-regular had an enormous influence on the development of modern linguistics; this justifies a rather detailed discussion of it. It is also the case that the argumentation, as given in Syntactic Structures (1957), 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 erroneous

conclusions have been drawn. 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.

For property X self-embedding is taken. Then step (a) in the argument is correct. The problem, however, resides in (b). One must now show for (b) that e.g. English is a self-embedding language. This is done by referring to self-embedding subsets of English, such as

- the rat ate the malt
- the rat the cat killed ate the malt
- the rat the cat the dog chased killed ate the malt, and so on.

It would not be difficult to think of other examples.

Chomsky, in Syntactic Structures, gives this as evidence that English is self-embedding, and therefore is not a regular language. The self-embedding property of English is however, not demonstrated by the examples above, in spite of appearance of the contrary. The only thing which has been proved 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 non-empty strings over a given alphabet V_T, i.e. L_T^+. A grammar for L is $G = \langle V_N, V_T, P, S \rangle$, where $V_N = \{S\}$ and P contains $S \rightarrow a$, $S \rightarrow a\,S$ for all a in V_T. This is clearly a regular grammar. Since L contains all strings over V_T of positive lengths, it also contains all self-embedding languages over V_T. In conclusion, from the existence of self-embedding subsets it does not follow that a language is self-embedding.

Chomsky's original argumentation (1956), in the technical paper which preceded Syntactic Structures, is considera

bly more precise. There he explaned that it is not only necessary to show that the language containes self-embedding subsets, but also that a particular change in the sentences of a self-sembedding subset must always he accompanied by a certain other change, on pain of ungrammaticality.

Let us clarify this by a simple example. Take the construct if s_1 then s_2 in English. There is a self-embedding subset of English of the form

$$(*) \quad \{\text{if}^n\ s_1\ (\text{then}\ \ s_2)^n,\ \ n = 1, \ldots\}$$

In order to show that English is self-embedding, according to Chomsky, one has not only to show that all strings in the subset above are grammatical (i.e. are good, though awkward, English), but also that

$$(* *) \quad \text{if}^n\ \ s_1\ (\text{then}\ s_2)^m$$

is ungrammatical for all cases with $n \neq m$. This reasoning is correct. The interesting thing is, however, that Chomsky in the article quoted (1956) only shows the existence of self-embedding constructs of the form (*), and does not give data to support that all constructs of the form (* *) are ungrammatical. In fact, one might say that the latter condition does not hold at all, since grammatical examples of the form (* *) are

if John sleeps, he snores
John drank coffee, then he left

and so on.

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).

It has been proved by Bar-Hillel (see Hopcroft and Ullman, 1969) that the intersection of two regular languages is regular. So, if L is a language, T is a regular language, and $T \cap L$ is non-regular, then L is non-regular. Assume for L the English language, and construct T as follows:

$$T = \{\text{William (whom William)}^n \text{ succeeded}^m \text{ succeeded William} \mid n, m \geq 1\}.$$

This is a regular language, because it can be generated by the following grammar

$$G = \langle V_N, V_T, P, S \rangle$$

where $V_N = \{S, A, B, C\}$, $V_T = \{\text{William, succeeded, whom}\}$

$$P = \begin{cases} S \longrightarrow \text{William } A \\ A \longrightarrow \text{whom William } A \\ A \longrightarrow \text{whom William } B \\ B \longrightarrow \text{succeeded } B \\ B \longrightarrow \text{succeeded } C \\ C \longrightarrow \text{succeeded William} \end{cases}$$

G is so-called "right-linear" grammar: such grammars generate regular languages (see Hopcroft and Ullman, 1969).

Let us now 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 succeed so that the grammatical sentences in T are those for which $n \geq m$. In both cases ($n = m$, $n \geq m$), however, the intersection is self-embedding. Hence English is not 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. We cannot expect more evidence than such weak intuitions. 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, 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.

We have discussed at some length the problem of adequacy of regular languages, because a next step in lingui

stics has been to examine the observational adequacy of context-free languages. Postal (1964) "proves" the theorem (his term) that the North American Indian language Mohawk is not context-free, by following the argumentation schema of Syntactic Structures, i.e.:

(a) A language with property x is not context-free;
(b) Mohawk has property x
(c) Then Mohawk is not context-free.

As property x he takes the property of "string repetition", as in the language $\{W\,W\}$, where every sentence consists of a string followed by its repetition. Then (a) is true.

Postal then shows the existence of string-repetition phenomena in Mohawk, i.e. sentences of the form

$$a_1\ a_2\ \ldots\ a_n\ b_1\ b_2\ \ldots\ b_n$$

where a_i "corresponds" to b_i. From this, he concludes that Mohawk is not context-free. 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 that a language is not context-free from the existence of non-context-free subsets.

Again a more convincing proof can be carried out along different lines (Brandt Corstius, personal communication). It has been proven by Bar-Hillel (see Hopcroft and Ulhman, 1969) that the intersection of a regular language and a context-free language is context-free. So, if L is a language, T is a regular language, and $T \cap L$ is non-context-free, then L is non-context-free. Assume for L the English language, and construct T as follows:

$T = \{$ The academics, accountants, actors, admirals, ...,
in respectively Belgium, Bulgaria, Burundi, Brasil, ...,
are respectively calm, candid, canny, careless, ...$\}$

or abbreviated

$$T = \{\text{The } a^k, \text{ in respectively } b^m, \text{ are respectively } c^n \mid K, m, n \geq 0\}.$$

It is not difficult to write a regular grammar for T. Let

us now consider which sentences in T are grammatical English sentences. Intuitively these are the strings for which $K = m = n \geq 1$. However it is known (see Hopcroft and Ullman, 1969) that there is no possible context-free grammar for the language

$$\{a^n b^n c^n \quad n \quad 1\},$$

i.e. the intersection of T and English is not context-free. We therefore conclude that English is a non-context-free language. Again, this "proof" is as strong as the intuitions about the grammatical subset of T, which in this case are particularly weak. Much more convincing, at any rate, are other arguments against the context-free character of natural language. They are not based, however, on the above considerations about (weak) generative power, but on the less well defined notion of strong generative power.

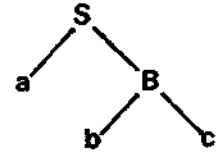
For a context-free grammar $G = V_N, V_T, P, S$ we define its strong generative power as the set of terminal leftmost derivations it generates, i.e. the set of derivations of the form

$$S \Longrightarrow \gamma_0 \Longrightarrow \gamma_1 \quad \cdots\cdots \Longrightarrow \gamma_{n-1} \Longrightarrow \gamma_n$$

where $\gamma_n \in V_T^+$, and $\gamma_i = \varphi A \psi$, $\gamma_{i+1} = \varphi \eta \psi$,

$A \longrightarrow \eta$ is in P, and $\varphi \in V_T^*$, for all $i = 0, 1, \ldots n-1$.

We may associate to each terminal leftmost derivation a labelled graph (also called "Phrase-marker"); for example to $S \Longrightarrow a\ B \Longrightarrow abc$ we associate the following graph



Each terminal leftmost derivation is a structured description of the terminal string or sentence it produces. The linguistic question, then, is whether a particular grammar can express in a satisfactory way what we feel the structu

re of a sentence is. Consideration about syntactic and semantic ambiguity in natural language often require that a sentence has two different structural descriptions. In some of these cases (e.g. for a sentence such as "Italian like opera as much as Germans") a context-free grammar cannot provide two different leftmost derivations which intuitively, correspond to the two readings of the sentence.

Moreover, also intuitive relations between different sentences (e.g. active and passive form of a statement) are often not directly expressible by means of type 2 and type 1 grammars. To express such relation linguists felt an increasing need for the possibility to simultaneously assign more than one phrase-marker to a sentence. These and many other similar problems led to abdication of the traditional context-free model, and for similar reasons of the context-sensitive model as well.

The next step in the Chomsky hierarchy is type-0 grammars. But these are equivalent to Turing machines, and there are good reasons not to give grammars such maximum power. This will be discussed in the next chapter.

1.3 Origins of the psycholinguistic approach.

Let us now, to conclude this general introduction, switch to psycholinguistics.

We have seen that the arguments in favour or against a certain variety of grammar were based on insights such as the grammaticality of strings, or the "fittingness" of a structural description. But when is a certain string of words "grammatical", and how "fitting" is a structural description? Clearly, these are linguistic intuitions, and Chomsky did not hesitate to state that linguistics is concerned with linguistic intuitions. These form the empirical domain in Chomskian linguistics. Not all linguists accept this view, but there are reasons to support it.

Two major problems, however, arise:

(1) Can we make explicit the relations between the formal linguistic theory on the one hand and linguistic judgements, i.e. expressions of linguistic intuitions on the ot

her hand? This will be the topic of the third chapter.

(2) What is the relation between a such defined grammar and models of the language user (speaker, listener)?

Initially, Chomsky, Miller and others conceived of this relation as follows. Intuitions, they said, express the (tacit) knowledge of speakers about their language; this knowledge, which they called linguistic competence, is at the basis of all actual language behavior or performance. So, if we have only determined the structure of linguistic competence, we can proceed to the study of performance, in which the competence plays a general and essential role. Of ccurse in a theory of performance additional psychological factors come in, such as motivation, memory span, and so on. It is their interaction with linguistic competence which is to be studied by psycholinguists. In our view this distinction between competence and performance is far-fetched, if not fully untenable. The data for competence research are linguistic judgements, which are forms of language behavior. It is not clear why just this type of language behavior (linguistic judgement) should have the privilege of leading to a theory, which has then to be built into the models for various other types of language bahavior, such as speaking or listening. In fact, the latter forms are much more direct or "primary" forms of language use, whereas linguistic judgement is a very secondary or derived form of language behavior.

Though this approach could not stand the test of time, it did originally stimulate much research in psycholinguistics. In fact between 1963 and 1967 at Harvard and MIT a number of psychologists and linguists (among which the present author) tried to show that the competence or knowledge as described by the linguists in their grammars, is "psychologically real", i.e. could be shown to operate in sentence understanding, in memorisation and speech. Aspects of the formal grammar, such as different types of rewrite rules, transformations, and so on, were tested for their psychological relevance in experiment upon experiment.

Let us consider one or two examples of the subjects of these early developments.

The correspondence between the various types of gram-

mars and automata led to considering various automata as models for the language user. In spite of the obvious finiteness of the human brain, the finite automaton was quickly dismissed as a model for the language user, as the regular grammar had been discarded as a linguistic model. More inspiring was the push-down automaton, which is in its nondeterministic form equivalent to a context-free grammar. The self-embedding property obviously attracted much attention; it was interesting to see if human finiteness would be reflected in a limited push-down capacity.

Severe limitations on the understanding of self-embedding were clearly demonstrated. One or two embeddings turned out to be disastrous for comprehension, as in the following examples:

(1) if if John comes Peter comes Charles comes.

(2) The dog the cat the mouse bit chased ate a lot.

Moreover, if limited push-down capacity (not lack of knowledge!) explained this, it should be equally hard to handle other types of embedding. But this turned out not to be the case, as it can be seen in the following examples:

(3) John, who saw everything, will tell it.

(4) John, who saw everybody you mentioned, will tell it.

So self-embedding seems to exhibit a special situation. It seems to be especially hard for the language user to interrupt a procedure by the same type of procedure.

