PERGAMON

Language & Communication 20 (2000) 197–253

LANGUAGE
&
COMMUNICATION

www.elsevier.com/locate/langcom

The transformations of transformations

Francis Y. Lin *

St. Hugh's College, Oxford University, Oxford, UK

Abstract

Zellig S. Harris's original idea of transformations has been transformed (changed) several times in Chomsky's work. In this paper I shall explicate all these transformations (changes). I shall argue that, though their motivations are highly understandable, these transformations (changes) are not necessary for understanding the workings of natural languages. In other words, I shall argue that Harris's transformational grammar is quite adequate. I shall also argue that the idea of 'discovery procedures', seen from a novel point view which I shall provide, can indeed lead to important discoveries about the secrets of human language and the human mind.

Keywords: Transformations; Discovery procedures; Evaluation procedures; Universal grammar; Knowledge of grammar

1. Introduction

Generative linguistics, as it is practised by many linguists today, started from the concept of 'transformational grammar'. The concept was first advanced by Zellig S. Harris. Harris thought that there are many transformations in a language, e.g. passivisation. He defined transformations in terms of the distribution of words in a language, and he sought to find the transformations in a language by what have become known as 'discovery procedures'. Over the past forty years, the concept of 'transformations' has undergone many changes. The idea of 'discovery procedures' was heavily criticised by Chomsky and abandoned in the linguistics community. From the very beginning the concept of transformations in Chomsky's work was 'somewhat different' from Harris's (the difference will be explained at the end of Section 4.1 below). Even this already 'somewhat different' concept has been changed many times. Transformations (in Chomsky's work) were transformed into 'principles

* Corresponding author.
E-mail address: francis.lin@st-hughs.oxford.ac.uk

0271-5309/00/$ - see front matter
PII: S0271-5309(00)00004-5

and parameters', which in turn were transformed into 'minimalist principles'. In this paper I shall explicate all these transformations (changes) of Harris's original conception of transformations. I shall argue that, though their motivations are highly understandable, these transformations (changes) are not necessary for understanding the workings of natural languages. In other words, I shall argue that Harris's transformational grammar is quite adequate. I shall also argue that the idea of 'discovery procedures', seen from a novel point view which I shall provide, can indeed lead to important discoveries about the secrets of human language and the human mind.

The structure of this paper is as follows. In Section 2 I shall present Harris's conception of transformational grammar, and explain the so-called discovery procedures. In Section 3 I shall expound Chomsky's criticisms of the procedural approach to language. Section 4 will explain Chomsky's own theory of grammar. Section 5 will provide an examination of Chomsky's argument, and will point out a fundamental error in his linguistic thinking. In Section 6 I shall argue that, once this fundamental error is corrected, we will see that discovery procedures can indeed discover the secrets of human language and the human mind. Conclusions and discussions will be offered in the final section, Section 7.

2. Harris's transformational grammar

2.1. An overview

Harris's original conception of transformational grammar is easy to understand. It was based on two very intuitive ideas. Intuitively, a sentence is either a simple one or a complex one, and a complex sentence is made up from simpler sentences. Harris (1956) called simple sentences 'base' sentences. He proposed that complex sentences are constructed from base sentences by means of 'transformations'. A base sentence can be transformed into a sentence or a phrase; the latter was called the 'transform' of the former. We shall see some examples in due course.

The other intuitive idea was that sentences form types. For example, the two sentences *They sought fame* and *He saw the supervisors* are of the same type $N_1 V\ N_2$ (Harris, 1956, p. 385).[1] The first sentence can be transformed into *Fame is what they sought*, and the second sentence into *The supervisors are (the people) who he saw*. We would not want to say that in this case the original two sentences have undergone two different transformations. Rather, we would want to say that they have undergone the same transformation, which is: $N_1 V\ N_2 \longleftrightarrow N_2$ *is wh-pro-*$N_2 N_1 V$ (*wh-pro-*N_i indicates words like *what, who,* depending on whether N_i refers to persons or objects) [Harris (1956), p. 385]. Sentence types, such as the two just discussed, were called 'sentence forms' by Harris (p. 387), and they were also called 'constructions'

[1] N_1 and N_2 stand for 'noun phrases' and V for 'verb'. Harris (1956, p. 385) actually uses the type $N_1 t\ V\ N_2$, where t stands for 'tense'. But I shall omit information about tense in this paper, for ease of exposition.

(pp. 383–384). So, we can either say that a sentence is transformed into another, or that a sentence form is transformed into another. But in general, a transformation $X \longleftrightarrow Y$ is between sentence forms, as shown in the example just given.

Complex sentences are formed from base sentences by means of transformations. A base sentence is one that is not thus formed. Alternatively, we can say that a base sentence is a sentence formed by a special transformation on the sentence itself. This special transformation is called the 'identity transformation' [Harris (1956), p. 387; (1957), p. 443]. For example, we can say that *The boy came* is transformed from the sentence itself by the identity transformation [Harris (1956), p. 387]. This seemingly superfluous terminology enabled Harris to state his idea of transformational grammar succinctly (see below).

Harris's idea of transformational grammar was this. Every sentence form (or sentence) can be factored (or analysed, decomposed) into transforms of the base sentence forms [Harris (1956), p. 387]. (If a sentence is already a base sentence, it can be seen as the identity transform of itself.) Among all the sentence forms (or sentences) in a language, the base sentence forms (or sentences) form the 'kernel' (p. 388). Kernel sentences are base sentences, they do not really undergo transformations. But we can say that they undergo the identity transformation. So the kernel can also be defined as 'those sentence forms which go into the identity transformation' [p. 388; see also Harris (1966), p. 606]. Kernel sentences are 'the ones into which every sentence can be decomposed via transformations' [Harris (1966), p. 606] (including the identity transformation). In other words, 'All sentences of the language are obtained by applying one or more transformations to one or more kernel sentences' [Harris (1956), p. 388] (including the identity transformation). Thus the grammar of a language consists of a kernel part (or base) and a transformation part. In other words, 'The language structure then consists of a set of kernel sentence forms (or sentences) and a set of transformations' [1956, p. 388; see also Harris (1959), pp. 458–461).[2]

2.2. More on the kernel

The kernel consists of base sentence forms (or sentences) - those which are not formed by transformations (apart from the identity transformation). Kernel sentence forms (or sentences) are the simplest ones. What are the kernel sentence forms in a language, say English? Harris thought that there are only a small number of kernel constructions in English, and they include (cf. Harris (1957), p. 445; Harris (1964a), p. 479; Harris (1964b), p. 484):

N V	A tree fell.
N V N	A rock struck a tree.

[2] Harris's transformational grammar was later developed into *operator grammmar* (Harris, 1988, 1991), to make the grammar more mathematically elegant. See Matthews (1999) for a concise exposition. Mathematical elegance is one thing, the nature of the grammar of a human language is another. I shall not discuss Harris's operator grammar in this paper.

N V P N	A child relies on luck.
N V N N	The man gave Shaw a letter.
N V N P N	A man attributed the picture to Vasari.
N be *A*	The man is good.
N be *P N*	The box is near the corner.
N be D_{loc}	The man is here.
N be *N*	Man is a mammal.
It V_{it}.	It rained. (V_{it} means weather verbs, such as *rain* and *snow*.)
There V_{th} *N*	There's hope. (V_{th} refers to verbs such as *be* and *appear*.)

In addition there are a few minor constructions, such as [Harris, (1957), p. 445]:

N is between *N* and *N*.

Harris (1957) regarded certain combiners as belonging to the kernel (p. 444). The combiners he talked about include items such as *if ... then ...* (p. 427), *and, or, either ... or ...* (p. 428), and *wh-* (for forming relative clauses) (p. 431). But in Harris (1964a,b), these combiners were treated as (elementary) transformations (1964a, p. 480; 1964b, pp. 492–493).

One might propose a slightly different set of kernel sentences (for English). 'Each difference would have to be compensated for by some corresponding alteration in the transformations which operate on [the kernel]' [Harris (1964b), p. 485]. But 'The over-all differences would be small, and the character of the theory would not be affected' (p. 485). In other words, this would not 'make a great deal of difference for the picture of the structure of the language, and even less for the structure of language in general' [Harris (1957), p. 445].

2.3. More on transformations

The kernel is a finite set of base sentence forms. Each kernel sentence form can be subjected to a number of transformations (including the identity transformation). All sentences in a language can be produced by applying transformations to the kernel sentences. The transformations in a language, say English, fall into several major types.

2.3.1. Unary transformations

Unary transformations are 'transformations between two sentence-forms'' [Harris (1965), p. 540]. They 'send a sentence of one grammatical form ... into another grammatical form ...' [Harris (1956), p. 384]. Below is a list of some unary transformations presented by Harris (1965, pp. 540–546):

I like this.	This I like.[3]
He learned a lesson.	He learned his lesson.

[3] This is an example of the transformation: $N_1 V N_2 \rightarrow N_2 N_1 V$. But for ease of exposition I shall only present here the examples, and leave out the exemplified transformations.

The man came.	He came.
It has a value.	It is of value.
He saw the man.	The man was seen by him.
He cut the meat with a knife.	The knife cut the meat.
The whole room laughed.	All those in the room laughed.
He studies eclipses.	He is a student of eclipses.
He kicked the door.	He gave a kick to the door.

2.3.2. Non-sentence-forming transformations

There are transformations which do not transform one sentence into another sentence. Rather, they transform a sentence into a phrase, which can then be used as a part of another sentence. Below are some examples [Harris (1956), pp. 385–388; Harris (1965), p. 551]:

Books are interesting.	interesting books
The bulbs are from Holland.	bulbs from Holland
The bulbs are from Holland.	the bulbs which are from Holland
The book has a value.	the book's value, the value of the book
He read the note.	his reading of the note
The hour is late.	The hour being late [, we left]
Everyone ran forever.	With everyone running forever [we felt something was wrong]
He smiled.	Smiling [, he gave the book back to me]

2.3.3. Binary transformations

Binary transformations combine two sentence (forms) to produce a new sentence (form). Below are some examples [Harris (1965), pp. 546–552]:

The man talked *and* the man drove.[4]
I came *because* he arrived.
I came *while* he arrived.
If he is free, he will go.
The man *who* drove talked.

Binary transformations can also combine a sentence (form) with a transform of another sentence (form). This often happens when the transform acts as an adverbial in the resultant sentence. For example:

I talked + I drove →
I talked while driving [Harris (1965), p. 548].

[4] This is the result of combining two sentences with the combiner *and*. The transformation involved is: $S_1 + S_2 \rightarrow S_1 and\ S_2$. Here I shall only list the resultant sentences and omit the transformations involved.

He smiled + He gave the book back to me →
Smiling, he gave the book to me (1965, p. 551).
Everyone ran forever + We felt something was wrong →
With everyone running forever, we felt something was wrong (1956, p. 386).

According to Harris (1965), the transformations of English fall into the major types just discussed. And this 'seem[s] to be the case for many other language too' (p. 540). In the above we only presented some examples of the transformations in English.

2.4. Discovery procedures

In the last few subsections we have been talking about kernel sentence forms and transformations. We have also seen that kernel sentence forms and transformations were represented by symbols such as *N, V, A, P*, etc. All this is quite intuitive. But Harris was not content with intuitive talk. He wanted to base his study on a firm scientific footing. On this point, Harris had been heavily influenced by Edward Sapir and Leonard Bloomfield; witness his acknowledgement: 'The work here starts off from the distributional (combinational) methods of Edward Sapir and of Leonard Bloomfield, to both of whom I am glad to restate my scientific and personal debt' (1991, p. vi). The methods referred to in this quotation are explained below.

The question Harris wanted to ask was this: How do we know that there are categories such as *N, V, A, P*, etc., that there are kernel sentence forms (as mentioned earlier), and that there are transformations (as discussed above)? Harris's answer was that we can follow certain 'discovery procedures'. In this subsection we shall discuss Harris's 'discovery procedures'.

The flow of speech is not without structure. Rather it is highly structured. At the time Harris worked on syntax, there had already been techniques for breaking the flow of speech into segments, e.g. phonemes, and morphemes. The procedures were mechanical, and the approach was very successful in phonology and morphology (Harris, 1944, 1955). Harris wanted to extend the method to deal with syntax. In this subsection I shall only be concerned with his method for finding the syntactic structure of a language.

'From morpheme to utterance' (Harris, 1946) was probably the most important paper Harris wrote on syntax, another very important paper being Harris (1957). But the idea of procedures of classification, segmentation and substitution, which was to attract heavy attacks from Chomsky, was already clearly illustrated in Harris (1946). To understand Harris's work and also Chomsky's, it is vital to understand these so-called 'discovery procedures'.

In Section 2.3 above we saw symbols such as *N, V, A, P*, etc. They stand for syntactic categories. But how do we know that there are such things in a language (say, English)? And how can we find out how many such things there are in a language? Harris (1946) offered an procedure for answering these questions. The procedure 'consists essentially of *repeated substitution*' (p. 102; italics added). For example, we can substitute *child* for *young boy* in *Where did the — go?* More precisely,

> we take a form *A* in an environment of *C* — *D* and then substitute another form *B* in the place of *A*. If, after such substitution, we still have an expression which occurs in the language concerned, i.e. if not only *CAD* but also *CBD* occurs, we say that *A* and *B* are members of the same substitution-class, or that both *A* and *B* fill the position *C* — *D,* or the like [Harris (1946), p. 102].

For example, both *Where did the child go?* and *Where did the young boy go?* are English sentences, so *child* and *young boy* belong to the same substitution-class.

With this procedure of substitution, we can discover many 'morpheme classes' [Harris (1946), p. 102]. If certain single morphemes occur in the same environments, i.e. they can be substituted for one another, then they belong to the same morpheme class. The syntactic categories we have seen, such as *N* and *V*, etc. are just morpheme classes. For example, *N* is the class of morphemes which 'occur before plural-*s* or its alternants, or after *the* or adjectives', and they include morphemes (in this case words) such as *hotel, gain, one, two* (p. 105). Similarly, *V* is the class of words which occur before *-ed* past or its alternants, or before *-ing*; or after *N should* (*will, might,* etc.). So words like *go, gain, take, think, have, do,* belong to the class *V*. Within a class, we can also distinguish several sub-classes, also using the substitution procedure. For example, we can have V_b for the words *be, appear, become, get, keep*, which occur in a sentence of the type *The stuff will-fresh*; V_c for *stop, try, be,* which occur in *Mac will — walking*; V_d for transitive verbs; V_e for intransitive verbs; V_f for *consider, make,* which go with a sentence like *I'll — this book a best seller*; and so on (p. 106).

Not only can a single morpheme be substituted by another single morpheme, it can also be substituted by a 'sequence of morphemes'. In the latter case, the single morpheme and the sequence of morphemes in question also belong to the same substitution class [Harris (1946), p. 104]. For example, in *Please put the book away*, we can substitute *book* with *brandy bottle, silly green get-up,* etc. (p. 104). In this way, we can discover what kinds of words or sequences of words belong to the same substitution class. For example, *good boy* can be substituted for *man* anywhere; so we have $AN = N$ (pp. 108–109). This equation means that a group formed by an adjective and a noun functions as a noun. Putting an adverb *D* in front of *AN* does not alter this property; so we have $DAN = AN = N$. Similarly, a transitive verb plus a noun is like an intransitive verb: $V_dN = V_e$ (p. 112). [Harris (1946), pp. 109–114] presented many such equations.

Such equations naturally led to the immediate-constituent formulations of (English) syntax. To see this, we only need to reverse the equations. So for example, instead of $DAN = AN = N$, we now have $N = AN = DAN$. Both equations mean that *N, AN,* and *DAN* all belong to the same class (call it *NP*). Then the new equation suggests that an *NP* consists of an *N*, or an *A* plus an *N*, or a *D* plus an *A* plus an *N*. So we can now talk about the immediate-constitute structure of noun-phrases. [Harris (1963), pp. 131–135] listed many ways a noun-phrase can be constructed. For example, a noun phrase can be *N's N,* or *to V,* or *Ving OBJ,* and so on. Similarly, a verb-phrase can also have an immediate-constituent structure. A verb-phrase can be V_o, which takes no object: *exist, sleep*. Or it can consist of a verb belonging to one of certain types plus the appropriate object. The types (subclasses) of verbs include V_n,

taking an N as object: *take (it)*; V_{pn}, whose object is P NP: *rely (on him)*; V_{tv}, whose object is *to V*: *want (to go)*; V_{na}, whose object is *N A*: *consider (him foolish)*; V_{nsa}, whose object is *N S(ubject) A*: *view (him as (being) foolish)*; and so on (p. 136).

Not only do phrases (noun-phrases, verb-phrases, etc.) have internal structure, but so do sentences. The major assertion sentence type has the structure (p. 132):

NP V OBJ[5]

We have seen, in the preceding paragraph, the structure of *NP*, the types of *V* and types of *OBJ* which go with them. (Harris also discussed other sentence types such as yes-no question, wh-question, imperative, etc. But I shall omit that here.)

Thus, by applying the procedure of substitution, we can, Harris hoped, discover an *inventory* of elements (classes and subclasses of words, sequences of classes, etc.), and analyse sentences in terms of them so as to explicate the syntactic structure of sentences. In Harris's words,

> We begin with a set of sentences in a language, each provided with an analysis into constituents; these constituents in turn may be provided with an analysis into sub-constituents, and so on. The constituents are collected into *classes*, and we can say that a given sentence is a case of a particular *sequence* of constituent classes [Harris (1956), p. 383; italics added].

Not only can a procedure be set up to discover the syntactic structure of a sentence, Harris suggested that we can also devise a procedure for finding out whether a sentence is a transform of another sentence. There are various types of transformations, as we have seen in Section 2.3 above. Now, what is the procedure for finding transformations? This question was answered in Harris (1957) (see below).

The relation between a sentence and its transform is a formal relation, which is 'based on comparing the individual co-occurrences of morphemes' [Harris (1957), p. 390]. 'By investigating the individual co-occurrences ... we can define transformation' (p. 390). A procedure based on co-occurrences can then be used to discover transformations. Before going into the details of the procedure based on co-occurrences for finding transformations, let us note that this procedure is not that different from the procedure of substitution we just discussed: the procedure of substitution is just a procedure based on co-occurrences. To see this, first note that classes such as *N*, *V*, etc., can be defined in terms of co-occurrences:

> some morphemes have very similar (though not identical) sets of co-occurrents; thus, the set of co-occurrents of *cloth* - e.g. *The ( ) tore, The ( ) was torn, Get me a ( ) quick* — may have many morphemes in common with the set for *paper*,

[5] Harris's formulation was actually *APP NP APP V OBJ APP*, where *APP* refers to 'sentence-appendices', such as *at such times, quietly, in general, They being late (we left)*, etc. [Harris (1963), pp. 132, 136–137]. I omit *APP* here for simplicity. Note also that, when *V* is intransitive, i.e. when it is V_o, it cannot have an object. This was clear in Harris's discussion, though it was not shown in the formula just mentioned.

certainly many more than with the set for *diminish*. This suggests that morphemes can be grouped into classes in such a way that members of a class have rather similar sets of co-occurrents, and each class in turn occurs with specific other classes to make a sentence structure. ... *cloth* and *paper* both occur, say, in the environment *the ( ) is* (i.e. after *the* and before *is*), where *diminish* does not appear; we call this class N. And *diminish* and *grow* both occur, say, in *It will ( )*, where *paper* and *cloth* do not; we call this class V [Harris (1957), p. 391].

Secondly, note that the constituent structure can also be defined in terms of co-occurrences. For example,

Given a construction (which is recognised by its place within larger constructions, up to a sentence), we can see some relations or different statuses among the participating classes by noting details of class occurrence for each class. For example, given the A N construction (*slight hopes*), we note that it is substitutable for N alone but not for A alone: both *slight hopes* and *hopes* occur in *Their ( ) faded*. [So, we have] the constructional equivalence A N = N [Harris (1957), p. 393].

Now, the idea of co-occurrence can be extended to define and discover transformations:

If two or more constructions (or sequence of constructions) which contain the same n classes (whatever else they may contain) occur with the same n-tuples of members of these classes in the same sentence environment (see below), we say that the constructions are transforms of each other, and that each may be derived from any other of them by a particular transformation. For example, the construction N v V N (a sentence)[6] and N's V*ing* N (a noun phrase) are satisfied by the same triples of N, V, and N (*he, meet, we; foreman, put up, list;* etc.); so that any choice of members which we find in the sentence we also find in the noun phrase and vice versa: *He met us, his meeting us ...; The foreman put the list up, the foreman's putting the list up* ...Where the class members are identical in the two or more constructions we have a reversible transformation, and may write N_1 v V $N_2 \longleftrightarrow$ N's V*ing* N_2 (and the set of triples for the first = the set for the second) [Harris (1957), p. 394].

Having got a definition of transformation, how can we proceed to discover the transformations in a language? 'To establish the transformations in any given language', stated Harris (1957), 'we need methods, and if possible an organised *procedure*, for seeking ... what constructions may contain identical co-occurrences; these methods should if possible be general, but additional ones may be based on special

[6] (This is my note.) The small *v* here stands for 'auxiliaries', such as *may* and *will*. See Harris (1957, p. 407).

features of the language' (p. 397; italics added). The procedure consists of finding constructions containing the same classes, collecting sets of tentative transforms, checking whether the co-occurrences are identical in the suspected transforms by asking native speakers for judgement and using the methods of 'eliciting', etc. (pp. 397–401).

With this procedure Harris (1957) expected that all transformations in a language can be found. He further noticed that some transformations are combinations of certain more basic transformations. The most basic transformations, called 'elementary transformations', are those which are 'not obtainable by combining the effect of two or more other simpler transformations' (p. 442). Thus, 'The existence of elementary transformations makes it possible to regard all transformations as compoundings of one or more elementary ones' (p. 442). In this light, all sentences can be seen as made up from a set of very simple sentences, or 'kernel sentences', by a set of transformations (pp. 441–447). At this stage, we have arrived at Harris's transformational grammar, which we have outlined in Section 2.1 above.

2.5. The properties of transformations

We have been discussing Harris's transformational grammar and his methods for discovering it. The transformational grammar consists of kernel sentences and transformations. But what is the nature of such structures? In this subsection we shall discuss Harris's answer to it.

2.5.1. Distributional structure

In Section 2.4 above we saw procedures based on co-occurrences, such as the procedure of substitution, for discovering grammatical categories, immediate-constituent structures, and transformations. Another way of expressing the idea of co-occurrence is to use the term 'distribution'. As Harris stated, 'The distribution of an element will be understood as the sum of all its environments. An environment of an element A is an existing array of its co-occurrents, i.e. the other elements, each in a particular position, with which A occurs to yield an utterance' [Harris (1954), p. 775]. A language can be described by grammatical categories, immediate-constituent structures, and transformations. We can put this in another way, i.e.: 'a language can be described in terms of a distributional structure, i.e. in terms of the occurrence of parts (ultimately sounds) relative to other parts' (p. 775). For this reason, the methods discussed in Section 2.4 above have often been referred to as 'distributionalism'.

2.5.2. The psychological reality of distributional structure

Is the distributional structure of a language, discovered by certain procedures, psychologically real? Harris's (1954) answer was positive. He thought that 'A reasonable expectation is that the distributional structure should exist in the speakers in the sense of reflecting their speaking habits' (pp. 778–779). On this point, Harris was echoing Jespersen's view. According to Jespersen, there are many 'sentence types' in a language [Jespersen (1924), pp. 18–29], which are very similar to Harris's sentence

structures and transformations; the sentence types are a set of grammatical habits [Jespersen (1924), pp. 19, 25 and 29]; and they 'come into existence in the speaker's subconsciousness as a result of his having heard many sentences' [Jespersen (1924), p. 20]. [See Lin (1998) for a detailed exposition.] But Harris went further than Jespersen in providing an explanation of how the speaker forms his speaking habits and how the distributional structure comes into his mind. Harris suggested that the speaking habits 'are presumably based, like the linguist's analysis, on the distributional facts' [Harris (1954), p. 780].