In spite of this, the push-down automaton model is still of some use in psycholinguistics. Masters (1970), for instance, has studied in this way the language of schizophrenics. From the literature on schizophrenic language it was known that these patients use (1) less different words, (2) less adjectives, (3) shorter sentences, (4) more incomplete sentences, (5) more adjectives per verb, (6) more objects per subject, (7) less modifiers per verb, and so on. Masters wrote a context-free grammar of English and casted it in the format of a push-down automaton. By limiting the size of the push-down store to less there 6 elements it turned out that the language generated (or accepted) by it

showed all the mentioned seven characteristics of schizophrenic language. However, the main interest of the model has gone to the study of transformations, i.e. how do people cope with passive sentences, negative sentences, question sentences and so on. A review of this work can be found in Levelt (1974).

These early developments of psycholinguistics eventually led to very little, and faded away. It was mainly due to extraneous developments, especially in Artificial Intelligence, that a new approach in psycholinguistic theory evolved. Computer scientists and linguists tried to develop programs for understanding and producing natural language. Thome's work in Edinburgh was a first big step. Others followed, in particular Sager, Woods and Vinograd.

At the basis of these programs is a structure called augmented transition network. In its simplest form it is a finite automaton expanded with a push-down memory. In a more sophisticated form all sorts of conditions on transitions can be specified, thus obtaining the power of a Turing machine. It is possible to write a transformational grammar in such terms. In this way the grammar is no more an abstract body of knowledge, which may or may not be "consulted" by the hearer or speaker, but it is, in a sense, the accepting (or generating) mechanism itself.

2. THE GENERATIVE POWER OF TRANSFORMATIONAL GRAMMARS.

In the first chapter we discussed how linguists felt the urge to move up in the hierarchy. Initially they tried to show that the more restrictive forms of grammar such as regular grammars and context free grammars could not be sufficient to generate all and only the sentences of a natural language. The main argument, however, to shift to more complicated grammars was the lack of descriptive adequacy of grammars up to the level of context-sensitive.

The next step, therefore, was to move to certain types of strictly type zero grammars. They were called transformational grammars for reasons that we will discuss presently.

Beforehand, we must make one or two remarks in order to show that this move is not without problems, and that certain precautions have to be taken. For this we have to consider again the fundamental aims of a linguistic theory. We mention two of them:

(1) A linguistic theory should be descriptive for the linguistic intuition of a native speaker. One intuition concerns grammaticality. Native speakers can recognize sentences from the language as being elements of the language. But at the same time, they can equally well recognize non-sentences as not belonging to the language. In terms of grammars this would mean that native speakers have the disposal of a decision procedure. For any string x in V_T^* they can in a finite time decide whether $x \in L$ or $x \notin L$. A linguistic grammar, therefore, should be recursive, not only recursively enumerable.

(2) A linguistic theory should be explanatory in the sense that it can explain how the grammar is caused. In formal terms: the grammar should be such that it is learnable-in-principle, i.e. there should be a conceivable inference procedure for the grammar. In the last chapter we will show that this requirement comes down to the condition that the grammar is primitive recursive.

Since the difference between recursive and primitive recursive is small, and has no linguistic interpretation we will conclude from these two aims that any grammar for a natural language should be decidable or recursive.

2.1 The structure of Chomsky's transformational grammar.

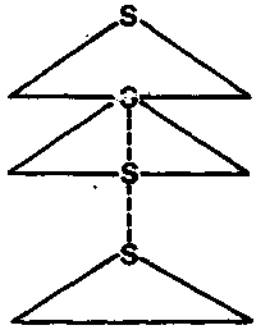
Various transformational grammars have been developed. Most influential has been (and is) Chomsky's formulation "Aspects of the Theory of Syntax" (1965), but there are interesting other examples such as Joshi's et al (1972) string adjunct grammar, and dependency grammars. The present discussion has to limit to Chomsky's model.

Chomsky wanted, on the one hand, to keep the various advantages of phrase structure grammars, such as Context Free Grammars and certain Context Sensitive Grammars (parsing, etc.) and at the same time expand the descriptive potentialities of the grammar. Necessary expansions, as we have seen in the first part, are required for generating more then one tree diagram or phrase marker per sentence in order to take account of certain ambiguities, deletions and relations between sentences. In all cases it is necessary to define relations between tree graphs or P-markers. These relations are called transformations. In principle a transformation maps a tree graph on a tree graph. It is a rule with tree graphs as input and output.

The rough structure of Chomsky's "Aspects"-grammar, then, is a context-sensitive grammar generating terminal tree graphs, which are called base structures. These base structures form the input for the transformational rules. For some of them these rules generate an output which is called a surface structure. Its terminal string is the sentence. Base structures transformationally leading to surface structures are called deep structures. All other base structures are said to have been filtered out.

The context sensitive base grammar generates an infinite set of strings. It is constructed in such a way that recursion can only take place through the recursive initial

symbol S. Recursive rewriting of S leads to base P-markers of the form:



Subsentence 1

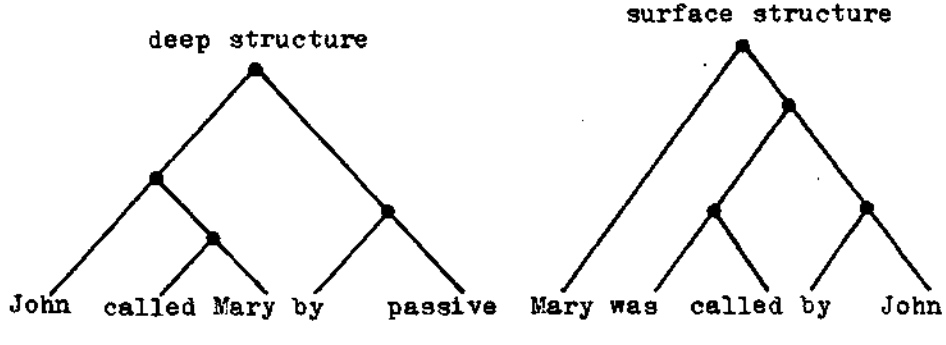
Subsentence 2

n

Were each triangle represents a subtree resulting from rewriting S up till recursion of S.

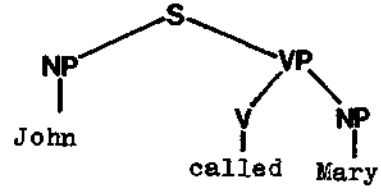
The transformation rules operate on such base structures in a special fashion. Transformations form an ordered list, they are tried out one by one, starting at the top of the list and ending at the bottom. This cycle is first applied to the most deeply embedded subsentence (n), then it turns to the next higher one (n - 1), a.s.o. until the top sentence (1) has been reached. (Additionally it seems necessary to assume the existence of some pre- and post-cyclic rules). If there is an output, it is called a surface structure. Its terminal string is called a sentence.

The structural description of a sentence is the pair of tree graphs consisting of deep and surface structure. So, for instance for the sentence Mary was called by John we have, in simplified form (node labels omitted) the pair:



The mediating transformation here is called the passive transformation. The pair of structures nicely expresses the relation between the sentence John called Mary and the passive Mary was called by John. The first, active, sentence is already more or less present in the deep structure of the passive sentence.

It should be obvious that transformations are type-zero rules. This is easy to see by writing trees as strings, namely as labelled bracketings. Let us take as an example John called Mary, which has as deep structure:



by rules
$$\begin{cases} S \rightarrow NP + VP \\ VP \rightarrow V + NP \\ NP \rightarrow \text{John, Mary} \\ V \rightarrow \text{called} \end{cases}$$

This can alternatively be written as:

$$({}_{S}\ ({}_{NP}\ \text{John}\ )_{NP}\ ({}_{VP}\ ({}_{V}\ \text{called}\ )_{V}\ ({}_{NP}\ \text{Mary}\ )_{NP}\ )_{VP}\ )_{S}$$

Transformations are rewritings of such strings. In fact, it is easy to replace the base grammar by a grammar which generates such labelled bracketings. Namely in the following way:

$$S \rightarrow ({}_{S} + NP + VP +)_{S}$$
$$VP \rightarrow ({}_{VP} + V + NP +)_{VP}$$
etc.

In this general form, transformations can replace any such string by any other string. This is obviously not very interesting. It is, actually, necessary to put a severe limit on transformations. In 'Aspects" Chomsky limits transformations to operations that either add a factor (substring) to a labelled bracketing, replace a factor, or delete one.

It should be immediately obvious that the latter two operations, which are essentially erasure operations can be

strict type-zero rules, because they shorten a given string. It is at this point that our above mentioned caution is required. What sort of condition has to be put on transformations, in order to keep decidability? Or to state it differently: what sort of condition on transformations is required in order that a Turing Machine given a string can decide whether the string can or cannot be generated by the Transformational Grammar?

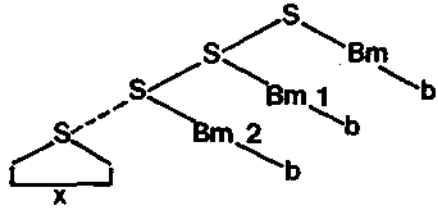
Chomsky was not very explicit about this problem. He formulated a condition, which he called the principle of recoverability of deletions. In essence, the condition states that given the string, and given the transformation from which it emerged, there should be only a finite set of strings that could have been the input to the transformation. This was secured by requiring that either (a) the deleted substring would after transformation still be present at some other place in the string, (b) the deleted string would be one of a finite set, i.e. the condition would specify the finite set for the transformations. An example of the former would be the derivation of John and Mary chased the dog from John chased the dog and Mary chased the dog by a coordination transformation. The first chased the dog is deleted, but this substring is still present at another place in the string. An example of the latter is the derivation of the imperative shut the door from you shut the door. The imperative transformation only allows for deletion of the element you (which is certainly a finite set).

It has been proven by Peters & Ritchie that this condition fails to preserve the recursiveness of the grammar. In fact they proved that transformational grammars of this sort are equivalent with type-zero grammars. They generate all and only the type-zero languages.

A rough outline of the proof is as follows: it has two parts. The first is to prove that every transformational grammar is type-zero. This we have more or less seen. The more interesting part is the converse, namely that every type-zero language can be generated by a transformational grammar of this type. It consists of 3 steps:

(1) Be L type-zero and G a grammar for L. We first construct a context-sensitive grammar G' which generates strings xb^m for all x is L. It is derived from G by changing the shortening rules of G by adding a row of b's.

(2) Context-sensitive G' is equivalent to G" in Kuroda's normal form. Kuroda proved the theorem that context-sensitive grammars are equivalent to grammars with the limited rule format ($S \rightarrow SB$, $CD \rightarrow EF$, $G \rightarrow H$, $A \rightarrow a$). This is a format in which for generating $x = a_1 \ldots a_n$. One has first to apply $S \rightarrow SB$ n - 1 times giving $SB_1 B_2 \ldots B_{n-1}$, which has the length of x, and then to replace or interchange the elements by applying further rules. In the Kuroda form, therefore, the string xb^m has a structure such a



(3) We now create a transformational component by which the b's can be erased, in order to leave us with the set of strings $\{x\}$, which form the type-zero grammar L. This can easily be done. A single transformation suffices. It is constructed in such a way that it applies to a (sub-) sentence which is factorizable in two substrings, the last of which is a b. The transformation consists of erasing the b, i.e. the second factor of the sub-sentence. The nice thing about using Kuroda's normal form is that in this way indeed all b's get erased. It should be remembered that the transformations are first applied to the most deeply embedded subsentence. The b is erased and the next cycle starts. In this way all b's are erased because each b is a second and last factor of an S. By this procedure our TG can generate type-zero lan-

guage L. Does the transformation conform to the principle of recoverability of deletions? Yes, bacause b is the only element that can be deleted. Thus, given the output string and the transformation, the input string can always be recontructed.