Distributional structure, which is expressed by syntactic categories, immediate-constituents and transformations, exists in language [Harris (1954), p. 778]. It also exists in the speakers (see the preceding paragraph). So, transformations are real. A transformation 'once established is not normally falsifiable by further research; and the existence of the transformational relation and the general properties ... of the set of sentences under this relation, and of the set of transformations, are not shaken by individual problems concerning transformations' [Harris (1965), pp. 553–554]. This is a major difference between Harris's transformations and Chomsky's; we shall come to this point in Section 4.1 below.

2.5.3. The creativity of language

The idea that language is creative had already been expressed by Jespersen (1924). Jespersen distinguished between 'formulas' and 'free expressions'. Formulas are expressions for which 'no one can change anything in them' [Jespersen (1924), p. 18]. In contrast to formulas, most expressions we hear or utter are not repetitions of what we have heard or uttered before: they are 'free expressions' (p. 19). A language consists mainly of free expressions — this is why language is said to be creative. Jespersen explained the creativity of language in terms of 'sentence types', which we mentioned in Section 2.5.2 above. According to Jespersen, free expressions are created after sentence types (p. 20), by inserting 'the words that fit the particular situation' (p. 19). Sentence types are thus generative. From any sentence type a speaker can produce many, and potentially infinitely many, sentences of the same pattern. In Jespersen (1937) he presented a system of syntax based on the notion of sentence type. See Lin (1998) for a detailed exposition.

Harris's explanation of language creativity was similar to Jespersen's. Jespersen's explanation was based on sentence types, while Harris's was based on kernel sentences and transformations. Thus, Harris stated:

> the kernel (including the list of combiners) is finite; all the unbounded possibilities of language are properties of the transformational operations. This is of interest because it is in general impossible to set up a reasonable grammar or description of the language that provides for its being finite. Though the sample of the language out of which the grammar is derived is of course finite, the grammar which is made to generate all the sentences of that sample will be found to generate also many other sentences, and unboundedly many sentences of unbounded length [Harris (1957), p. 448].

3. Chomsky on the taxonomic approach

Chomsky (1975a) referred to the approach of discovering the grammar of a language using mechanical discovery procedures as the 'taxonomic approach'. In Chomsky's words, 'The taxonomic approach, which encompassed all theoretical studies of language known to me that were concerned with the fundamental problem of projection and with the precise characterisation of the general concepts of linguistics, sought to define such concepts as "phoneme," "morpheme," "category," etc., by procedures of segmentation and classification applied at successive levels of generality, providing what might reasonably be called "a grammar of lists"' (p. 32).

From the very beginning of generative linguistics, Chomsky was completely dismissive of the taxonomic approach to grammatical analysis. The heart of the taxonomic approach to grammatical analysis is the concept of discovery procedures. A major part of Chomsky's early writings on linguistics was devoted to exposing the problems of discovery procedures. The importance of Chomsky's argument against discovery procedures cannot be overstated, for it was on this basis that Chomsky proposed his famous 'evaluation procedures' to replace 'discovery procedures' (see Sections 4.1 and 4.2 below). Understanding this argument is absolutely essential for understanding Chomsky's linguistic theory. In this section I shall present Chomsky's argument in detail.

3.1. No general procedures for discovering grammars

The taxonomic linguists tried to find some general, mechanical procedures for discovering the grammar of a language. But as Chomsky (1964a) pointed out, there were no such general procedures: 'In constructing a grammar for a particular language, one of the decisions to be made concerning each class of sentence is whether to consider them to be kernel or derived sentences. I know of no general mechanical procedure for arriving at the answer to this question, just as I am unacquainted with any *general* ... mechanical procedure for arriving at a phonemic, morphological, or constituent analysis for the sentences of a language' (p. 223). Instead, a completely different approach is needed, proposed Chomsky: 'To answer all of these questions, we must apparently do what any scientist does when faced with the task of constructing a theory to account for particular subject matter — namely, try various ways and choose the simplest that can be found' (p. 223).

3.2. Discovery procedures themselves need to be justified

We have seen in the preceding subsection that there were (yet) no general procedures for arriving at correct grammatical analyses. The situation with discovery procedures was rather depressing. In Chomsky's words, 'In brief, it is unfortunately the case that no adequate formalisable techniques are known for obtaining reliable information concerning the facts of linguistic structure (nor is this particularly surprising). There are, in other words, very few reliable experimental or data-processing procedures for obtaining significant information concerning the linguistic intuition of the native speaker' [Chomsky (1965), p. 19]. If there are no adequate, general

procedures to give us correct results, then what we will need is criteria for developing discovery procedures and choosing among them. As Chomsky states:

> It is important to bear in mind that when an operational procedure is proposed, it must be tested for adequacy (exactly as a theory of linguistic intuition — a grammar — must be tested for adequacy) by measuring it against the standard provided by the tacit knowledge that it attempts to specify and describe. Thus a proposed operational test for, say, segmentation into words, must meet the empirical condition of conforming, in a mass of crucial and clear cases, to the linguistic intuition of the native speaker concerning such elements. Otherwise, it is without value. The same, obviously, is true in the case of any proposed operational procedure or any proposed grammatical description [Chomsky (1965), p. 19].

Discovery procedures themselves therefore need to be justified and evaluated. Elsewhere, Chomsky put this same idea as follows:

> A common view appears to be that to justify a grammatical description it is necessary and sufficient to exhibit some explicit procedure (preferably, purely formal) by which this description could have been mechanically constructed from the data. This view I find very strange. Why, first of all, should it be required that there be some relatively straightforward, completely general procedures for arriving at grammars from the data? ... Why, furthermore, should exhibiting such a procedure be considered a justification of the result to which it leads? There are undoubtedly perfectly general and straightforward procedures for arriving at the most wild descriptions — e.g. we can define a *morpheme* in a perfectly general, straightforward, and formal way, with no mixing of levels, as any sequence of three phonemes. *Clearly, it is necessary somehow to justify the procedure itself* [Chomsky (1964a), p. 241, n. 27; italics added].

Not only do discovery procedures need to be justified and evaluated from a conceptual point of view, but taxonomic linguists were actually proposing evaluative criteria for the procedures, and hence for the grammars discovered by the procedures. Chomsky (1957, p. 52, n. 3), after citing Hockett (1952) and Wells (1947), remarked:

> Although discovery procedures are the explicit goal of these works, we often find on careful examination that the theory that has actually been constructed furnishes no more than an evaluation procedure for grammars. For example, Hockett states his aim in "A formal statement of morphemic analysis" as the development of "formal procedures by which one can work from scratch to a complete description of the pattern of a language" (p. 27); but what he actually does is describe some of the formal properties of a morphological analysis and then propose a "criterion whereby the relative efficiency of two possible morphic solutions can be determined: with that, we can choose the maximally efficient possibility, or arbitrarily, any one of these which are equally efficient but more efficient than all others" (p. 29) [Chomsky (1957), p. 52, n. 3].

3.3. In conflict with scientific inquiry

What the taxonomic linguists wanted to do was to devise mechanical procedures, which when applied to linguistic data, would produce the grammar of a language directly. As Chomsky (1957) said: 'As I interpret most of the more careful proposals for the development of linguistic theory ... they attempt to state methods of analysis that an investigator might actually use, if he had the time, to construct a grammar of a language directly from the raw data. I think that it is very questionable that this goal is attainable in any interesting way, and I suspect that any attempt to meet it will lead into a maze of more and more elaborate and complex analytic procedures that will fail to provide answers for many important questions about the nature of linguistic structure' (pp. 52–53). According to Chomsky, a grammar should be treated as 'a theory of linguistic intuition' [Chomsky (1965), p. 19]. 'There are few areas of science', pointed out Chomsky, 'in which one would seriously consider the possibility of developing a general, practical, mechanical method for choosing among several theories, each compatible with available data' (pp. 52–53). Chomsky (1959) expressed the same idea as this:

> But it is a rather peculiar idea to require that justification of a grammar must take the form of a demonstration that this grammar could, in principle, have been arrived at by applying direct mechanical procedures to the data, procedures, which are, furthermore, nearly unique in the result they produce and of quite general applicability. Such a justification of systems as abstract and theoretical as grammars has never been thought possible in any other field, and there seems to be little reason to impose this very extreme condition on justification in linguistics, at least in this very early stage in its development (p. 215).

And Chomsky (1964a) repeated the same point at another place, when he said:

> But a methodological requirement that the concepts of the theory of language be so chosen that simple and direct procedures of analysis can be used to determine their application in particular cases is no more justifiable in linguistics than in any other science, and it is likely to have the effect of distorting our conception and understanding of linguistic structure (p. 245).

3.4. Failure to answer many important questions

As quoted in the preceding subsection, Chomsky (1957) stated that 'I suspect that any attempt to meet it [i.e. the goal of arriving at the grammar of a language by discovery procedures] will lead into a maze of more and more elaborate and complex analytic procedures that will fail to provide answers for many important questions about the nature of linguistic structure' (pp. 52–53). What important questions did Chomsky refer to here?

There seem to be at least three fundamental questions concerning the nature of linguistic structure Chomsky had in mind. One is the creativity of language, another is the nature of linguistic knowledge, and the third is language acquisition.[7] I shall discuss these three questions in turn.

3.4.1. The creativity of language

The speaker of a language can utter a sentence he has never uttered or heard before; he can also judge whether an arbitrary sentence is grammatical or not. This shows that language is creative. And this is a fundamental feature of language. So any adequate linguistic theory must be able to account for the creativity of language. But 'structural linguistics has rarely been concerned with the "creative" aspect of language use, which was a dominant theme in rationalistic linguistic theory. It has, in other words, given little attention to the production and interpretation of new, previously unheard sentences - that is, to the normal use of language' [Chomsky (1965), p. 205, n. 30]. What taxonomic linguistics had been typically concerned with was 'the much narrower problem of constructing several inventories of elements in terms of which utterances can be represented, and had given little attention to the rules that generate utterances with structural descriptions' [Chomsky (1961b), p. 223].

To be fair, might theories of immediate-constituent analysis not be interpreted as phrase structure grammars, which are indeed generative (i.e. which can generate an infinity of sentences)? Chomsky's answer to this question was affirmative. But he stressed that 'the suggestion that the various theories of immediate constituent analysis might be interpreted as generative, phrase structure grammars: as in Chomsky, [1956, 1964a], or Postal [1964a] certainly goes beyond what is explicitly stated by linguists who have developed these theories, and very likely beyond their intentions as well' [Chomsky (1965), p. 205, n. 30].

How can the creativity of language be explained? Chomsky proposed that the explanation lies in finding a universal grammar among all human languages. We shall see this in Section 4.5 below.

3.4.2. The language learning device

The taxonomic approach assumes that the child is equipped with certain mechanical procedures (and probably also procedures of generalisation, analogy, etc.), and that by applying the procedures to linguistic data presented to him he will acquire the grammar of his language. So the language acquisition device is assumed to be just a set of such procedures. But Chomsky believed that such procedures are useless in accounting for how the child acquires the language. See Section 4.6.1 below for details.

3.4.3. Knowledge of language

The taxonomic approach assumes that the knowledge of language the child has is acquired using procedures of segmentation, classification, generalisation, analogy,

[7] Chomsky seemed also to have in mind the question of how to define grammaticalness and degrees of grammaticalness. See Note 10 below.

etc. But Chomsky argued that linguistic knowledge cannot be explained in this way. Rather, linguistic knowledge is, thought Chomsky, largely innate. See Section 4.4 below for details.

4. Chomsky's theory of grammar

Section 3 can be summarised as follows. According to Chomsky, the taxonomic approach is simply not feasible: in no other scientific inquiry would one talk about a general, mechanical, procedure for finding a theory; the procedures themselves have to be justified, evaluated. But feasibility apart, the taxonomic approach fails to, and cannot, explain many important questions about the nature of linguistic structure, such as the creativity of language, the acquisition of language, and the nature of linguistic knowledge. So, what should the linguist do?