This equivalence of TG's and type zerogrammars shows that they are not, in general, recursive as we had required. Such grammars, therefore, are unfit for linguistic description or explanation.

It should be clear what makes the grammar undecidable for a Turing Machine. Given a string and the transformation it is possible to reconstruct the previous string, but the problem is that the TM does not know the transformation and worse, whether a transformation was applied at all. There is, given a string x of length $|x|$ no upper bound on the size of the deep structure for x. It therefore requires an infinite set of operations to test for all possible deep structures for x whether they are generated by the base grammar.

One could be inclined to ascribe this state of affairs to the combination of the apparently not adequate principle of recoverability of deletions and the string (i.e. context free) base grammar. This is only partly correct. In a further paper Peters & Ritchie proved that even a regular grammar as base grammar was sufficient for the generation of all type-zero languages. In fact this could be a highly trivial grammar, namely :

1. $S \rightarrow S\#$
2. $S \rightarrow a_1 \ldots a_n b \#$, where $\{a_1, \ldots a_n\}$ is the terminal vocabulary of the language. It generates strings of the form $a_1 \ldots a_n b \# \; n$.

In order to proof this Peters and Ritchie had to make use of the filter-function of transformations which we mentioned above. So, one also has to repair the filter-mechanism.

A final objection one could make is that such trivial grammars can never be descriptively adequate. But even at

this point Peters and Ritchie were able to show that these grammars are able to account for grammaticality, ambiguity and paraphrase, i.e. for the most important structural intuitions.

In conclusion, the Chomskian transformational grammar severely fails both descriptively (it cannot account for ungrammaticality intuitions), and explanatory. It is not learnable and also it fails in handling the universal base problem. The idea of the latter is that what is universal to language is the base grammar. Languages would mainly differ with respect to their transformational structure. In Chomsk's formalism this is trivially true. The above trivial grammar can be used to generate any language. The universal base hypothesis is not any more an empirical issue.

At the present moment these problems of generative power have not yet been solved. There is only one other completely formalized transformational grammar, namely Joshi's mixed adjunct grammar. It goes back to the adjunction grammar of Harris. It is nicely recursive and it seems attractive to apply some of Joshi's notions to the Chomskian grammar. One is the so-called trace-condition. It says that each transformation leaves a trace, i.e. an element or string which cannot be erased by any further transformation. In this way for a given string x, there is an upper bound on the number of transformations which can have been applied in its derivation. A Turing Machine, therefore, has only to retrace a finite set of derivations, i.e. there exists a decision procedure. It has to be made convincing, however, that such a trace-condition can be linguistically interpreted, i.e. has a meaningful relation to linguistic data. This is still an open empirical issue.

Finally, it is amazing to see that younger linguists like McCawley and Lakoff are not at all bothered by the problem of generative power of their grammars. In fact, what they did was changing Chomsky's transformational grammar in such a way, as to even remove restrictions, i.e. to make the grammar more powerful. The resulting quibles between them and the Chomsky adherents are therefore clearly issues which are undecidable. Both have grammars as powerful as Turing Machines.

3. GRAMMARS AND LINGUISTIC INTUITIONS

3.1 The unreliability of linguistic intuitions

The empirical touchstone in the tradition of transformational linguistics is the linguistic intuition, either of the linguist himself or of an informant. This is also the case in other linguistic traditions, but not in all. Some linguists write grammars for a given corpus, at times on principle, and at times because they are forced to do so for lack of informants. Without taking position on the problem of whether or not intuitions constitute a sufficient basis for a complete language theory, we can in any case propose that their importance in linguistics is essentially limited by the degree to which they are unreliable. It is a dangerous practice in linguistics to conclude from the lack of psychological information on the process of linguistic judgment that intuitions are indeed reliable. Although incidental words of caution may be found in linguistic literature, their effect is negligible. Chomsky warns his readers that he does not mean "that the speaker's statements about his intuitive knowledge are necessarily accurate" (Chomsky 1965), and further states that

> in short, we must be careful not to overlook the fact that surface similarities may hide underlying distinctions of a fundamental nature, and that it may be necessary to guide and draw out the speaker's intuition in perhaps fairly subtle ways before we can determine what is the actual character of his language or of anything else.

Chomsky (1957) emphasizes that, as far as possible, grammars should be constructed on the basis of clear cases with regard to grammaticality. If the grammar is adequate for those cases, the status of less clear cases can be deduced from the grammar itself, and the intuitive judgment is no longer necessary.

After the first phase of the development of transforma

tional generative linguistics, little seems to remain of these two directives in linguistic practice. Instead of an increasing number of cases in which the theory decides on the grammatical status of halfacceptable sentences, we find an enormous increase of examples in which sentences of doubtful grammaticality are applied as tests of syntactic rules.

Even if all problems of doubtful grammaticality just mentioned have been solved, we must still ask what the linguist can do with his reliable data. Data would offer the linguist the opportunity to test his theory, but this does not work only in one direction. The theory (grammar) determines which data are relevant, or, in other words, which linguistic intuitions must be investigated in order to justify certain conclusions. This theory may be said to indicate how the data (intuitions) are represented in the model (the grammar). In this respect the theory of interpretation fills the same function in linguistics as measurement theory in the social sciences (cf. Krantz, et al.1971)

For the direct investigation of the descriptive adequacy of a grammar, that is, for the investigation of the correctness of the structural descriptions, intuitive judgments of another nature are needed; we call them STRUCTURAL INTUITIONS.

Here we shall discuss a type of structural intuition which is sometimes used in linguistic practice and which can offer direct insight into the structure of the sentence intuitions on syntactic cohesion. Cohesion intuitions are expressed in judgments on whether or not words or phrases belong together in a sentence. Chomsky (1965) uses cohesion intuitions for the study of relations between the main verb and prepositional phrases:

> It is well known that in Verb-Prepositional Phrase constructions one can distinguish various degrees of "cohesion" between the verb and the accompanying Prepositional Phrase.

He illustrates this with the sentence He decided on the boat which can be read in two ways. On the boat refers either to the place or to the object of the decision. This

is clear when we compare it with the following nonambiguous sentence: He decided on the boat on the train. Chomsky writes that in the latter sentence "the first prepositional phrase ... is in close construction to the verb", and he modifies the base grammar to agree with this insight. Cohesion is a direct and potentially valuable structural intuition, but its use in linguistics demands a theory of interpretation which establishes the relation between syntactic structure and cohesion judgment.

3.2 The interpretation problem: some empirical studies

Let us start the discussion taking as example the simple sentence John breaks in. There is a gamut of methods for having subjects judge how strong the syntactic relations are among the three words of this sentence. For example the subject can be asked to rank the three word pairs - (John, braks), (breaks, in) and (John, in) - according to relatedness. The most probable result is (from strong to weak): (breaks, in), (John, breaks), (John, in). For longer sentences, where the number of pairs becomes quite large, the task can be facilitated in several ways. One of these is TRIADIC COMPARISONS, in which the subject must indicate for every triad of words from the sentence which pair has the strongest relation in the sentence, and which has the weakest. The triads may be presented, for example, as shown in Figure 2.1. The subject marks his judgment in every triangle by placing a plus sign (+) at the side of the triangle showing the strongest relation, and a minus sign (-) at the side showing the weakest relation. When every triad for the sentence has been judged, each word pair can be assigned a number which represents the relatedness judgment. This can also be done in various ways. One of these consists of counting the number of times a word pair is judged as stronger than other word pairs. Thus, in Figure 1., the pair (breaks, in) is judged as more strongly related than either (John, breaks) or (John, in); this gives a score of 2. The pair (John, breaks) has a score of 1, because it is more strongly related than only one other pair, (John, in), which in turn has a score of 0. If there are more than three words in the

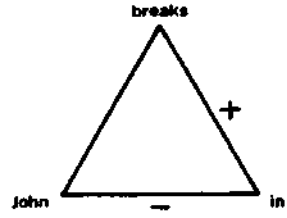


Fig. 1 An example of triadic comparison

sentence, the scores are added for all the triads in which the word pair occurs, yielding the final score for the pair. Other methods of determining the final score are also possible, but we need not describe them here.

An interpretation theory is necessary in order to connect relatedness judgments to a linguistic theory. The purpose is, of course, to test the linguistic theory on the basis of as plausible an interpretation theory as possible.

A general formulation as follows: the constituents of a sentence vary in cohesion, and the cohesion of a constituent is smaller than the cohesion of its parts. This is still nothing other than a faithful explicit representation of a more or less implicit linguistic notion. Without changing anything essential in the formulation, we can define the concept of cohesion mathematically as follows:

DEFINITION (Cohesion): A real-valued COHESION FUNCTION α is defines over the nodes of a phrase marker P, with the following property: if $A \rightsquigarrow B$, then $\alpha(A) < \alpha(B)$, for all nodes A, B in P, where $A \rightsquigarrow B$ means that there is a descending path in P from A to B. The COHESION of a constituent C, $\alpha(C)$, is defined as $\alpha(K)$, where K is the lowest node in P which dominates C and only C.

It follows from the definition that for every path from root to terminal element, the cohesion values of the nodes increase strictly. Consequantly the cohesion of a constituent is necesserily smaller than that of its parts.

The following step is the formulation of the theory of interpretation. This theory must indicate how the strength

of the relation between two words, as judged by an informant, is connected with sentence structure. Let us imagine that we have performed such an experiment for a given sentence, and that the results of the experiment are summarized in a relatedness matrix R, in which the strength of the syntactic relation is indicated for every word pair in the sentence. Thus matrix element r_{ij} in R is the score for the degree of relatedness between words i and j. The score is obtained in one of the ways described in the preceding paragraph. The interpretation theory must attempt plausibly to relate the observed r-values to the (theoretical) cohesion values α. An obvious place to begin would be to find the smallest constituent for every word pair (i, j) to which both words belong, and to compare their degree of relatedness with the cohesion value of the constituent. Let us call that constituent the SMALLEST COMMON CONSTITUENT, SCC, of the word pair. Each word pair in the sentence evidently has one SCC and only one.

The most careful approach, therefore, is to establish no direct relationship between r-values and α's, but only between the rank order of the r-values and the rank order of the α's. The following interpretation axiom states that the rank order of the r-values must agree with the rank order of the α's of the smallest common constituents concerned.

Interpretation axiom: For all words i, j, k, l in the sentence,

$$r_{ij} < r_{kl} \iff \alpha(SCC_{ij}) < \alpha(SCC_{kl}).$$

In this axiom, $\iff$ stands for "if and only if", and SCC_{ij}(SCC_{kl}) stand for the "smallest common constituent of words i and j (k and l)".

Given the interpretation axiom, we can study which phrase marker is most fitting for the observed relatedness values for a given sentence. If we have no particular theoretical expectation concerning the phrase marker, we can draw up a list of the predicted equalities and inequalities for every possible phrase marker in order to find the phrase

marker which best agrees with the relatedness data. In doing so we should remember that different phrase markers for a single sentence do not always lead to the same number of equalities and inequalities. In general, however, we will certainly have particular theoretical expectations concerning syntactic structure, and it will be possible to limit the test to alternatives within that theoretical domain. The following is an experimental example of this.