4.1. A change of aspect

We have seen the problems Chomsky identified with the taxonomic approach to theory of grammar. In addition, the taxonomic methods 'are of such incredible complexity that even if they were somewhere near successful, no one would seriously consider the possibility of actually applying them in discovering a grammar. Furthermore, if the researcher gains a sudden and useful insight into the structure of language, he will surely not reject it because it was not arrived at by any procedure' [Chomsky (1959), p. 215]. Why do we have to discover grammars *directly,* by using mechanical procedures, especially when mechanical procedures do not seem to offer any hope of reaching adequate grammars? Why can't we approach grammars *indirectly*, by constructing various grammars in our heads, as it were, and then devising ways of choosing among them, especially when such an indirect method enables us to gain a great deal of insights into the nature of linguistic structure? After all, this is what a scientist would do in other scientific inquires! (see Section 4.6.2 below). What Chomsky thought that linguists needed was a 'change of aspect', in Wittgenstein's terminology. In Chomsky's own terms, what was needed was a 'far-reaching conceptual change' [Chomsky (1975a), p. 30]. This change of aspect was the fountain of Chomsky's whole linguistic thinking.

To explicate this change of aspect, I shall quote extensively Chomsky's own, detailed, and (I believe) accurate account, adding only some brief explanations. Chomsky (1975a) said:

> While working on "discovery procedures" for linguistics, I was troubled by a number of nagging doubts. In the first place, if (as I assumed) refinements with no far-reaching conceptual change would suffice to provide a complete and formalised system of procedures, then linguistics as a theoretical field would have a kind of "terminal character". That is, by applying these methods, we would be able to derive mechanically the grammar of any language, given an adequate corpus. Obviously, it must be possible to achieve such a result; every

child serves as an "existence proof," in that he acquires knowledge of his language, and the knowledge acquired is, to a very good approximation, identical to that acquired by others on the basis of their equally limited and somewhat different experience. Nevertheless, it seemed impossible that grammars derived by methods of the sort under investigation could really express and characterize knowledge of language, which posed innumerable mysteries. My own work on Hebrew, though only rudimentary beyond the morphophonemic level, sufficed to suggest to me that something central was missing.[8] The failure of inductive, data-processing procedures at the syntactic level became more obvious the more I worked on the problem [Chomsky (1975a), p. 30].

By 1952, I was working on generative grammar of English, and shortly obtained results that I found quite exciting, though *they were entirely divorced from the system of procedures of analysis* on which I was working at the same time; in particular, results on the system of auxiliary verbs in simple declaratives, interrogatives, negatives, passives, and on complex verb constructions such as "want (NP) to VP," "consider NP (to be) Predicate," etc. As in the case of my earlier work on morphophonemics of Hebrew, it was possible, so it became clear, to discover systems of rules that made sense of the distribution of forms, principles that served to explain a collection of superficially chaotic and anomalous facts. In this case, too, investigation led to more abstract underlying structures that were far removed from anything that might be obtained by systematic application of procedures of analysis of the sort that I was investigating[9] [Chomsky (1975a), p. 31; italics added].

Now, here was the change of aspect:

It became increasingly clear to me that the methodological limitation to procedures based on substitution, matching, and similar "taxonomic" operations was arbitrary and unwarranted. *One might approach the problem of projecting a corpus to a language of grammatical sentences in an entirely different way, with a procedure for evaluating a completed system of categories rather than a procedure*

[8] (This is my note, not Chomsky's.) To understand this point, it is useful to have a look at the following remark by Chomsky:

> But the elements that I was led to postulate in studying the generative grammar of Hebrew were plainly not within the range of such [taxonomic] approaches. However refined, these were essentially procedures of segmentation and classification. They were designed to isolate classes of phones, sequences of these classes, classes of these sequences, etc., until ultimately, sentences are characterized in terms of their constituents. But the elements that were needed in the optimal grammar simply did not have this character. *They were not classes, sequences of classes, or anything of the sort, but were simply abstract elements forming strings that could be mapped into phonetic representation by deeply ordered rules of considerable generality* [Chomsky (1975a), pp. 29–30; italics added].

See also the discussion of Chomsky (1964a) below.

[9] (This is my note.) The categories Chomsky used to describe Hebrew grammar, such as *NP, VP*, etc. were not classes, sequences, or sequences of classes, *discovered* using some mechanical procedures. See the discussion of Chomsky (1964a) below. See also Note 8 above.

for constructing these categories step by step by taxonomic methods. ...Two approaches to the specific problem of defining syntactic categories were thus counterposed: a constructive, taxonomic approach and an alternative, no less rigorous or formalisable, that was concerned essentially with the properties of a completed solution [Chomsky (1975a), pp. 31–32; italics added].

It was clear at once that the problem of defining syntactic categories was only a special case. The taxonomic approach, which encompassed all theoretical studies of language known to me that were concerned with the fundamental problem of projection and with the precise characterisation of the general concepts of linguistics, sought to define such concepts as "phoneme," "morpheme," "category," etc., by procedures of segmentation and classification applied at successive levels of generality, providing what might reasonably be called "a grammar of lists." Alternatively, one might try to define the notion "grammar" directly in terms of a set of primitive notions applicable to a corpus of data; the phonemes, morphemes, categories, etc., would be the elements that appear in the highest-valued grammar of the appropriate form meeting the empirical conditions determined by application of the primitive notions to a corpus of data. *There would be no reason to regard phonemes, morphemes, categories, and other elements to be segments, classes of segments, sequences of classes, sets of phenomenal properties (e.g. phonetic distinctive features), and so on. Rather, they would be elements in various abstract systems of representation* [Chomsky (1975a), p. 32; italics added].

Given a system of syntactic categories of various levels of abstractness, selected by an evaluation measure, *it seemed possible to make inroads into defining "grammaticalness" and "degree of grammaticalness" ("deviation from grammaticalness").*[10] One might further consider the system of syntactic categories to be one of several "levels of representation," each providing certain descriptive mechanisms that are in principle available for grammar construction. *The language and its structure, then, might be determined, given an adequate corpus, by a system of general principles of language structure that characterise linguistic levels, the form of possible grammars, the structures (systems of levels) generated by a grammar, and the evaluation procedure for selecting among such grammars.* In this way, I gradually come to believe, one might not only solve the "inductive problem" (now abandoning all inductive methods, in a strict sense), but also, more important, develop and justify grammars with a range of explanatory power that clearly escaped the limits of taxonomic theory [Chomsky (1975a), p. 32; italics added].

According to Chomsky, the change of aspect took place no later than 1953:

[10] Cf. Note 7 above.

> By 1953, I had abandoned any hope of formulating taxonomic "discovery procedures" and turned my attention entirely to the problems of generative grammar, in theory and in application. It was at that point that I began writing LSLT,[11] bringing together and extending the work I had begun on various aspects of generative grammar, but now *with conviction* as well as enthusiasm [Chomsky (1975a), p. 33; italics added].

To help understand the above change of aspect, it is useful to look at an essentially identical account Chomsky (1964a) gave in 'A transformational approach to syntax' (originally presented in 1958). That account illustrated the change of aspect by a discussion of the nature of syntactic rules, and it went as follows. Suppose that we try to discover the immediate-constituent structure of sentences, using classes, sequences of classes, etc. Then we shall find categories, phrases, and sentences. The constituents of a sentence will be very likely to be something like an *NP* plus a *VP*. An *NP* may consist of a *that* and another sentence, it may also be made up by a *to* and a *VP*, and so on. A *VP* may consist of just a verb, or a verb plus an *NP*, and so on. So we can express this information in the form:

$$Sentence \longrightarrow NP + VP$$
$$NP \longrightarrow that + Sentence$$
$$NP \longrightarrow to + VP$$
$$NP \longrightarrow T + N \text{ (Here } T \text{ means article)}$$
$$VP \longrightarrow V$$
$$VP \longrightarrow V + NP$$

These rules are very similar to the kernel sentences and transformations Harris discussed. 'So far', remarked Chomsky, 'we have done nothing more than modify Harris' "Morphemes to Utterance" procedures (generalised to account for selection, in accordance with his intentions, in fact showing how these ideas can provide us with a grammar which generates the sentences of the language in a uniform way, with a structural description automatically provided for each generated utterance)' [Chomsky (1964a), p. 215].[12] But notice:

> This modification, however, suggests *a shift of interest* that leads to some essential re-interpretation of the basic notions of constituent analysis. We now regard the goal of linguistic analysis of a particular language (on the IC-level) as the formulation of a set of rules of the form $X \longrightarrow Y$, such that each sentence (and no nonsentence) has a derivation from *Sentence* in terms of these rules, and such that a structural description of the generated sentences ... is uniquely constructable from the derivation. *We are not at the moment concerned with the*

[11] (This is my note.) LSLT refers to Chomsky's *The Logical Structure of Linguistic Theory*, i.e. Chomsky (1975a).

[12] Harris wanted to discover the structure of sentences using certain mechanical procedures. Chomsky here said that we can give a structural description of each sentence, not by using mechanical procedures, but by developing a generative grammar. The aims were the same, but the approaches were different.

> *question how the rules are discovered; we do not, for example, require that they be arrived at by some sort of substitution procedure, or by study of suprasegemental features, inflections, and so on* [Chomsky (1964a), pp. 215–216; italics added, except *Sentence*]

To summarise, according to Chomsky (1959), there should be 'a shift of interest from linguistic methodology to linguistic theory' (p. 211), i.e. a change of aspect. We should give up the hope 'of finding simple and direct mechanical procedures for the discovery of the optimal grammar of a language' (p. 211). Rather, we should construct grammars ourselves, based on intuition, experience or skill, and then choose among them. We still use symbols such as *NP, Sentence, VP,* etc., but they are nothing like classes or sequences of classes of concrete elements, discovered by some mechanical procedures. They are rather things we invented ourselves for certain theoretical purposes. In other words, 'They are simply elements in a system of representation which has been constructed so as to enable us to characterise effectively the set of English sentences in a linguistically meaningful way' [Chomsky (1964a), p. 216].

This is a good place to add a discussion of the differences between Harris's transformations and Chomsky's. As we saw in Section 2.5.2 above, distributional structures, such as kernel sentences and transformations, are regarded as real entities, existing in language and in the speaker. But the status of kernel sentences and transformations in Chomsky's theory was rather different. Kernel sentences and transformations in Chomsky's theory are not real entities, they are mere abstract 'elements in a system of representation which has been constructed so as to enable us to characterise effectively the set of English sentences in a linguistically meaningful way' [Chomsky (1964a), p. 216; quoted above]. Harris's kernel sentences and transformations are based on classes, sequences, and sequences of classes, but 'The approach that I will describe here', Chomsky made it clear, 'bears little formal resemblance to this conception' [Chomsky (1961a), n. 23]. The terms 'kernel sentence' and 'transformation' are 'used here in a sense somewhat different from (Harris's) and in a different general framework' [Chomsky (1964a), p. 223, n. 23]. Thus, the term 'kernel sentence' was 'borrowed from Harris but necessarily given a somewhat different sense' [Chomsky (1975a), p. 44]. And 'In LSLT transformations are understood in a very different sense; it probably would have been preferable to select a different terminology instead of adapting Harris's in this rather different context' [Chomsky (1975a), p. 43]. Harris's transformations, as we have seen in Section 2.5.2 above, are not normally falsifiable by further research once established, because they are seen as real entities. By contrast, Chomsky's transformations are theoretical constructs, so 'a transformation is definitely falsifiable by further research, since a transformation is not the statement of a particular observed co-occurrence relation but rather forms part of a system, a generative grammar, which as a whole has certain empirical consequences' [Chomsky (1975a), p. 44]. To make the difference more vivid, Chomsky added: 'Such a system may be refuted by new evidence or may be replaced by a more highly-valued grammar with a different choice of transformations and other rules. No matter how well established a transformation appears today,

tomorrow's discoveries may reveal that it was entirely misconceived' [Chomsky (1975a), p. 44].