For the sentence the boy has lost a dollar, only the phrase markers in Figure 2 are worth consideration. In an experiment described elsewhere (Levelt 1967a), twenty-four

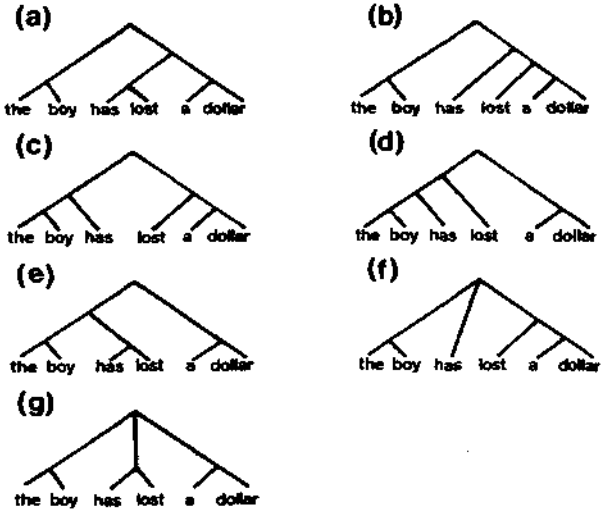


Fig. 2 Possible phrase markers for the sentence the boy has lost a dollar (node labels omitted).

native speakers of English judged this sentence by means of the method of triadic comparison.

Table 1. Shows the relatedness values obtained for the various word pairs. The value for a word pair was obtained by adding the scores for that pair in each triad and for each subject, it is expressed in a percentage.

Table 1. Relatedness Values for the Sentence the boy has lost a dollar.

	the	boy	has	lost	a	dollar
the	-	99	43	29	19	16
boy		-	63	65	16	31
has			-	86	31	40
lost				-	42	70
a					-	94
dollar						-

Table 2. shows the number of inequalities predicted by means of the interpretation axiom for phrase markers (a) to (g), as well as the violations of these given Table 1. (also expressed in percentages in order to facilitate comparison of the models).

Table 2. Number of Predicted and Violated Inequalities for Phrase Markers (a) to (g) in Figure 3.

Phrase marker	(a)	(b)	(c)	(d)	(e)	(f)	(g)
Predicted Inequalities	64	67	58	67	64	46	36
Violations	9	11	7	12	8	5	0
Percentage of Violations	14	16	12	18	13	11	0

The predicted equalities are not taken into consideration here, but even without a statistical test it is quite clear that the results in this respect are in conflict with the expectations.

The problem is thus reduced to the following question:

given a formal grammar, which properties must matrix R of relatedness values have in order to be able to find an accurate structural description within that grammatical model?

We shall at this point find that critical property for the constituent model. Let a, b, and c be three random (but different) elements (words) of a sentence s. Let us imagine the three smallest common constituents for a and b, b and c, and a and c, respectively.

It is quite clear that for the three smallest common constituents, one and only one of the four hierarchical relations in Figure 3 must apply.

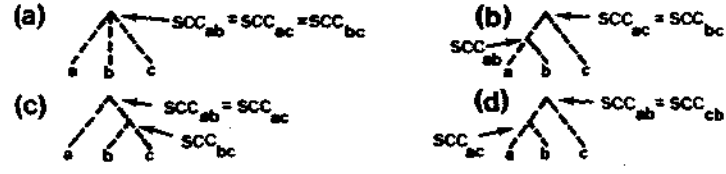


Fig. 3. The four possible hierarchies for the three elements in a phrase marker (dotted lines indicate paths which can contain other nodes)

If (a) is the case for the phrase marker of s, we have the following definition of cohesion:

(1) $\alpha(SCC_{ab}) = \alpha(SCC_{ac}) = \alpha(SCC_{bc})$

If it is (b) we have the following relation:

(2) $\alpha(SCC_{ab}) > \alpha(SCC_{ac}) = \alpha(SCC_{bc})$

If the hierarchical relation is as in (c), we have:

(3) $\alpha(SCC_{ab}) = \alpha(SCC_{ac}) < \alpha(SCC_{bc})$

If (d) is the case, we have:

(4) $\alpha(SCC_{ab}) = \alpha(SCC_{cb}) < \alpha(SCC_{ac})$

By the interpretation axiom, it follows from (1) to (4) that one and only one of the relations (5) to (8) must hold for the observed degrees of relatedness of a, b, and c.

(5) $r_{ab} = r_{ac} = r_{bc}$

(6) $r_{ab} > r_{ac} = r_{bc}$

(7) $r_{ab} = r_{ac} < r_{bc}$

(8) $r_{ab} = r_{cb} < r_{ac}$

These relations (5) to (8) simply mean that r_{ab} must be equal to or greater than the smallest of the two other relations r_{ac} and r_{bc}. This may be summarized as in (9):

(9) $r_{ab} \geq \min(r_{ac}, r_{bc})$

It follows from considerations of symmetry that the inequality also holds for every permutation of a, b and c. (9) is called the ULTRAMETRIC INEQUALITY. In whichever way a, b, and c are chosen, the relatedness values in R must satisfy the condition of ultrametric inequality, if representation by phrase marker is to be possible. In a different context, S.C.Johnson (1967) showed that this is not only a necessary condition, but also a sufficient one: if the matrix is ultrametric, there is a tree diagram which agrees with that matrix.

To summarize, then, it holds that the formal constituent model can be tested by establishing whether relatedness matrices satisfy the condition of ultrametric inequality (9) for all triads. If this is not the case within the measurement error, when the interpretation axiom is maintained, the constituent model must be rejected as such.

We shall limit the discussion to constructions of the type article+noun (the child, a policeman, etc.). Whether

we test the parsing of the surface structure or that of the deep structure, article and noun in the cohesion determinant phrase marker will always be connected at a relatively low level in the hierarchy. Only at a higher level does the noun phrase as a whole come to be related to the other elements of the sentence. But this means that for every third element x in the sentence the smallest common constituent of article and x is the same as that of noun and x. It follows from the interpretation axiom that with the same degree of cohesion the same relatedness value should be expected for these pairs. For the sentence the child cried for help, for example, the theory predicts the following equalities:

$$r(\text{the, cried}) = r(\text{child, cried})$$

$$r(\text{the, for}) = r(\text{child, for})$$

$$r(\text{the, help}) = r(\text{child, help})$$

This holds, no matter what the sentence structure is, provided that the smallest common constituent of the and child includes no other smallest common constituent. Any theory which allows the contrary is a priori in disagreement with current relatedness data, for the relation between the article and its corresponding noun is always stronger than any other relation in an experimental matrix. But the reader can clearly see that the predicted equalities conflict with intuition; one feels that the relations with the article are systematically weaker than those with the noun, and this is indeed what is regularly found in judgment experiments. For the dozens of sentences with article/noun pairs which we have investigated, we have always found, without exception, that the average strength of the relation between the noun and the other words of the sentence is considerably greater than that between the article and the other words. An example of this is the following. The Dutch sentence Meester geeft de doos aan Jetty of aan Thea ('Teacher gives the box to Jetty or to Thea') was presented to eight subjects, who judged the word pair relations on a seven-point scale. The relatedness values (total scores) for de 'the' and doos 'box' are given in Table 3.

Table 3. Experimental relatedness values for the relations between de 'the' and doos 'box' on the one hand, and on the other, the remaining words in the sentence Meester geeft de doos aan Jetty of aan Thea ('Teacher gives the box to Jetty or to Thea')

	Meester 'Teacher'	geeft 'gives'	aan$_1$ 'to'	Jetty 'Jetty'	of 'or'	aan$_2$ 'to'	Thea 'Thea'
de 'the'	10	11	9	9	9	10	9
doos 'box'	38	45	20	38	9	22	35

r (de, doos) = 55

The relations with doos 'box' are systematically stronger than those with de 'the'. Only the minimal relation with of 'or' shows the predicted equality. This result is also characteristic for the strength of the effect: the relations with the article are always close to the absolute minimum score (the minimum score is 8 for eight subjects), while those with the noun tend to cluster around the middle of the scale. It is possible to produce systematic deviations from ultrametricity by introducing article/noun constructions into the test sentence. In general, relations with the head of an endocentric construction are systematically stronger than those with the modifiers.

We may then conclude that the transformational extension of the constituent model must also be rejected when the interpretation axiom is maintained. The model is not capable of accounting for either the strong relation between the article and the corresponding noun, or the weak relation between the same article and the other words in the sentence. Yet this result is not surprising to the intuition. It shows that the relation between article and noun is asymmetric; the article is dependent on the noun, and the noun is the head of the noun phrase. A phrase struc

ture grammar or constituent model is not suited for the representation of such dependencies. An obvious alternative is to use a dependency grammar as a linguistic theory, and to adapt the formulation of the interpretation axiom accordingly.

3.3 A Dependency Model for Relatedness Judgments (1)

In the preceding paragraph we found that relatedness judgments are more a reflection of the relations in the deep structure than of those in the surface structure. We suppose in the present paragraph that the dependency model must be a transformational model. Here, too, the theory has two aspects: a linguistic definition and an interpretation axiom. In a dependency grammar the equivalent of cohesion consists of the two notions of dependency and connectedness. We define a dependency function over the nodes of a dependency diagram:

DEFINITION (Dependency). A real-valued DEPENDENCY function α is defined over the nodes of a dependency diagram D, with the property that if $A \rightsquigarrow B$, then $\alpha(A) < \alpha(B)$ for all nodes A, B in D, where $A \rightsquigarrow B$ means that B is directly dependent on A.

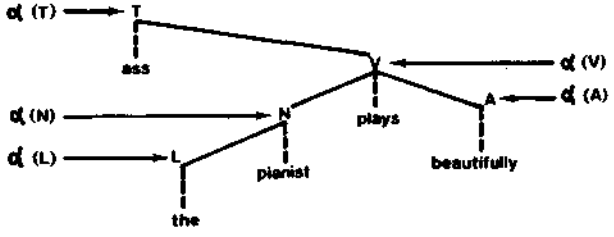
The nodes of a dependency diagram thus have values expressed as real numbers; these values increase in all descending paths of the diagram. The head (the start symbol of the grammar) has the smallest degree of dependency.

If we suppose, by convention, that every element in a dependency diagram is dependent on itself, then for every pair of elements there is at least one element on which both are dependent. The FIRST COMMON HEAD FCH of two elements in a dependency diagram is the element with the high

(1) The suggestion of a dependency model as well as other considerations in this paragraph originated in the work of Mr. E.Schils.

est dependency value α, on which both elements are dependent. This may be illustred by the following example.

Figure 4 gives a dependency diagram for the underlying structure of the sentence the pianist plays beautifully, and an FCH table for all pairs of elements in the diagram. N and A, for example, are both dependent on V, but also indirectly on T. The first common head of N and A is the element with the highest dependency value. It follows from the



FCH	T	D	N	V	A
T	T	T	T	T	T
D	T	L	N	V	V
N	T	N	N	V	V
V	T	V	V	V	V
A	T	V	V	V	A

Fig. 4 Hypothetical dependency diagram for the sentence the pianist plays beautifully, with degrees of dependency and FCH table.

definition of the dependency function that V has a higher dependency value than T, and V is therefore the first common head of N and A. Or consider nodes D and N. They are both dependent on V, but also on N and T. Because $\alpha(N) > \alpha(V) > \alpha(T)$, $FCH_{DN} = N$, as may be seen in the FCH table.