4.2. Evaluation procedures

According to Chomsky, we do not know what the grammar of a language is — it needs to be discovered; and we cannot hope to discover the grammar by using some mechanical procedures. How can we proceed? Chomsky (1959) concluded that we must make a change of aspect in linguistics. He urged us (i.e. linguists) to abandon the approach of 'formulat[ing] precise procedures which will, in a mechanical, step-by-step manner, lead to the optimal grammatical description for an arbitrary language' (p. 218). 'This is hardly surprising', remarked Chomsky, 'It would be strange indeed if it were possible to develop methods for discovering something whose nature and structure is so complex and still so little understood as a grammar of a language' (p. 218). Instead, we should take a completely different approach. We can devise some abstract syntactic categories, syntactic rules, etc. to construct grammars. We do not need to 'consider the symbols *NP, Sentence, VP*, and so forth, that appear in these rules to be names of certain classes, sequences, or sequences of classes, and so on, of concrete elements' [Chomsky (1964a), p. 216]. 'We are not at the moment concerned with the question how the rules are discovered; we do not, for example, require that they be arrived at by some sort of substitution procedure, or by study of suprasegemental features, inflection, and so on' (p. 216). The syntactic categories and rules we employ are 'simply elements in a system of representation which has been constructed so as to enable us to characterise effectively the set of English sentences in a linguistically meaningful way' (p. 216). In short, 'One may arrive at a grammar by intuition, guess-work, all sorts of partial methodological hints, reliance on past experience, etc. ' [Chomsky (1957), p. 56].

Now, isn't this new approach completely subjective and arbitrary? It seems that according to this new approach we can construct the grammar of a language in whatever way we like, and that we can construct as many different grammars for a language as we can. But surely we would not want to say that any wild grammar thus constructed will be an adequate grammar for a language. So evidently we must have some criteria for judging the adequacy of grammars and for choosing among grammars. It is obvious that if there were no such criteria, then constructing grammars would have no meaning at all. Hence, it is absolutely essential to formulate such criteria to go hand-in-hand with the construction of grammars. What is necessary in linguistic theory is then an 'evaluation procedure' — Chomsky's term for the criteria for judging and choosing among grammars. According to Chomsky (1957), 'it is unreasonable to demand of linguistic theory that it provide anything more than a practical evaluation procedure for grammars' (p. 52). Even just providing an evaluation procedure 'is still strong enough to guarantee significance for a theory that meets it' (p. 53). But what does an evaluation procedure consist of, and how can one know what is needed in an evaluation procedure?

Chomsky (1965) talked of various 'evaluation measures'. For example, 'minimisation of feature specification' or 'the measure based on abbreviatory notations'

(p. 39). Whether the rules of a grammar should be unordered or ordered in some specific way could be an evaluation measure. 'Length' may also be a measure of evaluation (pp. 42 and 45). 'Simplicity' may also be taken as an evaluation measure (pp. 37–40). But this measure is quite problematic. The reason is that 'There is no known absolute sense of "simplicity" or "elegance," developed within linguistic theory or general epistemology' (p. 39).

What is the relationship between evaluation measures and evaluation procedures? Chomsky did not answer this question clearly. But the text suggests that an evaluation procedure consists of one or more evaluation measures. I shall adopt this interpretation here. (Adopting a different interpretation would not affect the following discussion.) I shall also sometimes not distinguish between evaluation measures and evaluation procedures, as Chomsky did.

Our (i.e. the linguists') task would be relatively easy if we knew what an evaluation procedure is before we set off to find the grammar for a language. In that case, we would simply need to apply the evaluation procedure to the many possible grammars and choose among them accordingly. But Chomsky (1965) stressed that there is no ready-made evaluation procedure: 'It must first of all be kept clearly in mind that such a measure is not given a priori, in some manner' (p. 37). The reason is that there can be various evaluation procedures, and they would select different grammars and would lead to different conclusions about the nature of linguistic structure and language learning. As Chomsky put: 'Given primary linguistic data D, different choices of an evaluation measure will assign quite different ranks to alternative hypotheses (alternative grammars) as to the language of which D is a sample, and will therefore lead to entirely different predictions as to how a person who learns a language on the basis of D will interpret new sentences not in D' (p. 37). What this means is that 'choice of an evaluation measure is an empirical matter' (p. 37).

Now, it seems that we are in deep trouble. We want to select among possible grammars, which are all compatible with the primary linguistic data, by using an evaluation procedure. But now we are told that there are various evaluation measures and that we have to find what they are and select among them. This effectively means that we must select among possible evaluation procedures. It then seems that we need a meta-evaluation procedure for selecting evaluation procedures. Would we not require a meta-meta-evaluation procedure for selecting meta-procedures? If the answer were yes, then we would have in front of us an endless regression.

But luckily, we do not need second-order (or above) meta-evaluation procedures: we can formulate *the* meta-evaluation procedure for selecting, or constructing, an evaluation procedure. One major meta-criterion for constructing an evaluation measure for grammars is 'that of determining which generalisations about a language are significant ones; an evaluation measure must be selected in such a way as to favor these' [Chomsky (1965), p. 42].

The whole issue of constructing an evaluation measure now turned on whether it allows 'significant generalisations'. Chomsky argued that 'minimisation of feature specification' and the measure based on abbreviatory notations mentioned just now do lead to some significant generalisations (pp. 43–45). These generalisations imply that 'a person learning a language ... will select grammars containing these

generalisations over other grammars that contain different sorts of generalisations' (p. 45).

It is apparent that constructing evaluation measures or procedures is by no means an easy matter. We do not know what specific evaluation measures or procedures are needed. All we know is that they must lead to 'significant generalisations'. Choice of an evaluation measure or procedure is, to repeat what Chomsky emphasised, 'an empirical matter'.

4.3. From evaluation procedures to universal grammar

Recall that the idea of an evaluation procedure is to choose among a store of possible grammars. This idea was expressed in Chomsky (1957), and it was argued there that 'it is unreasonable to demand of linguistic theory that it provide anything more than a practical evaluation procedure for grammars' (p. 52). It looks as if the primary task in linguistic theory was to provide an evaluation procedure. But this did not become the primary task in generative linguistics. Instead, generative linguistics became almost exclusively the search for universal grammar. Why did this happen?

This is to be explained by a shift in the conception of the nature of linguistic theory construction. Chomsky (1957) was primarily concerned with how *the linguist* can arrive at the grammar of a language. But Chomsky (1965) devoted a great deal of attention to the question how *the child* can acquire (arrive at, or construct) the grammar of a language, given the linguistic data exposed to the child. These two conceptions are by no means in conflict with each other. In fact, Chomsky argued that they are essentially the same [see Chomsky (1965), pp. 24–25]. But this shift of attention did give rise to the search for universal grammar (see below).

If we consider how the linguist can arrive at the grammar of a language, then it is natural to think that he must have an evaluation procedure for choosing among a large number of possible grammars. But if we consider how the child can arrive at the grammar of a language, then it will seem to be problematic if we assume that all the child has is an evaluation procedure. How does a child acquire (learn, arrive at, construct) a grammar?

As Chomsky (1965) pointed out, there are two approaches to language acquisition: the *empiricist* approach and the *rationalist* approach (pp. 47–48). The empiricist approach assumes that there are general learning mechanisms, analogy, induction, etc.; taxonomic principles of segmentation and classification are also regarded as empiricist principles (p. 47). The rationalist approach holds that 'there are innate ideas and principles of various kinds that determine the form of the acquired knowledge in what may be a rather restricted and highly organised way' (p. 48). Chomsky argued that the empiricist approach is fundamentally wrong and he concluded that the rationalist approach must be taken in order to understand the nature of language [see Section 4.6.1 below; see also Lin (1999) for a detailed exposition].

The child is exposed to what Chomsky (1965) called 'primary linguistic data'. This data 'must include examples of linguistic performance that are taken to be well-formed sentences, and may include also examples designated as non-sentences, and

no doubt much other information of the sort that is required for language learning, whatever this may be' (p. 25). 'On the basis of such data', Chomsky pointed out, 'the child constructs a grammar - that is, a theory of the language of which the well-formed sentences of the primary linguistic data constitute a small sample. To learn a language, then, the child must have a method for devising an appropriate grammar, given primary linguistic data' (p. 25). What is the child's method for devising an appropriate grammar then?

Taking the rationalist approach, Chomsky (1965) concluded that the child must possess 'a linguistic theory that specifies the form of the grammar of a possible human language' (p. 25). In other words, the child must be equipped with 'a schematism that determines in advance the general properties of human language and the general properties of grammars that may be constructed to account for linguistic phenomena' [Chomsky (1975a), p. 12]. The 'schematism' or 'linguistic theory' was later on referred to by the term 'universal grammar' (UG) or 'the initial state of the language faculty'.

But how does a 'linguistic theory' or 'schematism' that specifies the forms of the grammar of a possible human language help the problem of selecting among grammars? Recall that a language can be characterised by many mutually inconsistent grammars. When learning a language, the child eventually arrives at just one grammar, i.e. the grammar of the language in question. UG implies that the child has 'innate predisposition ... to develop a certain kind of theory to deal with the evidence presented to him' [Chomsky (1965), p. 26]. For example, the child 'cannot help constructing a particular sort of transformational grammar to account for the data presented to him' (p. 59). What this means that is that the child will only consider certain kind of grammars — grammars which are specified by UG, but not other kinds of grammars. UG thus 'has a far-reaching effect in cutting down the number of possible alternatives' [Chomsky (1964a), pp. 241–242]. For example, 'if we have decided to construct a grammar with a phrase structure and a transformational part, a great many possible descriptions are already eliminated' (p. 242)[13] Thus, 'a part of the problem of choosing among grammars is absorbed by the theory of the form of grammar itself [i.e. UG]' (p. 243).

We have been talking about UG. We have also been speaking of the selection of a grammar by the child (i.e. the grammar of the language spoken in the child's community) among the store of potential grammars. But why should there be more than one potential grammar? Why can't it be the case that UG is so rich that the child cannot help but construct a single grammar, namely the grammar of the community, out of the data presented to him? That is, why can't UG permit only just one grammar compatible with the available data?

Chomsky (1965) regarded this as being 'logically possible' (p. 37). But he thought that 'It is difficult to imagine how in detail this logical possibility might be realised, and all concrete attempts to formulate an empirical adequate linguistic theory certainly leave ample room for mutually inconsistent grammars, all compatible with primary data of any conceivable sort' (p. 38). He then remarked:

[13] For example, finite state grammars or linear grammars will be eliminated in this case. See Chomsky (1957).

> All such theories therefore require supplementation by an evaluation measure if language acquisition is to be accounted for and selection of specific grammars is to be justified; and I shall continue to assume tentatively, as heretofore, that this is an empirical fact about the innate human *faculté de langage* and consequently about general linguistic theory as well [Chomsky (1965), p. 37].

So it is more reasonable to assume that UG cuts down the number of possible grammars, and that an evaluation procedure is still needed to select the highest-valued one among the remaining possible grammars. This explains Chomsky's remarks in several articles. For example, 'It is thus suggested that the language learner (analogously, the linguist) approaches the problem of language acquisition (grammar construction) with a schematism that determines in advance the general properties of human language and the general properties of grammars that may be constructed to account for linguistic phenomena. His task is to select the highest-valued grammar of the appropriate form compatible with available data' [Chomsky (1975a), p. 12]. In other words, the child must 'possess, first, a linguistic theory that specifies the form of the grammar of a possible human language, and second, a strategy for selecting a grammar of the appropriate form that is compatible with the primary linguistic data' [Chomsky (1965), p. 25]. In short, 'the child develops a grammar by employing the evaluation procedure to select among grammars that conform to a universal grammar' [Chomsky (1975b), p. 148].

There are thus two ways of cutting possible grammars (down to one): providing a UG and providing an evaluation procedure. As Chomsky (1965) put it, 'We can attempt to refine the evaluation measure for grammars or to tighten the formal constraints on grammars so that it becomes more difficult to find a highly valued hypothesis compatible with primary linguistic data' (p. 46). Though both ways can work, 'the latter, in general, [is] the more promising' (p. 46). 'Thus the most crucial problem for linguistic theory seems to be', suggested Chomsky, 'to abstract statements and generalisations from particular descriptively adequate grammars and, wherever possible, to attribute them to the general theory of linguistic structure, thus enriching this theory and imposing more structure on the schema for grammatical description' (p. 46). It is UG and not individual grammars that is the linguist's primary concern: 'wherever possible, general assumptions about the nature of language should be formulated from which particular features of the grammar of individual languages can be deduced' (p. 46). The 'long-term task for general linguistics' is to develop 'an account of this innate linguistic theory that provides the basis for language learning' (p. 25). It thus becomes apparent that the primary task for general linguistics was not to provide an evaluation procedure but to search for the innate universal grammar.