We now define the notion of connectedness negatively as follows:

DEFINITION (Disconnectedness). The DEGREE OF DISCONNECTEDNESS of two elements A and B, $\delta(A, B)$ in a dependency diagram is defined as follows: $\delta(A,B) = [d(A) - d(FCH_{AB})] + [d(B) - d(FCH_{AB})] = d(A) + d(B) - 2d\ (FCH_{AB})$.

Two situations can occur here. The first is that in which FCH_{AB} is different from A and B themselves. In Figure 4, that is the case for D and A: $FCH_{DA} = V$ and $\delta(D,A) = d(L) - d(V) + d(A) - d(V)$. This is the sum of the two reductions in dependency which occur when we pass from the two elements to V. The other case is that in which one of the elements is the FCH of both. This holds, for example, for D and V in Figure 4, where V is the first common head of D and V. The disconnectedness in thus $\delta(D,V) = [d(D) - d(V)] + [d(V) - d(V)] = d(D) - d(V)$, which is the difference in the degree of dependency of D and V. In both cases δ is a nonnegative real number.

We must now give the interpretation theory which relates experimentally measured degrees of relatedness to this linguistic theory of dependency and connectedness.

Interpretation Axiom. $r_{ij} < r_{kl} \Leftrightarrow \delta ij > \delta kl$, for all words i, j, k, l, in a sentence.

It should be noted that the degree of connectedness of two words is considered to be equal to that of the syntactic category which dominates them directly.

The degree of relatedness of two words is therefore greater to the extent that their connectedness in the dependency diagram is stronger, and vice versa.

It is not difficult to see that, on the basis of the definitions of dependency and connectedness, the following should be the case: If two elements B and C lie in the path between two other elements A and D, then the connectedness

between B and C is greater than that between A and D. By the interpretation axiom, it follows from this that $r(B,C) > r(A,D)$. This holds likewise when the two pairs have one element in common: with a path A-B-C we find $\delta(A,B) < \delta(A,C)$, and therefore $r(A,B) > r(A,C)$.

Within the context of the investigation of another problem, we examined the way in which degrees of relatedness behave under pronominalization (cf. Visser-Bijkerk, unpublished undergraduate thesis, 1969). Every reasonable linguistic theory recognized that the boy gave the ice cream to a child and he gave the ice cream to a child have the same structure, with the exception of the substitution of he for the boy. Likewise, the substitution of it for the ice cream, or of him for a child, will also leave the structure unchanged. Three noun phrases can thus be pronominalized in this sentence. Alternate pronominalization of one, two, or all three of those noun phrases will produce seven new sentences, beside the original complete sentence. The eight sentences (including the original) will all have the same structure, with the exception of the pronominalizations. We examined this in the context of the constituent model as well as within that of the dependency model. In the experiment this sentence (in Dutch) was used together with seven others, all with corresponding syntactic structure. The eight sentences were for following:

de jongen gaf het ijsje aan een kind
'the boy gave the ice creem to a child'

de man betaalt het geld aan een agent
'the man pays the money to a policeman'

de miljonair schonk het schilderij aan een pastoor
'the milionaire presented the painting to a priest'

de directeur stuurde het honorarium aan een advocaat
'the director sent the fee to a lawyer'

de meester leende het boek aan een leerling
'the teacher lent the book to a pupil'

de slager overhandigde het vlees aan een klant
'the butcher handed the meat to a customer'

de eigenaar vermaakte het huis aan een invalide
'the owner bequeathed the house to an invalid'

de grossier leverde het hout aan een timmerman
'the wholesaler delivered the wood to a carpenter'

With all the pronominalizations, this gave sixty-four experimental sentences. Each subject was presented with all the forms of pronominalization, and asked to judge them on seven-point scales. Each form was derived from a different sentence content, and the sixty-four sentences were distributed in such a way to eight subjects that each sentence was judged only once. We shall limit our discussion to the results of each form of pronominalization, that is, the totals for the various forms over subject and sentence content; therefore we shall indicate the various words with their category symbols. The sentences on which no pronominalization has been carried out have the form $D_1N_1VD_2N_2$ to D_3N_3; those in which the first noun phrase has been pronominalized have the form he VD_2N_2 to D_3N_3, and so forth. Note that the three articles are all different in Dutch (de, het, een), and thus no confusion was possible.

Analysis showed that the data obtained seriously conflicted with the constituent model. The principal deviation had to do with the predicted equalities for the relations with article and noun. With one exception, the relations with the noun are stronger than those with the corresponding article, quite in agreement with that which was discussed in the preceding paragraph.

There were also great deviations from the constituent model concerning inequalities. The ultrametricity of the matrices was limited, and alternative phrase markers were always found for the various forms of pronominalization.

The experiment, reported here by way of example, is no proof of the correctness of the dependency model. Further experimentation will certainly lead to modifications and additions. The purpose of this chapter was to show that to an explicitly formulated grammar an equally explicitly formulated interpretation theory could be added, making it possible

to investigate the descriptive adequacy of the linguistic theory. We found that a transformational grammar with a phrase structure grammar as its base is not descriptively adequate in a number of regards, and that a dependency grammar as base avoids many of the difficulties. In both cases, the linguist can set these findings aside by rejecting the interpretation theory. To do so, however, will oblige him to find a better interpretation theory, and it is by no means excluded that this is possible. In that case, the linguist will finally have to attend to a matter which he usually neglects, namely, the theory of the relationship between formal linguistic model and concrete linguistic data.

I am deeply grateful to Dr. Paolo Legrenzi, who managed to compose this chapter from the written and printed parts and pieces that were handed in by me. W.L.

4. SYSTEMS, SKILLS AND LANGUAGE LEARNING

4.1 Language as skill

Language behavior, like any other complex human activity, can be approached from a variety of viewpoints. One could be mainly concerned with the actual or potential output of such behavior, i.e. with the structure of a corpus or language. Alternatively, attention could be directed to the communicative function of language, the transmission of intentions from speaker to hearer and the interpersonal variables that play a role in such communication.

Somewhere between the purely linguistic and the purely social-psychological points of view is the approach which considers language as a human skill. A skill analysis of language borrows from linguistic analysis in that the linguistic structure of the input or output message is systematically varied in order to measure its effects on speed, accuracy, timing and other aspects of linguistic information decoding and encoding. In its turn, knowledge of language as a skill is required for effective analysis of language as interpersonal communication. It is especially important to have an understanding of the mechanism of selective attention and motivation in the transmission of linguistic information in order to fully appreciate the facilitative or inhibitory effects of interpersonal variables in the functional use of language.

Apart from bridging the gap between a more structurally and a more functionally directed study of language, the skills approach to language behavior has the definite advantage of leading to a natural integration into an already existing body of psychological knowledge. The study of human skills, including symbolic skills, has been intensive and quite succesful since World War II. This is not the place to review the enormous developments in the post war study of "human factors", nor to outline the deep influence of cybernetic thinking on the analysis of skills. The reader may be referred to a recent volume on one symbolic skill, human

problem solving (Newell and Simon, 1972)., to get an appre ciation of this revolution in psychological thinking.

Herriot (1970), who was one of the first authors to stress the analogies between language behavior and other skills, especially mentioned the following features of skills which have been intensively studied, and which are equally central to language.

(a) Hierarchical organization. It is not necessary to con vince linguists of the hierarchical nature of language,we will return to this in section4.3.But many other skills are hierarchical in structure. The succesful completion of a task is, in almost all skills, dependent on the accurate performance of subtasks, plus the correct temporal or spa tial integration thereof.

(b) Feedback. Nearly all human performance is controlled by comparing the behavioral effects with some internal standard or aim. The difference is then reduced by taking appro priate measures. This is especially salient in problem sol ving behavior, but it is also true for many aspects of lan guage. A speaker's behavior, for instance, depends to a lar ge degree on signs of understanding on the part of the listener.

(c) Automation. After a skill has been acquired it is to a large degree automatic, i.e. it does not require conscious control of each of its subtasks. Automobile driving is an example in case: during normal driving, one's attention is free for even rather complicated discussions. Skilled language use is similar in that there is no conscious attention to articulatory movements, or even to choice of senten ce schemes. Attention is normally mainly with the semantic contents, and sometimes with the choice of appropriate lexi cal "core" terms.

(d) Anticipation. In skill research subjects often "react" before the appropriate stimulus is given. The accurate timing of the concert soloist is not by rapidly reacting to the conductor's sign, but by anticipating the critical moment. Any skill which involves planning also allows for an ticipation. Speech perception is "being ahead of the spea-

ker". This is possible because all speech is redundant. To the degree that the listener is familiar with the theme, he is able to anticipate by making hypotheses about what the speaker is going to say. As for any skill, this does not require much of a conscious effort. Anticipation is not necessarily a conscious phenomenon.

One could easily add other typical skill features that are equally essential in language behavior. Instead of expanding this issue any further in the present context we will finish this paragraph by mentioning two more reasons why the skill point of view can be especially fruitful for the study of language.

Of all psychological study of skill the major part concerns skill acquisition. Much is known about factors which facilitate or interfere with the learning of skills (see e.g. Bilodeau, 1966). It should be interesting to know how much of these findings can be generalized to language acquisition. Especially the study of second language learning should profit from this viewpoint, because almost all skills are learned on the basis of already existing skills, just as in second language learning. The degree of compatibility between the old and the new skill has been a very central issue in the study of skill acquisition.

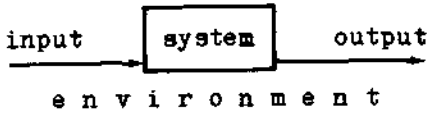
Finally, the cybernetic revolution in skill research has led to a high degree of theoretical modelling in the analysis of skill, and especially to the introduction of very general formal systems for the description of skilled behavior. Skill research is increasingly profiting from what is known as systems analysis or system theory, of which some basic notions will be introduced in the next section. Such formal models are specifically developed for the theoretical representation of features such as feedback, hierarchy, anticipation, control, automation, learning. It is therefore, surprising that no systems analysis of (apects of) human language behavior has ever been envisaged. The remainder of this chapter is intended to give some general thoughts on this issue. We will first introduce some central notions of system theory (section 4.2). Next, we will devote a few words to a stratified description of the language user

(section4.3). Instead of staying in this general mode, we will select one stratum, the syntactic level, for further analysis in terms of systems (section4.4). It will be shown that empiristic and rationalistic models of language acquisition can be theoretically analyzed in such terms and that both are wrong-in-principle (section4.5). Finally attention is given to some more global aspects of second language acquisition (section4.6). This chapter does not present any new empirical finding; its only aim is to present a way of thinking about matters of language acquisition which, though not new in itself, might lead to fruitful theoretical integration of grammar, skill research and applied linguistics.

4.2 System theory: some basic notions

There are many rather different definitions of the notion "system" (see e.g. Bertalanffy, 1969). Throughout this chapter we can neither be complete, nor go into much mathematical detail. In this section we will arbitrarily choose the following description of what we mean by a system. A system is any part of the real world which is considered apart from the rest of the world. This latter, the complement of the system, is called the system's environment. The environment may influence the system by means

Figure 1
System and environment



of what is called input into the system. In its turn the system may affect the environment by means of a certain output. The system may be in any of a finite or infinite number of states. The state is the present condition of the system. It is defined in such a way that for all possible

cases it is true that given the state of the system as well as the input it receives in that state, it is fully determined what the next state and the next output will be.

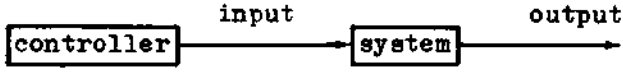
Different classes of systems can be distinguished dependent on the types of input, output and state descriptions one chooses. If input, output and state transition is to be considered as occuring at discrete moments in time, the system is called a discrete-time system. Successive instants can then be numbered, and the behavior of the system can be completely described by the state transition function, which gives the state at the next instant as a function of the present state and the present input, and the output function, which gives the next output as a function of the present state and the present input. If moreover, the set of elementary inputs (i.e. inputs that can be applied at one given instant) and the set of elementary outputs are finite, the system is called an automaton. The automaton is finite if the set of states of the system is finite, it is infinite otherwise.

It is, in the present context, useful to think of systems in terms of automata, because most language behavior is characterized by discreteness in time and finiteness of input and output vocabulary. It should be kept in mind, however, that this limitation is not essential in system theory.

Essential in system theory is the notion of control. Assume that the state space of the system contains a designated initial state, s_0, as well as a desiguated arbitrary final state s_f. The initial state s_0, is controllable if there is a string of inputs which leads the system from s_0 to s_f. The system is controllable if every state of the system is controllable.

The idea of control is that we want to bring the system in a desired state (giving a desired output), and the question is whether we can do it, and if so, what string of inputs should be applied in order to obtain this goal. This can be depicted as follows:

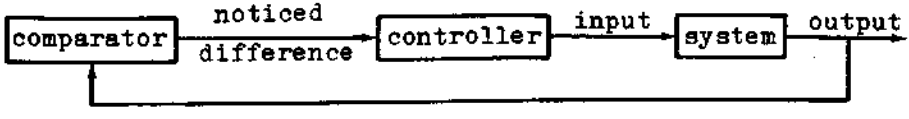
Figure 2 - Diagram of system control



This notion of control will be used in section 4, where we will consider the listener as the system, the state of the listener in which he accepts the message as the desired state, and the speaker as the controller who has the task of leading the listener into this desired state, by choosing an appropriate input string of words.

The notion of feedback comes in if the controller is able to compare the factual output of the system with the desired or reference output. This is depicted in Figure 3: For the purpose of

Figure 3 - Diagram of control through feedback

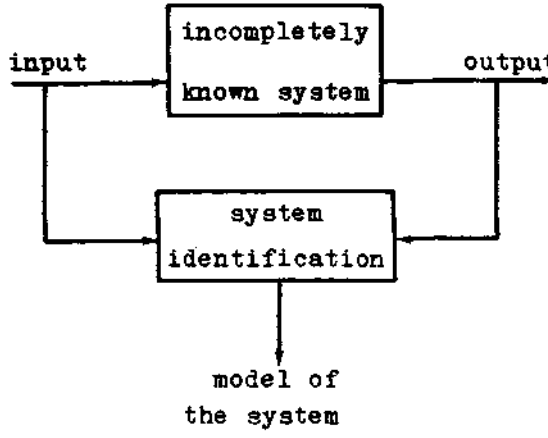


clarity the comparison of factual and comparison of factual and desired output has been set apart in a separate box. The controller acts on the basis of the noticed difference and chooses an input which may lead to a decrease of the difference.

An interesting chapter of system theory is concerned with the so-called identification problem. If our knowledge of a certain system is limited, how can we learn to control the system without opening it? In that case we have to estimate as accurate as possible the structure, or parameters, of the system, by sistematically sampling input/output pairs. Another way of formulating the identification problem is: can we devise a procedure which gives us an accurate model of the system, by observing a finite set of input/output

strings. If an accurate model, i.e. a model which simulates the system perfectly, can be derived, we can approach the control problem by trying to solve it for the model. The identification problem, which will be related to the problem of language acquisition in section 4, is summarized in the diagram of Figure 4.

Figure 4 - Diagram of system identification



It is often possible to organize the description of a system in terms of sub-systems and their interrelations. There are several different notions of hierarchy in system theory, we will limit ourselves to one: the notion of a stratified hierarchical system. One can consider the same system on different levels of detail. Figure 5 is not taken from a linguistic or psycholinguistic text, but from a text on hierarchical systems (Mesarović et al., 1970).

One may consider one and the same system, for instance a speaker delivering a lecture, from a very detailed point of view (e.g. as a producer of a sequence of elementary sounds), or from a global point of view (as a producer of a certain textual composition), or from several intermediate levels of detail. Each level of description has its own sets of inputs, outputs and states. On the level of sentences, for instance, the elements are words (or morphemes),

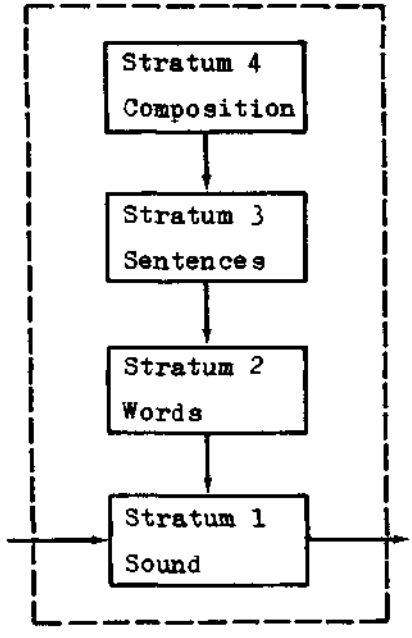


Figure 5 - A four-strata diagram of a text generating system.

but it is irrelevant whether these words are written or spoken, or spoken by a male or a female voice, etc. The latter features, however, are essential for a stratum 1 - description.

In general, the description of one stratum cannot be derived from the description on another stratum. Each level has its own concepts and principles. It is, especially, impossible or unfeasible to describe a high level stratum in terms of a low level stratum. One cannot derive processes of human problem solving from principles of neural interaction, or the principles of text composition from syntax. But one should keep in mind that in a stratified description it is the same system which is described on different levels. A state of this system is the composition of the different states of the subsystems at a certain in

stant in time. The state of a lower level subsystem is co-determined by the output of a higher level stratum. This influence is called intervention, and is depicted in Figure 5 by downward arrows. The intervention of stratum 4 upon stratum 3 means that the text generating system does not generate a random sequence of sentences, but that successive sentences are chosen such as to produce a coherent text.

There are some general principles that hold for all stratified systems: (a) The higher level is concerned with larger positions and broader aspects of the system's behavior, (b) decision times on the higher level are usually longer than decision times on the lower level, (c) the higher level is concerned with the relatively slow aspects of the system's behavior (d) description of a higher level is usually less structured, less certain, and more difficult to formalize than the description of low level behavior of the system.

4.3 The language user as a system

The structure of a human language user is so complicated that we have little a priori knowledge about its possible states, state transition function or output function. A complete and detailed description of such a huge and complew system is excluded from the beginning. On the one hand one wants to create a model of the language user's global behavior, i.e. his communication with other language users about certain aspects of the real world. On the other hand, one has to fill in all the details of such behavior on all levels of functioning. In such cases the system theorist resorts to a stratified description. He defines different levels of detail and tries to create more explicit models for each of the sybsystems. The subsystems should be chosen in such a way that their functioning is as much as possible independent from other subsystems. This description can then be extended by a specification of the intervention and other relations between levels and subsystems. It is, therefore, completely legitimate to choose a certain stratum for further analysis. One should only keep in mind that it is a

part of a larger system, and that its description should, in the long run, be integrated in a more general characterization of the system.

There is nothing new here for linguists. Linguistics is a highly stratified science with various levels of description such as phonology, morphology, syntax, semantics, more or less comparable to the strata of the system in Figure 5.

Also in psycholinguistics the use of hierarchical models for speaker or listener are increasingly common. This is especially so in studies directed toward computer simulation of natural language understanding. The reader is referred to Winograd's (1972) system as a recent example. It consists of a hierarchy of subsystems, each having its own principles of functioning, but nevertheless cooperating in a global and sometimes surprisingly "human" manner.

In this section we will not propose any stratified model for a language user. Instead, we will arbitrarily select one level of description, the syntactic level, for the purpose of discussing the contributions system theory can make to the problem of (second) language acquisition. The syntactic level is selected because results are most clear-cut in that area, not because this stratum is the most important for understanding language acquisition. In fact it will be shown in the next paragraph that a syntactic account of language learning is unfeasible. But the syntactic level is certainly the highest level for which such results could be obtained through formalized analysis.

4.4 Some system aspects of the syntactic stratum

Consider the listener as a system. Though for the system as a whole the usual input is a text, and the desired final state is one of understanding of that text, on the syntactic level this input/output relation reduces to a sentence as input and a syntactic structural description as output. The syntactic subsystem reaches a final state if a correct structural description of the sentence has been

created. One calls this state the accepting state. Generally, the listener does not overtly output the structural description, so that the speaker does not know whether the accepting state has been reached. However, control is nevertheless often possible since the speaker shares the language with the listener and can therefore plan the input in such a way as to be sure that an accepting state is indeed obtained. The speaker/listener situation so far can be represented by the elementary control-diagram of Figure 2; where the system is the listener, and the controller the speaker. If we call the state of the listener before the utterance is presented the initial state, according to system theory this initial state is controllable if there is an input string which brings the listener into the accepting state. It is interesting to notice that in the ideal case, i.e. where the listener has unlimited memory, etc., the set of all input strings by which the system can be controlled in the initial state is the language itself. The linguistic notion of grammaticality, therefore, is a special case of the notion of controllability in system theory.

The notion of feedback comes in if the speaker is not completely with the listener's linguistic outfit. Important cases are the child, talking to his mother, and the beginning second language learner who tries to make himself understood by a native speaker of thet language, or more typically by his language teacher. In such cases it is very important for the controller to get feedback, as in Figure 3, about the state of the listener. If a certain utterance is not understood or accepted by the listener, the speaker could try a different wording if only the listener gives some clue with respect to his state of understanding. From the purely syntactic point of view this amounts to feedback with respect to whether a certain input string has led the listener into the accepting state or not.

This brings us to our main theme, the systems approach to language acquisition. In terms of system theory, language learning is a case of the identification problem. The language learner is confronted with an incompletely known

system, the fluent language user, i.e. speacker/listener. In order to "control" this system, i.e. to communicate in the new language, the learner has to make hypotheses about the system's structure and parameters and test such hypotheses by checking sample of input/output pairs. This is exactly the situation depicted in Figure 4. The system identification box represents the language learner, who infers a model of the system by observing a set of input/output pairs. Again limiting our attention to the syntactic stratum, such a pair consists of, on the one hand, a string of morphemes or words and, on the other hand, some indication of whether the string is acceptable or non-acceptable to the system. If the system is a syntactically ideal system, this indication simply means, as we have noticed before, that the corresponding string is either grammatical or ungrammatical. Here it is immaterial whether the unknown system is a listener or a speaker of the language. Syntactically this amounts to an inversion of input and output, whichdoes not affect the essential character of the pairs: they always consist of a string and a plus or minus-sign. If the sign is positive, the particular pair is called a positive example, i.e. the learner knows that the particular string is a sentence in the language. Because a syntactically ideal speaker always produces grammatical text, a positive example is best imagined as drawn from a speaker-system. If the learner is exclusively presented with positive examples, i.e. a sequence of grammatical sentences, one calls such a sequence a text presentati If, however, the sign is negative, i.e. if the string is not a sentence of the language, the pair is called a negative example. If we consider the unknown system as an informant to whom we present strings with the question whether they belong to the language or not, we will sample a mixture of positive and negative examples: some strings turn out to be grammatical and others are faulty. Such a mixture of positive and negative examples is therefore called an informant presentation.