4.4. Universal grammar

The search for UG dominated research in generative linguistics since 1965. The years that followed saw a large number of publications by Chomsky on UG and related topics in linguistics and philosophy. In linguistics, Chomsky tried to formulate 'general assumptions about the nature of language ... from which particular

features of the grammar of individual languages can be deduced' [Chomsky (1965), p. 46; quoted above]. In philosophy, he argued that these general assumptions must be innately given, and cannot be learned on the basis of generalisation, analogy, or taxonomic principles of classification and segmentation, and hence that the empiricist approach must be wrong. These two endeavours (in linguistics and in philosophy) had to go hand-in-hand. If the empiricist approach could account for the linguistic phenomena Chomsky concentrated on, then there would be no strong reason for opting for the rationalist approach Chomsky took. On the other hand, to argue that the empiricist approach is wrong, Chomsky had to present some linguistic phenomena and to argue that only the rationalist, and not the empiricist, approach can explain them. In the past thirty years a series of abstract principles have been formulated, such as Principle of Structure-Dependency, Specified Subject Condition, A-over-A Principle, Complex-NP Condition, Tensed Subject Condition, Subjacency, Binding Principles (A, B, and C), Theta-Criterion, Case Filter, X-bar Theory, and so on (Chomsky, 1964b,c, 1973, 1976, 1977b, 1980b, 1981, 1986, 1995b; Ross, 1967; Jeckendoff, 1977; Rizzi, 1982). These principles have been attributed to UG. (The version of UG which contained such principles was called 'principles and parameters' (P&P). But hereafter I shall not distinguish between UG and P&P.) They illustrate the same point, according to Chomsky, that such knowledge cannot be learned by procedures of generalisation and analogy, etc., and that it must be innately given. Chomsky's arguments concerning these principles are of the same nature. So I shall here consider only three principles, namely Principle of Structure-Dependency, Specified Subject Condition and Binding Principle B, and the related arguments by Chomsky.

4.4.1. The Principle of Structure-Dependency

Chomsky (1975b) invited us to 'consider a familiar example, perhaps the simplest one that is not entirely trivial' (p. 30). Suppose that we want to discover the child's knowledge of forming questions in English. We first observe that the child can form such questions as those of (1), corresponding to the associated declaratives:

(1) the man is tall — is the man tall?
the book is on the table — is the book on the table?

From this observation, we might think that the child follows the following rule (this we treat as a hypothesis):

> **Hypothesis 1.** The child processes the declarative sentence from its first word (i.e. from "left to right"), continuing until he reaches the first occurrence of the word "is" (or others like it: "may," "will," etc.); he then preposes this occurrence of "is," producing the corresponding question (with some concomitant modifications of form that need not concern us) [Chomsky (1975b), p. 31].

But we will soon find that this hypothesis does not work, as we further observe that the child forms question (2) and not (3):

(2) the man who is tall is in the room — is the man who is tall in the room?
(3) the man who is tall is in the room — is the man who tall is in the room?

So we must revise our hypothesis. The correct hypothesis is of course something like this:

> *Hypothesis 2*: The child analyses the declarative sentence into abstract phrases; he then locates the first occurrence of "is" (etc.) that follows the first noun phrase; he then preposes his occurrence of "is," forming the corresponding question. (pp. 31–32).

The difference between the two hypotheses is this. Hypothesis 1 holds that the child is employing a 'structure-independent rule', while Hypothesis 2 holds that the child is employing a 'structure-dependent rule'. But 'By any reasonable standards, hypothesis 2 is far more complex and "unlikely" than hypothesis 1' (p. 32).

But the child's linguistic behaviour shows that he unerringly makes use of the structure-dependent rule postulated in hypothesis 2, rather than the simpler structure-independent rule of hypothesis 1. So the question is now why this is the case. One answer is that the child follows the structure-dependent rule because this rule facilitates communication. But Chomsky found this answer useless:

> let us try to account for it [i.e. Principle of Structure-Dependency] in terms of communication. I see no way of doing so. Surely this principle enters into the function of language; we might well study the ways in which it does. But a language could function for communication (or otherwise) just as well with structure-independent rules, so it would seem. For a mind differently constituted, structure-independent rules would be far superior, in that they require no abstract analysis of a sentence beyond words. ... To account for or somehow explain the structure of UG, or of particular grammars, on the basis of functional considerations is a pretty hopeless prospect (pp. 57–58).

The only possible answer is, argued Chomsky, that Principle of Structure-Dependency is innate:

> There seems to be no explanation in terms of "communicative efficiency" or similar considerations. It is certainly absurd to argue that children are trained to use the structure-dependent rule, in this case. In fact, the problem never arises in language learning. A person may go through a considerable part of his life without ever facing relevant evidence, but he will have no hesitation in using the structure-dependent rule, even if all of his experience is consistent with hypothesis 1. The only reasonable explanation is that UG contains the principle that all such rules must be structure-dependent. That is, the child's mind ... contains the instruction: Construct a structure-dependent rule, ignoring all structure-independent rules. The principle of structure-dependence is not learned, but forms part of the condition for language learning. (p. 32).

4.4.2. Specified subject condition and binding principle B

Chomsky (1986, p. 8) presented the following example:

(4) John ate an apple.
(5) John ate.
(6) John is too stubborn to talk to Bill.
(7) John is too stubborn to talk to.

In (4) an apple was the thing that John ate. The object of 'ate' in (5) is missing, and the sentence means that John ate something. We may generalise from sentences like (5) that:

(8) Whenever the object is missing, an arbitrary object is meant.[14]

Now, (6) means that John is so stubborn that he will not talk to Bill. Let us now look at (7), and we find that object of 'talk to' is missing. If we apply (8), we should regard (7) as meaning that John is so stubborn that he will not talk to an arbitrary person. But (7) actually means that John is so stubborn that an arbitrary person will not talk to him.[15]

Chomsky points out that children are capable of knowing the correct meaning of all the sentences in (4)–(7). He asks: are children trained to do such things? The answer is, he suggests, no. He says that it is not the case that children interpret (7) on the analogy of (5) and then hear from their parents or teachers that the derived interpretation is incorrect, and he adds that 'it is doubtful that anyone has undergone this experience and it is certain that not everyone who knows the facts has done so' [cf. Chomsky (1988), p. 21 and 24]. He further points out that the differences shown in (4)–(7) are not noted in 'even the most compendious traditional or teaching grammar' [Chomsky (1986), p. 9]. Children simply know how to derive the correct meaning of (4)–(7), and this is 'knowledge without training or relevant experience' [Chomsky (1986), p. 8].

Consider another example Chomsky (1986, p. 8) offered:

[14] This generalisation is not entirely correct. When an object of a verb is missing, it is often the case that a specific type of object, not just any arbitrary object, is meant. For example, in 'Have you eaten yet?' the missing object is 'the meal appropriate to the given time of day', and not an arbitrary object. But such qualifications are not central to the point under discussion.

[15] The above is Chomsky's argument. One may say that there is a fault in this argument. The fault is that the argument fails to distinguish between the case where the object of a verb is missing and the case where the object of a 'verb + preposition' is missing. Generalisation (8) applies in the former case but not in the latter case: it is not true that 'John talks to' means that John talks to an arbitrary person. So one may say that (7) cannot be interpreted in the same way as (5), and therefore that Chomsky's argument does not hold for this reason. But I think that Chomsky can easily side-step this attack by replacing 'talk to' in (6) and (7) with a word such as 'persuade'. In that case, the missing object will be the object of the verb (e.g. 'persuade'), and (8) will be able to apply. So I shall not pursue the just-mentioned criticism of Chomsky's argument.

(9) [The men expected to see them].
(10) I wonder who [the men expected to see them].

The problem posed by this example is that, though the bracketed phrases in (9) and (10) are the same, the pronoun *them* receives different interpretations. In (10) the pronoun refers to either the men or some other persons, but in (9) it can only refer to some other persons but not the men. Chomsky said that if the child understood language using analogy, then the child would interpret the pronouns *them* in (9) and (10) in the same way because the bracketed expressions in which they occur are exactly the same. But the child knows the difference between (9) and (10), and he knows this difference without having been explicitly taught about it. The child does not even need to have heard the specific sentences (9) and (10) before. The child's knowledge is 'knowledge without training or relevant experience' [Chomsky (1986), p. 8]. The idea of analogy is not only 'useless' but also 'wrong' in such cases.

The above two examples cannot, according to Chomsky, be explained in terms of generalisation, analogy, and the like. Rather, they can only be explained by the principle Specified Subject Condition. This principle states that a pronoun (e.g. *him*, *them*) must be free in the domain of the nearest subject. This condition gave rise to Binding Principle B, which says that a pronoun must be free in its governing category. The governing category of a pronoun is the minimal domain containing the pronoun, its governor and an accessible subject. X governs Y if and only if (i) X is a governor and (ii) X c-commands Y and Y c-commands X. The notion of c-command is defined in terms of the abstract X-bar syntax: X c-commands Y if and only if (i) X does not dominate Y and Y does not dominate X, and (ii) the first branching node dominating X also dominates Y. A node X dominates a node Y if X is higher up in the tree than Y and there is a line from X to Y only downwards. See Haegeman (1994) for these definitions. The details should not concern us here. What is important is that the Specified Subject Condition and Binding Principle B are apparently abstract: they are not known to ordinary speakers and can only be assumed to be innate (or derived from some innate principles).

We have seen some examples of UG principles. According to Chomsky (1975b), such principles cannot be learned. Rather, they are 'simply innate to the language faculty, as part of the schematism that determines admissible grammars and the way in which their rules apply, thus determining the class of languages accessible to humans by application of the language faculty' (p. 91). Such principles are 'of considerable importance', explained Chomsky, 'They restrict the class of possible rules, and also, the possible application of established rules. Therefore, they contribute to solving the fundamental problem of linguistic theory ... namely: to constrain the class of "learnable systems" so that it becomes possible to explain the rapidity, uniformity, and richness of learning within cognitive capacity' (p. 111).

4.5. From TG, through UG, to MP

In the discussion so far in this section (Section 4) we saw that it was very natural for Chomsky to propose evaluation procedures in place of discovery procedures. It

was also very natural for him to concentrate on developing UG principles instead of formulating evaluation procedures. But why didn't he start off, as it were, with developing UG? In other words, why did he instead spend so much energy in developing transformational grammar (TG), as described in Chomsky (1975a, 1957, 1964a, 1965), etc.?

This question can be answered in two ways. One way is to look at the 'short-term goal' of the writings just cited and the 'long-term goal' Chomsky had in mind. The short-term goal and the long-term goal were these:

> The purpose of LSLT, as described at various points in the manuscript "is to construct an integrated and systematic theory, which, when applied rigorously to linguistic material, gives the correct analysis for the case where intuition (or experiment, under more desirable circumstances) make a clear decision" (§93.2, p. 415). "Our problem is to carry out this development, to bring to light the formal pattern underlying the sentences of a language, and to show how these observed regularities might account for particular decisions about which sequences are grammatical and how these are to be understood" (Introduction, p. 63) ... *The primary question for linguistic theory is: how can speakers produce and understand new sentences?* (§75, p. 293) [Chomsky (1975a), p. 6; italics added].

> The fundamental long-term task is to provide "an explanation for the general process of projection by which speakers *extend their limited linguistic experience* to new and immediately acceptable forms" (§110, p. 519) [Chomsky (1975a), p. 6; italics added].

So the short-term goal was to explain the creativity of language, i.e. 'how can speakers produce and understand new sentences?'. This explains why Chomsky was interested in transformational grammar: 'The investigations of generative grammar described in [Chomsky, 1957] were motivated in part by an interest in the problem of accounting for the ability of a speaker to produce and understand an indefinite number of new sentences (or for that matter, to recognise them as properly formed, or as deviating from well-formedness in one or another aspect) ([Chomsky, 1957, p. 15]), a task that he performs regularly with great facility' [Chomsky (1961b), p. 222].[16] The transformational grammar is finite, but it can generate an infinite number of sentences. It can also account for the structure of, say English sentences. (A finite state grammar, which is finite, can also generate an infinite number of sentences, but it cannot account for the structure of a human language, see Chomsky, 1957.)