As we have seen in section 2, the essential problem of system identification is whether we can devise a procedure which can generate an accurate model of the system by observing a finite set of examples. On the syntactic level, such

a model is called a grammar of the language, and the question is then if a correct grammar of the language can be derived from a finite text or informant presentation. If the answer is affirmative, such a procedure could be an ideal model of the language learner, and actual language acquisition could be studied on the basis of such an ideal model (1). If the answer is negative, however, it makes no sense whatsoever to even try to understand the acquisition of syntax as a relatively autonomous process. Before we study processes of language acquisition, we should first solve what Chomsky (1965) called the adequacy in principle of a theory of language learning. If there is no conceivable procedure to output a grammar on the basis of a finite presentation of the language, be it text of informant presentation, then any theory in such terms must be wrong, since children and adults do acquire languages.

Before we introduce, in the next paragraph, some substantial results with respect to this adequacy-in-principle, we must add two more notions which are essential for a discussion of theories of language acquisition.

System identification is impossible without some a priori knowledge of the structure of the system. One should, for instance, have some knowledge of the sort of input accepted by the system, or linguistically speaking, the learner must have some idea about the class of languages that should be considered.

The set of models, or syntactically speaking: grammars, which agree with this a priori knowledge is called the hypothesis space in system identification. It is obvious that language acquisition is greatly facilitated if the hypothesis space is made very narrow. This means that the learner already has very detailed a priori knowledge of the language to be learned.

(1) The construction and testing of ideal models is common practice in many areas of psychology. Compare for instance the ideal perceiver models in signal detection theory.

Another way to speed up learning is to make the learner very "clever". He could be endowed with very powerful heuristics which allow him to scan the hypothesis space in a very systematic way, and to process huge amounts of observations in very short time.

4.5 Adequacy of empiristic and rationalistic acquisition models

The system identification procedure presented so far can be seen as a schema for organizing the discussion about language acquisition in terms of the syntactic stratum. It corresponds to what Chomsky and Miller became to call a language acquisition device, LAD (Miller and Chomsky,1957; Chomsky, 1962). But there are two important points to keep in mind before we proceed this discussion.

First, LAD is a schema which is limited to the syntactic stratum. As we have seen in section 2, concepts and principles can be quite different for different strata of the system and there is no reason whatever to expect that substantial results for the syntactic stratum will be valid for other strata as well. We should not expect to solve the language acquisition by solving it at the syntactic level. This is in sharp disagreement with Chomsky's position. Chomsky (1962) tries to minimize the additional role of the semantic stratum in language acquisition. He writes "For esample, it might be maintained, not without plausibility, that semantic information of some sort is essential even if the formalized grammar that is the output of the device does not contain statements of direct semantic nature. Here care is necessary. It may well be that a child given only the input of Figure 2 (i.e. of LAD) as nonsense elements would not come to learn the principles of sentence formation. This is not necessarily a relevant observation, however, even if true. It may only indicate that meaningfulness and semantic function provide the motivation for language learning, while playing no necessary part in its mechanism, which is what concerns us here. And Chomsky repeats this

argument in Aspects (1965, p. 33). In a moment we will discuss how much of this position can be maintained.

Second, LAD is nothing else than a schema for the discussion of language acquisition procedures. LAD is only meant to be a hypothetical system identification procedure endowed with a hypothesis space and a set of heuristics, with a text or informant presentation as input and a grammar, i.e. a model of the system, as output. At this point the literature is badly confused and quite misleading. The confusion mainly relates to the distinction between empiricistic and rationalistic acquisition models, which we will now introduce. In Aspects, Chomsky formulates this distinction in terms of LAD as follows.

The empiricistic model of language acquisition says that there is hardly any limitation with respect to the hypothesis space of LAD, it has little a priori knowledge of the system's grammar. Language learning occurs through strong heuristic principles by which the grammar is derived from observations.

The rationalistic model, on the other hand, assumes that LAD's hypothesis space is very narrow or specific; there is a large a priori knowledge of the system's grammar. A relatively small set of observations will suffice for LAD to derive the system's grammar.

Both models, therefore, are special conceptions of LAD's structure. The main confusion in the literature resulted from contaminating the LAD discussion schema with the rationalistic assumptions about LAD. The most outstanding example in this respect is McNeill (1970), but many others made the same short circuit, often to their own disadvantage. Braine (1971), for instance, weakened his argument against syntactic acquisition models by making the same contamination, as we will see.

A second source of confusion is the identification of rationalistic with innate, and empiricistic with learned. Though it is not implausible that the a priori knowledge of the grammar is innate in some sense, it is exactly equally plausible to suppose that the strong heuristics in

an empiricistic model are innately given. Innateness has no intrinsic relation with the dichotomy under concern. Here we will not go into the innateness issue. We refer the reader to Levelt (1973), where it is treated in much detail.

Let us put the discussion straight. The first question concerns the adequacy-in-principle. Can one conceive of whatever procedure which derives the grammar from a finite text or informant presentation? Only in the affirmative case it makes sense to pose the second question: how does the child, or second language learner, compare with such an ideal procedure? Chomsky (1965) makes a very one-sided statement with respect to these questions. He writes: "In fact, the second question has rarely been raised in any serious way in connection with empiristic views... since study of the first question has been sufficient to rule out whatever explicit proposals of an essentially empiricist character have emerged in modern discussions of language acquisition". The facts are, however, that the question of constructability of a language acquisition procedure had not been solved at all in 1965. Substantial results in this respect have only been obtained by Gold in 1967 and by Horning in 1969. These latter solutions have been completely ignored by both linguists and psycholinguists, so that it makes sense to give a very short summary of the main results. Technical detail, however, must be left out in the present context. The interested reader is referred to the original publications, or to Levelt (1973), chapter 8.

Gold (1967) could prove the following. With text presentation an error-free acquisition procedure can only be constructed if the hypothesis space is limited to finite languages. That is, with text presentation, a language can be learned in principle if and only if the learner knows in advance that the language is finite.

Since natural languages are quite clearly not finite, they cannot be learned by text presentation in Gold's sense. Gold's mathematical results were extended by Horning. Instead of discussing the error-free case, Horning discussed a stochastic version of the identification procedure. He proved that the difference between the grammar derived

by LAD, and the "real" grammar of the system can be made arbitrarily small in the case of (stochastic) text presentation, if LAD knows in advance that the system's grammar is of the non-ambiguous context free type. Natural languages are clearly of a more complicated type, be it alone for the fact that natural languages are ambiguous,and the question is what the results would be for more complicated stochastic languages. This has not yet been solved. But for our purpose it is not too important to wait for such solutions. With respect to the second question, the factual properties of the acquisition procedure, Horning could prove that even for the context free case, where acquisition is possible in principle, the procedure is so time-consuming as to be completely unrealistic as a model for human language acquisition: "grammars as large as the ALGOL-60 grammar will not be attainable simply by improving the deductive processing"."But adequate grammars for natural languages are certainly more complex than the ALGOL-60 grammar". So, even with the strongest heuristics, a text presentation model for natural language acquisition is excluded as a realistic model.

How is the situation for informant presentation?This is very much better. Gold could prove that even if LAD only knows that the language is primitive recursive,which is probably true for all natural languages, it can derive a correct grammar for the language. Though this might seem to be a hopeful alternative to the text presentation model, in this case we hit upon too much empirical countereviden ce. This has most clearly been formulated by Braine (1971). He argues that the language learning child is at best presented with positive examples. If presented with ungrammatical utterances, these are hardly ever marked as such. In our terms, Braine argues that the child is, at best, in a text presentation situation. We mention some of several arguments: (1) The speech of many children is never corrected, i.e. marked as grammatical or ungrammatical. Nevertheless all children finally acquire their language. (2) If such marking occurs, it seems to be highly ineffective as a means for language improvement. This is clear from experiments by Braine (1971) and Brown (1970). Therefore, the

"this-is-ungrammatical" -output of the adult can hardly be considered as input for the language identification procedure. It should be noted that the same is true for second language acquisition. Experiments by Crothers and Suppes (1967) show that presentation of negative syntactic information does not improve the acquisition of certain syntactic forms in Russian. (3) Informant presentation in Gold's sense requires, roughly speaking, that every ungrammatical string will, in the long run, occur in LAD's observations. This, however, is highly unrealistic, since it is known (see Ervin-Tripp, 1971) that the speech directed to young children is highly grammatical and hardly ever contains negative instances. It seems to me that this is also very much true for the second language learning situation in so-called natural teaching methods. Students are almost exclusively presented with positive examples. (4) One could think that non-reaction of adults to ungrammatical strings might constitute implicit negative information for the language learning child. This can certainly not be the case. Initially, almost all utterances of the child are ungrammatical in the adult's sense. Nevertheless, the adult reacts if he can derive the child's intention. This means that many ungrammatical strings are "marked" as positive. This should confuse any language acquisition procedure. This situation is fully comparable to the learning of a language in a foreign country, or by means of most "natural" methods. Conversation is not interrupted for reasons of ungrammaticality, but mostly for inunderstandability only.

If these arguments are sufficiently convincing, it follows that the language learning child, as well as the second language learner in a foreign country (still the quickest way to learn a second language!), are essentially in a text presentation condition.

But since the work of Gold and Horning we know that there is no conceivable real-time acquisition procedure for natural languages within the syntactic stratum. The conclusion therefore must be that the adequacy-in-principle question must be answered in the negative for all modeld of the LAD-family, i.e. not only for the empiristic models,

but also for the rationalistic models.

It is now interesting to look back at the literature. From the citation above, it is clear that Chomsky (1965) rejects the empiristic model, without answering the adequacy-in-principle question. Even according to his own writing, however, the latter issue should have been solved first. It is only due to this lack of substantial results that Chomsky, and with him McNeill and many others, could keep believing in the adequacy of a rationalistic model. On the other hand Braine (1971) quite correctly rejected the rationalistic model by arguing that it is unfeasible with text presentation. He then made a case for an empiricistic model. But it should by now be clear that the test argument relates to the adequacy-in-principle of the LAD-schema as such, and that Braine's argument therefore leads to rejection of both versions of LAD, i.e. including his own empiristic version.

The only safe conclusion is that all exclusively syntactic accounts of language acquisition must fail for principled formal reasons, be they empiricistic or rationalistic. Chomsk's assumption which was cited at the beginning of this section, saying that an essentially syntactic account of language learning might suffice, cannot be maintained. This is, moreover, little surprising from the system theoretical point of view, and even less so from what we know about language teaching.

One note could be added. This discussion did not solve the rationalist/empiricist controversy. It can be reformulated on another, especially a higher stratum of the system description. Even about the level of intention and meaning one could ask whether a child or second language learner acquires such structures by analyzing his observations by means of strong heuristic principles, or alternatively, whether he has strong advance knowledge of such structures and can easily select the correct structure by only making a relative small amount of observations.

4.6 Some global aspects of second language learning

In this final section we return from the syntactic stratum to some more global aspects of the language learner. More specifically, we will make some remarks on three points. The first is the question of facilitation and interference due to the first language. The second issue relates to the acquisition of hierarchical skills and possible conclusions for language learning. The third issue is some possible causes of failures in second language learning.