The long-term goal was, in Chomsky's later words, to solve Plato's problem concerning language. Plato's problem is 'the problem of explaining how we can know so much given that we have such limited evidence' [Chomsky (1986), p. xxv]. The study of human language presents a version of Plato's problem: How can the child acquire

[16] Notice the expression 'in part' in this quotation. This expression implies that Chomsky had other interests. My understanding is that Chomsky, when writing Chomsky (1975a) and Chomsky (1957), had also the long-term goal in mind. See below.

a grammar of remarkable complexity given so little exposure to the linguistic data? Chomsky's answer was that the child is endowed with UG, which significantly narrows down the range of possible grammars so that the child does not really need to do much choosing among them. Chomsky's long-term goal was, then, to develop UG.

Could one go directly to UG without going through TG? In other words, could one set off to discover UG principles without first considering the transformation rules of an individual language (or individual languages)? The answer is not explicit in Chomsky's writings. But it seems that the answer must be no. UG principles are highly abstract things, they are significant generalisations about the structure of language. To discover such abstract generalisations, it seems that one must start with some more concrete rules (hypotheses, generalisations) and then work upwards. This is after all standard practice in science, e.g. in physics. I think that this explains Chomsky's following remark:

> LSLT is an attempt to develop a theory of transformational generative grammar... it is argued that the competence attained by the normal speaker-hearer is represented by a transformational generative grammar, which determines the representation of each sentence on the levels of phrase structure and transformational structure ... These representations ... *provide the basis for the more general theory of language* ... The principles of this theory specify the schematism brought to bear by the child in language acquisition. They define the linguistic universals that constitute "the essence of language" ... and thus can be taken as one fundamental element in the characterisation of the innate "language faculty" [Chomsky (1975a), p. 45; italics added].

It must be noted that in LSLT the problem of language creativity and Plato's problem were treated differently. The former was explained by a transformational grammar, and the latter by UG. But in Chomsky (1965) he started to fuse these two problems, and hence the short-term and the long-term goals. He started to think that creativity in language must be explained in terms of UG. For example, he adopted the traditional view of the philosophical grammarians:

> Within traditional linguistic theory, furthermore, it was clearly understood that one of the qualities that all languages have in common is their "creative" aspect. Thus an essential property of language is that it provides the means for expressing indefinitely many thoughts and for reacting appropriately in an indefinite range of new situations ... The grammar of a particular language, then, is to be supplemented by *a universal grammar that accommodates the creative aspect of language use* and expresses the deep-seated regularities which, being universal, are omitted from the grammar itself. Therefore it is quite proper for a grammar to discuss only exceptions and irregularities in any detail. It is only when supplemented by a universal grammar that the grammar of a language provides a full account of the speaker-hearer's competence [Chomsky (1965), p. 6; italics added].

We have been talking about the relationship between TG and UG from the point of view of the short-term vs. the long-term goals. The discussion shows that the transition from TG to UG, seen from this viewpoint, is very natural. Now let us explore the other way of looking at the relationship, this time from the viewpoint of 'descriptive adequacy' vs. 'explanatory adequacy'. The meaning of 'descriptive adequacy' is this:

> A grammar can be regarded as a theory of a language; it is *descriptively adequate* to the extent that it correctly describes the intrinsic competence of the idealised native speaker. The structural descriptions assigned to sentences by the grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be immediately aware of this) in a substantial and significant class of crucial cases [Chomsky (1965), p. 24].

One can develop a transformational grammar to account for 'the linguistic intuition of the native speaker'. Different (transformational) grammars may do so with different degrees of descriptive adequacy. To make a grammar more adequate, one can make it more sophisticated, adding all sorts of specific rules to account for all sorts of data. But, as discussed in Section 4.2 above, there might be many competing (transformational) grammars with the same degree of descriptive adequacy. So one must choose among them. This is from the methodological point of view. For the conceptual point of view, one must consider the question, What is the ultimate aim of a grammar? The ultimate aim is to explain Plato's problem, i.e. how the child can know so much with so little exposure to linguistic data (see above). So one must make the grammar not only descriptively adequate, but also explanatory adequate, that is, it must explain the selection among descriptively adequate grammars and Plato's problem. Thus the term 'explanatory adequacy' means:

> To the extent that a linguistic theory succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data, we can say that it meets the condition of *explanatory adequacy*. That is, to this extent, it offers an explanation for the intuition of the native speaker on the basis of an empirical hypothesis concerning the innate predisposition of the child to develop a certain kind of theory to deal with the evidence presented to him [Chomsky (1965), p. 26].

It is easily seen that there is a conflict between descriptive adequacy and explanatory adequacy. For a grammar to be descriptively adequate, the grammar must be very sophisticated, detailed and concrete. But to make it explanatorily adequate, it must be highly abstract, it must contain significant generalisations about not only an individual language (or individual languages) but also human languages in general.

Chomsky was never really interested in the descriptive adequacy of a grammar. This is understandable. Descriptive adequacy is not an end in itself. The deeper and more intellectually challenging problem is explanatory adequacy. Thus, explanatory adequacy has always been the priority in Chomsky's thinking, witness:

> In short, the most serious problem that arises in the attempt to achieve explanatory adequacy is that of characterising the notion "generative grammar" in a sufficiently rich, detailed, and highly structured way. A theory of grammar may be descriptively adequate and yet leave unexpressed major features that are defining properties of natural language and that distinguish natural languages from arbitrary symbolic systems. It is for just this reason that the attempt to achieve explanatory adequacy — the attempt to discover linguistic universals — is so crucial at every stage of understanding of linguistic structure, despite the fact that even descriptive adequacy on a broad scale may be an unrealised goal [Chomsky (1965), p. 36].

Nonetheless, the conflict between descriptive adequacy and explanatory adequacy was a problem in Chomsky's theory. So how did he deal with it? Chomsky's ingenious solution was to let the more concrete grammars for individual languages be derivable from the more abstract UG principles. So, achieving descriptive adequacy and explanatory adequacy were no long seen as separate enterprises: one could still see a descriptively adequate grammar as consisting of various language-specific rules, but now such rules could be derived from UG principles. The tension between descriptive adequacy and explanatory adequacy was thus greatly reduced. Chomsky summarises this well, when he states:

> Concern for descriptive and explanatory adequacy is as old as the study of language. As soon as the two traditional goals were reformulated with modern generative grammar, a serious tension arose between them: the search for descriptive adequacy seems to lead to ever greater complexity of rule systems, varying among grammatical constructions and across languages, while the search for explanatory adequacy leads to the conclusion that language structure is largely invariant. It is this tension that has driven the research inquiry of generative grammar for some 40 years. The P&P framework suggests a way to resolve the tension, thus offering some conception of the form that a genuine theory might take [Chomsky (1998), p. 3].

Could one go directly to an explanatorily adequate grammar without ever considering a descriptively adequate grammar (or grammars)? The answer seem to be no. To say that a grammar is explanatorily adequate is to say that the grammars of individual languages are derivable from it. If one had no idea what the latter could be, then how could one talk about the former at all? So again, it seems very natural for Chomsky to explore TG before moving on to UG.

It is worth stressing the following difference between TG and UG. In TG, 'Each language is a rich and intricate system of rules that are, typically, construction-particular and language-particular' [Chomsky (1995b), p. 170]. For example, 'the rules forming verb phrases or passive clauses in English' are 'specific to *these* constructions in *this* language' (p. 170). But in UG, 'UG provides a fixed system of principles and a finite array of values for these parameters. The language-particular rules reduce to choice of values for these parameters' (p. 170). So, 'The

notion of grammatical construction is eliminated, and with it, construction-particular rules' (p. 170). For example, 'Constructions such as verb phrase, relative clause, and passive remain only as taxonomic artefacts, collections of phenomena explained through the interaction of the principles of UG, with the values of parameters fixed' (p. 170).

The rules in TG are derivable from UG principles, so they do not have independent existence. The same question can now be asked of UG principles, such as the ones mentioned in Section 4.2 above: Do these principles have an independent existence or can they be in turn derived from some more general principles? This question led to the development of the minimalist program (MP).

To understand MP, it is useful to consider another question, namely, What is language used for? A reasonable answer is this: 'The language is embedded in performance systems that enable its expressions to be used for articulating, interpreting, referring, inquiring, reflecting, and other actions' [Chomsky (1995b), p. 168]. So, 'We can think of [a linguistic expression] as a complex of instructions for these performance systems, providing information relevant to their functions' (p. 168). This suggests that there are (at least) two linguistic levels, namely, the level where linguistic expressions interact with the articulatory-perceptual systems and the level where they interact with the conceptual-intentional systems. The former level has generally been taken to be phonetic form (PF), the latter logical form (LF) (see p. 168).

It is also useful to consider still another question: What does a language consist of? It is reasonable to think that a language consists of a set of words and a process for building sentences out of words. In Chomsky's (1995b) terms, it is a standard assumption that 'a language consists of two components: a lexicon and a computational system' (p. 168). The lexicon 'specifies the items that enter into the computational system, with their idiosyncratic properties', while the computational system 'uses these elements to generate derivations and [linguistic expressions]' (pp. 168–169).

The first of these two questions looks at language from the outside, and the second looks at language from within. The considerations just mentioned suggest that PF, LF, the lexicon, and the computational system are the minimal requirement that a theory of grammar (UG) must satisfy. In other words, UG *must* have these levels and components. The existence of PF, LF, the lexicon and the computational system is thus justified. But what about the levels and principles postulated in P&P, such as S-structure, D-structure, X-bar theory, Case Filter, Subjacency, Binding, Government, and the like? 'Any [such] additional structure or assumptions require empirical justification', stressed Chomsky (1995b, p. 187). It turned out that such additional structure and assumptions can be derived from certain more general principles, such as Procrastinate, Greed, Last Resort, and so on [see Chomsky (1995b)].

I shall not go into the details of MP. The transitions from TG to UG and from UG to MP are clear enough. The motivation behind all this was to characterise, as much and as revealingly as possible, the innate language faculty, in order to explain the creativity and child acquisition of language.

4.6. Methodology

In this section (Section 4) we have been discussing why a change of aspect occurred in Chomsky's linguistic thinking and how it led to the development of Chomsky's linguistic theory. In this subsection, let us have a bird's-eye view of Chomsky's methodology. Needless to say, Chomsky's methodology was closely related to his linguistic thinking, or was even an integral part of it. Broadly speaking, the methodology consisted of two related parts: rationalism and scientificism. Both were designed to refute the taxonomic approach. We shall look at them in turn.

4.6.1. Empiricism vs. rationalism

As mentioned in Section 4.3, there are two approaches to the analysis of linguistic knowledge, or of knowledge in general, namely the empiricist approach and the rationalist approach [cf. Chomsky (1965), pp. 47–48]. In the case of linguistic (grammatical) knowledge, the rationalist approach is like this. A scientist S 'would try to discover the property of the mind P that enables a child endowed with P to acquire the grammar of the language' [Chomsky (1975b), p. 147]. 'Having gotten this far, S would now proceed to investigate the property P', continued Chomsky, 'he might proceed in the manner of R [rationalism], formulating a general schematism (call it "universal grammar") and an evaluation procedure which, he would postulate, jointly constitute P, or an essential element in it. Endowed with these systems in its initial state, the child develops a grammar by employing the evaluation procedure to select among grammars that conform to a universal grammar. If the latter is sufficiently restrictive, a small amount of evidence might lead quickly to the selection of a grammar that is very rich and complex and that goes well beyond this evidence' (pp. 147–148).

The rationalist approach R assumes that the child is endowed with UG, a set of innate grammatical principles, which enables the child to select the grammar of the language quickly among all the possible grammars (together with an evaluation procedure). It assumes that UG is 'psychological', 'real', 'genetically determined' [Chomsky (1975a), pp. 35–39]. These were there at the very beginning of Chomsky's linguistic thinking. For example, Chomsky remarked:

> In LSLT the "psychological analogue" to the methodological problem of constructing linguistic theory was not discussed, but it lay in the immediate background of my own thinking. To raise this issue seemed to me, at the time, too audacious [Chomsky (1975a), p. 35].

> In LSLT, the "realist" position is taken for granted [Chomsky (1975a), p. 35].

> These matters are not discussed in LSLT, but the issues lie in the immediate background of this work and have been the subject of considerable discussion and controversy since [Chomsky (1975a), p. 13].