(a) Facilitation and interference.

One of the most intensively studies phenomena in skill research is the role of compatibility in skill acquisition (see for instance Bilodeau, 1966, Fitts and Posner, 1967, Welford, 1968). The question is how much the learning of a new skill is facilitated by similarity with an already existing skill. If one has learned to perform some task (e.g. writing) with the right hand, how easy is it to learn to do the same task with the left hand? If a child wants to learn to drive a bicycle, is it advantageous if he already has some skill on the scooter? A very general summary of numerous experimental findings is the following: compatibility between old and new task is facilitatory in the sense that the initial skill at the new task is higher. However, compatibility hardly affects the speed of learning. Therefore, compatibility is not reflected in speed of learning, but only in the maintainance of the initial advantage.

If this general result can be extended to second language learning, one should expect that the learning of Japanese is not slower than the learning of French, but that, throughout learning, the proficiency in Japanese will be less than the fluency in French. The large effect of compatibility on second language learning has been demonstrated by Carroll and Sapon (1959), see also Carroll (1966).

Little is known about the causes of the compatibility effect. In terms of system theory one would suppose that the facilitatory effect of language similarity is due to the

restriction of the hypothesis space that the language learner can allow himself. An interesting aspect of such restriction is that there is no a priori lower limit. The apparent similarity between first and second language can easily induce the learner to over-restrict his hypothesis space. This results in what is known as interference in skill and second language research: the learner keeps making intrusions from his native language. I would not be surprised if it were shown that there exists an optimal similarity between languages: if the second language comes too close to the first, interference may become more important than facilitation. In that case the task for the language teacher would be to expand the hypothesis space by contrastive teaching. Newmark and Reibel (1968) reject this approach, but much more research is required to give a definite answer.

(b) Acquisition of a hierarchical skill

Fitts and Posner (1967) distinguish three stages in the acquisition of hierarchical skills. The first stage is learning of individual components. Each component initially requires full attention, therefore they can only be trained in succession. The second stage is called integration. Dependent on the depth of the hierarchy different or all components are organized in larger wholes. The learner tries to get familiar with the spatial and temporal relations between the subtasks. Finally the stage of automation is reached. In section 2 we noticed that in a stratified system slow decisions are feasible at the higher levels where the broader aspects of planning take place. All skilled behavior is characterized by full automation at lower levels, so that the subject's attention is available for controlling the performance as a whole.

All this applies to language learning as well. Initially the language learner has to give attention to all sorts of minor components of the skill: the pronunciation of individual sounds, the meaning of individual words, etc. Only then integration becomes possible. In its turn this leads to a higher level integrated component, e.g. a correctly pronounced and understood word, which requires further

syntactic integration, etcetera.

Horning (1969), after his negative conclusions with respect to language learning from text presentation, remarks that, in the case of the child, language learning probably proceeds quite differently. The child is not presented with the full blown language, but with a very limited subset of the language. Probably the child initially does have an extremely limited hypothesis space and the parents are nicely matching it by presenting the child with a very simple language. One could say that the child is learning a mini-grammar. Recent research (Ervin-Tripp,1971) has indeed shown that the language which adults direct to their very young children is extremely simple in structure: it does not contain conjunctions, passives, subordinate clauses, etc. Moreover, sentences are very short. Therefore Horning may be correct: the child is learning a mini-language, which is gradually expanded in later stages. In terms of skill integration: the initial language becomes a higher level component of the language in a later stage. In this way a growing set of already automated sentence schemes becomes available to the child, who in his turn keeps expanding his hypothesis space for whatever reason.

It is noteworthy that his idea has been around since a long time in second language learning practice. This is especially the case for the Berlitz-method (1967). Right from the first lesson a mini-language is learned which suffices to discuss some little subject. In later lessons this is gradually expanded by new words and forms, but at each stage one aims at maximal automation or fluency before proceeding to the next stage. This is fully comparable to the teaching of other symbolic skills such as arithmetic. One preferably starts with one operation (addition) in a limited domain (1-9), and gradually expands if sufficient automation has been acquired.

But again, much more research is required with respect to the optimal organization of the training of hierarchical skills. No general principles are as yet available.

(c) Some causes of failures in second language learning. From the systems point of view failures in language learning can be due to a variety of factors. We already mentioned interference through a too restricted choice of the hypothesis space. Contrastive teaching might be helpful. Also, certain parts of the system's behavior might not have been observed by the learner, and his model of the fluent language user would therefore remain incomplete. An example which has often occurred to me, but which does not seem to get much attention in language teaching is lip position. It is well known that in many cases exactly the same sound can be produced with different lip positions. In a language course one does learn to make the correct sound, but one is not taught that the natuve speaker makes a characteristic lip position with the sound. People tend to keep their "native" lip positions even if they pronounce faultlessly. Since looking at the speaking face is an important addendum to language understanding (see e.g. Campbell, 1970), such people may always be hampered in their verbal communication, as well ad recognized as foreigners.

As long as a task is not too difficult, performance may appear to be fully automated, whereas in fact the learner is still giving attention to several low level components. This is immediately revealed if the subject's attention is distracted, either by a secondary task or by stress (speeding up performance or otherwise). The less a skill is automated, the earlier it will break down. If tasks during second language teaching are kept too easy, the subject may seemingly acquire a high level of skill, but nevertheless fail at a stressful examination. During language courses, the teacher should from time to time "test the limits" in order to detect which components are most likely to break down, and are thus least automated.

Finally, some errors persist because the learner intends to "control" the native speakers in a very special way. He does not only want to make his intentions understood, but also the fact that he is a foreigner. This can often be quite advantageous for all sorts of social reasons. (See Diller, 1971, for discussion of this point).

REFERENCES

Bar-Hillel, J., Perles and Shamir, On formal properties of simple phrase structure grammars, Z. Phonetik Sprachwiss Kommunikationsforschung, 14, 143, 1961.

Berlitz Publication Staff, English: First Book, Berlitz Publications, New York, 1967.

Bertalanffy, L. von, General System Theory, Braziller, New York, 1968.

Bilodeau, E.A., Acquisition of skill, Academic Press, New York, 1966.

Braine, M.D.S., On two models of internazionalization of grammars, in The Ontogenesis of Grammar. A Theoretical Symposium, Slobin, D.I., Academic Press, New York, 1971.

Brown, R., Psycholinguistics. Selected Papers, Free Press, New York, 1970.

Campbell, H.W., Hierarchical ordering of phonetic features as a function of input modality, in Advances in Psycholinguistics, Flores d'Arcais, G.B. and Levelt, W.J.M., Eds., North-Holland, Amsterdam, 1970.

Carroll, J.B., Research in Language teaching: the last five years, Reports of the Working Committees, North east Conference on the Teaching of Foreign Languages, M.L.A. Materials Center, New York, 1966.

Carroll, J.B. and Sapon, S.M., Moderne Language Aptitude Test, Manual, The Psychological Corporation, New York, 1959.

Chomsky, N.A., Three models for the description of language, IRE Transaction on Information Theory, IT-2, 113, 1956.

Chomsky, N.A., Syntactic Structures, Mouton, The Hague, 1957.

Chomsky, N.A., Explanatory models in linguistics, in Logic, Methodology and Philosophy of Science, Proceedings of the 1960 International Congress, Nagel, E., Suppes, P. and Tarsky, A., Stanford University Press, Stanford, 1962.

Chomsky, N.A., Aspects of the Theory of Syntax, MIT Press, Cambridge, Mass., 1965.

Chomsky, N.A. and Miller, G., Introduction to the formal analysis of natural language, in Handbook of Mathematical Psychology, Luce, R.D., Bush, R.R., and Galanter, E., Eds., vol. 2, chap. II, 1963.

Crothers, E., and Suppes, P., Experiments in Second Language Learning, Academic Press, New York, 1967.

Diller, K.C., Generative Grammar. Structural Linguistics and Language Teaching. Newbury, Rowley, Mass., 1971.

Ervin-Tripp, S., An overview of theories of grammatical development, in The Ontogenesis of Grammar. A theoretical Symposium, Slobin, D.I., Academic Press, 1971.

Gold, E.M., Language identification, Information and Control, 10, 447, 1967.

Herriot, P., An Introduction to the Psychology of Language, Methuen, London, 1970.

Hopcraft and Ullman, S., Formal languages and their relation to automata, Reading, Addison-Wesley, USA, 1969.

Horning, J.J., A study of grammatical inference, Technical Report CS 139 Stanford Artificial Intelligence Project, Computer Science Developments, Stanford, 1969.

Johnson, S.C., Hierarchical Clustering Schemes, Psychometrica, 32, 241, 1967.

Joshi, A.K., Kosaraju, S., and Yamada, H.M., String adjunct grammars: I local and distributed adjunction, II Equational representation, null symbols and linguistic relevance, Information and Control, 21, 93-235, 1972.

Krantz, D., Luce, R.D., Suppes, P., and Tversky, A., Foundation of Measurement I, Academic Press, New York, 1971.

Levelt, W.J.M., Psychological Representations of Syntactic Structures, in The Structure and Psychology of Language (in preparation). Available as Heymans Bulletin HB-69-36 Ex, Department of Psychology, Groningen University, 1967.

Levelt, W.J.M., Formele Grammaticals in Linguistick en Taalpsychologie, 3 vol., Kluwer, 1973, in English traslation, Formal grammars in linguistics and psycholinguistics, 3 vol., Mouton, The Hague, 1974.

Masters, J.M., Pushdown Automata and Schizophrenic Language, unpublished report, University of Sydney, 1970.

McNeill, D., The acquisition of language. The study of developmental psycholinguistics, Harper and Row, New York, 1970.

Mesarovic, M.D., Macko D., and Takahara, Y., Theory of Hierarchical, Multilivel, Systems, Academic Press, 1970.

Miller, G.A., and Chomsky, N.A., Pattern Conception, Paper for Conference on Pattern Detection, University of Michigan, 1957.

Miller, G.A., and Chomsky, N.A., Finitary Models of Language Users, in Handbook of Mathematical Psychology, Luce, R.D., Bush, R.R., and Galanter, E., Eds., vol. 2, chap. 13, 1963.

Newell, A., and Simon, H.A., Human Problem Solving, Englewood and Cliffs, Prentice-Hall, 1972.

Newmark, L., and Reibel, D.A., Necessity and sufficiency in language learning, International Review of Applied Linguistics in Language Teaching, 6, 145, 1968.

Peters, P.S., and Ritchie, R.W., A note on the universal base hypothesis, Journal of linguistics, 5, 150, 1969.

Peters, P.S., and Ritchie, R.W., On restricting the base component transformational grammars, Information and Control, 18, 483.

Peters, P.S., and Ritchie, R.W., On the generative power of transformational grammars, Information Sciences, 6, 49, 1973.

Postal, P.M., Limitations of phrase structure grammars, in The structure of language: Readings in the philosophy of language, Fodor, J.A., and Katz, J.J., Eds., Prentice-Hall, Englewood Cliffs, USA, 1964.

Sager, N., Syntactic Analysis of Natural Language, in Advances in Computers, Alt, F., and Rubinoff, M., Academic Press, New York, 1967.

Thorne, A computer model for the perception of syntactic structure, Proc. Royal Society, B. 171, 377, 1968.

Winograd, T., Understanding natural language, Cognitive Psychology, 3, 1, 1972.

Woods, Transition network grammars for natural language analysis, Comm. ACM, 13,591, 1970.