By contrast, the empiricist approach E does not assume that there are innate *grammatical* principles. Instead, it assumes that 'The mind contains a system of

properties that provide an initial analysis of the data of sense. Systems of knowledge and belief are developed by procedures of generalisation, analogy, induction, association, and habit formation that are developed in the several varieties of empiricist psychology and philosophy' [Chomsky (1975a), p. 13]. It is these *general cognitive procedures* which are regarded as being innate in E.

But the notions of generalisation and analogy, etc. are too vague. This Chomsky sharply pointed out a long time ago: 'although there is frequent reference in the literature of linguistics, psychology, and philosophy of language to inductive procedures, methods of abstraction, analogy and analogical synthesis, generalisation, and the like, the fundamental inadequacy of these suggestions is obscured only by their unclarity' [Chomsky (1975), p. 31]. Later on he repeated the same point: 'There is no general notion of ''analogy'' that applies to these and other cases. Rather, the term is being used, in an extremely misleading way, to refer to properties of particular subsystems of our knowledge, entirely different properties in different cases' [Chomsky (1988), pp. 26–27].

What about taxonomic procedures of classification, segmentation, etc.? 'It is, I believe, appropriate', remarked Chomsky (1975a), 'to regard the procedural approaches of structural linguistics as an unusually refined, detailed, and sophisticated development of a theory of this general [empiricist] character' (p. 13). Some pages later Chomsky repeated: 'As already noted, the procedural approaches developed and applied to varied language material are among the most sophisticated and interesting efforts undertaken within a significant (i.e. nonvacuous) empiricist framework' (p. 36). Similarly, in Chomsky (1975b) he maintained that 'it would be reasonable to regard these as an instance of E, perhaps the most complex version that has yet been developed' (p. 138).

But as Chomsky (1975a) pointed out, 'There was some unclarity, in this work, as to exactly what was at stake. Procedures are sometimes presented simply as a device for giving a more organised and less redundant characterisation of the linguist's corpus of data. Thus one set of procedures, and the results they provide, are in principle no better or worse than others, apart from their utility for particular purposes. No empirical claim is associated with a particular system of procedures, under this interpretation' (p. 36).

But can't one think of the taxonomic procedures as being innately given? Chomsky's (1975a) answer is affirmative: 'However, a careful review of the work in question seems to me to show that a different and more interesting conception was implicit, despite disclaimers: namely, that the procedures determined the true structure of language, a system with ''psychological reality''' (p. 36). He suggested that 'we [can] interpret the methods of structural linguistics based on segmentation and classification as a ''learning theory'' (contrary to the intentions of those who developed these methods, so far as I know)' [Chomsky (1975b), p. 138]. Taxonomic procedures under this interpretation then are linguistic universals, which are innate, genetically determined: 'Thus, if one takes a realist interpretation of the work of post-Bloomfieldian theorists, they are proposing certain quite deep linguistic universals: the principles implied by their procedural methods. Under this interpretation, it is postulated that human languages must have the properties determined by application of these procedures to a corpus of data' (p. 37). If the taxonomic approach is successful,

then it will show how the child 'acquires language by applying procedures of the sort to the data available to him' (p. 11). And this will be a great discovery: 'If a general mechanical procedure for constructing grammars of the appropriate type from the raw data can be developed (and we are, I believe, very far from this), this will be a major achievement which will, for example, shed great light on the human ability to learn language' [Chomsky (1964a), p. 245].

But alas, the taxonomic approach would lead to a dead end, concluded Chomsky (see also Section 3 above):

> It seems to have been demonstrated beyond any reasonable doubt that, quite apart form any question of feasibility, methods of the sort that have been studied in taxonomic linguistics are intrinsically incapable of yielding the systems of grammatical knowledge that must be attributed to the speaker of a language (cf. Chomsky 1956, 1957, 1965; Postal [1962, 1964a, 1964b]; Katz and Postal 1964, §5.5, and many other publications for discussion of these questions that seems unanswerable and is, for the moment, not challenged) [Chomsky (1965), p. 54].

So, according to Chomsky (1975b), 'approaches of the character of E, which assigns a fundamental role to segmentation, classification, analogy, and generalisation, are mistaken in principle' (p. 152), and 'no such approach is feasible' (p. 153). The examples such as those shown in Section 4.4 above further demonstrate, according to Chomsky, that the empiricist approaches are hopeless in accounting for language. He concluded that 'a theory of the character of R is much more likely to be successful' (p. 153).

4.6.2. Scientific

In Section 3.3 above we saw that Chomsky found the taxonomic approach at odds with general scientific practice. In no scientific inquiry apart from linguistics would a scientist seriously consider the possibility of discovering a theory, or choosing among competing theories, by a general, mechanical procedure. So, in Chomsky's view, the taxonomic approach was not really scientific, contrary to what the taxonomic linguists thought. Chomsky wanted a genuine scientific theory. He stated that 'we must apparently do what any scientist does when faced with the task of constructing a theory to account for particular subject matter — namely, try various ways and choose the simplest that can be found' [Chomsky (1964a), p. 223]. A theory of grammar is 'analogous to a particular scientific theory, dealing with its subject matter (the set of sentences of this language) much as embryology or physics deals with its subject matter' (p. 233). This is because 'A grammar of L seeks to formulate laws (grammatical rules) in terms of theoretical constructs (particular transformations, phonemes, and so on) which govern the construction of sentences, i.e., which, in particular, correctly predict which physical events are and are not sentences acceptable to the native speaker, whether they have been observed or not' (p. 223).

In many of his writings Chomsky (1964b, 1966, 1972, 1975b, 1980b, 1986, 1995a) compared theory of grammar with physics, chemistry, biology, psychology, the brain sciences, or natural sciences in general. He regarded language as a natural

object and proposed to study it in the same way as natural sciences study the stars, the heart, the free-falling object, and so on.

5. An examination of Chomsky's argument

5.1. A summary of Chomsky's argument

I have given a detailed account of the development of Chomsky's linguistic thinking. The transition from taxonomic grammar and discovery procedures, through transformational grammar and evaluation procedures, through Principles and Parameters, to the minimalist program, seems to be rather natural and well motivated. Chomsky's thinking can be summarised as consisting of the following three parts.

1. Grammar is unknown to, and hidden from, the speaker.

'A human language is a system of remarkable complexity', pointed out Chomsky. 'To come to know a human language', he stated, 'would be an extraordinary intellectual achievement for a creature not specifically designed to accomplish this task'. He emphasised that 'A normal child acquires this knowledge on relatively slight exposure and without specific training'. He maintained that this is to be explained by assuming that the child 'quite effortlessly make[s] use of an intricate structure of specific rules and guiding principles'. And he claimed that such rules and principles 'lie far beyond the reach of will or consciousness' [all quotations in this paragraph are from Chomsky (1975b), p. 4].

Chomsky maintained that the linguist's task is to discover such rules and principles. He suggested that 'It is useful, at the outset, to make a distinction between *data* and *facts*': 'The linguist's data consists of certain observations about the form and use of utterances', and 'The facts of linguistic structure that he hopes to discover go well beyond these observations'. The difference between 'data' and 'facts' is this: 'An enormous amount of data is available to any native speaker; the deeper facts of linguistic structure, however, remain hidden to him' [all quotations in this paragraph are from Chomsky (1961b), p. 219].

Chomsky (1965, p. 18) reckoned that the 'facts' constitute the speaker's knowledge of the grammar. And he held that the grammar of a language is unknown to the speaker: 'Like most facts of interest and importance, this is neither presented for direct observation nor extractable from data by inductive procedures of any known sort'.

2. Knowledge of grammar had better be approached indirectly rather than directly.

Since we do not know what knowledge of grammar is, how can we discover it (i.e. knowledge of grammar)? One approach is the direct, empiricist approach, which assumes that there are general procedures of generalisation and analogy, or

mechanical procedures of segmentation and classification. But this is simply not feasible: 'It would be strange indeed if it were possible to develop methods for discovering something whose nature and structure is so complex and still so little understood as a grammar of a language' [Chomsky (1959), p. 218]. Besides, this idea simply runs against common scientific practice: 'There are few areas of science in which one would seriously consider the possibility of developing a general, practical, mechanical method for choosing among several theories, each compatible with available data' [Chomsky (1957), p. 53]; 'Such a justification of systems as abstract and theoretical as grammars has never been thought possible in any other field' [Chomsky (1959), p. 215]. But feasibility apart, that the mechanical, taxonomic approach will not work 'seems to have been demonstrated beyond any reasonable doubt', according to Chomsky (1965, p. 54).

3a. So we must consider the problem of justifying grammars.

A language can be characterised by many grammars. A child is presented with primary linguistic data. His task is to arrive at the grammar of his language, that is, to select that grammar from among all grammars compatible with the data. It might be thought that the child only needs an evaluation procedure (for selecting grammars). But it is more plausible to think that the child is endowed with UG, which significantly narrows down the number of potential grammars.

3b. We must also consider the problem of justifying UG principles.

In the process of linguistic research we make many linguistic generalisations (e.g. Principle of Structure-Dependency, Subjacency, Binding Principles), and we attribute them to UG. But do these principles have an independent existence? Are they really 'real'? Or are they derivable from some more fundamental principles? To achieve greater explanatory adequacy, we must, as happened in physics, search for the more basic principles. Hence the minimalist program.

These three parts can be put more succinctly as follows. Grammars are not known and are to be discovered. Knowledge of grammar is not learned and must be innate. Generative linguistics is a branch of natural science and principles postulated must be justified.

One must admit that Chomsky's linguistic thinking is very logical. If grammar is not known, then indeed it will be odd to hope for some mechanical procedures for discovering it. If knowledge of grammar cannot be explained by generalisation and analogy etc., then it will indeed seem to be right to assume that knowledge of grammar consists of innate linguistic principles. If (generative) linguistics is a branch of science, then it will indeed seem to be natural to search for more and more general and abstract linguistic principles. But there are three 'ifs' here. Is it really the case that the grammar of a language is not known? Is it really the case that generalisation and analogy etc. cannot explain knowledge of language? Is it really the case that generative linguistics is like a science? If these are not the case, then Chomsky's linguistic thinking will be, though logical, founded on false premises. In the rest of this section, I shall examine these very important questions, and argue that Chomsky has been mistaken on all of them.

But before we move on, I must clarify my use of the term 'know'. I am using the term in its ordinary sense. When I say that I know a grammar or a rule of grammar, I am using the word 'know' in exactly the same way as in 'I know that this is my right hand' and 'I know that London is the capital of Britain'. The word 'know' in this paper thus means 'explicitly know' or 'consciously know'. But Chomsky does not think that grammar rules are known to the speaker in this sense. (He thinks grammar rules to be government-and-binding rules or minimalist principles, which are so abstract and cannot be consciously known to an ordinary speaker, see Chomsky, 1986, 1995b, 1998). But Chomsky also sometimes speaks of grammar as being 'known' to the speaker. Here he is using the word 'know' in a very different sense; and it means 'tacitly know' or 'unconsciously know'. Of course, this usage can be confusing, so Chomsky prefers talking about the speaker's 'cognizing' the grammar to speaking of his (the speaker's) 'knowing' the grammar [see Chomsky (1980b), p. 70; Chomsky (1986), pp. 268–269]. But, in the ordinary sense of 'know', it is correct to say that Chomsky holds that grammar is unknown to the speaker.[17]

5.2. Is grammar of language unknown?

Chomsky argued that the grammar of a language is unknown, in two ways. One is that language is complex, and that only humans and not animals can learn a language with ease. But there are many other types of knowledge, e.g. knowledge of tennis games and knowledge of arithmetic, which are not simple, and which are easy to learn for humans and not for animals. And yet rules in such cases are really known. Just citing the complexity of language and the difference in learning a language between humans and animals is not sufficient to conclude that grammar is unknown.

Chomsky's second way is more forceful. He presented a series of linguistic phenomena and argued that they can only be explained by certain principles, such as Principle of Structure-Dependency, Subjacency, and Binding Principles. These principles, argued Chomsky (1986), belong to UG, and 'lie far beyond the reach of will or consciousness'. They constitute knowledge 'without training or relevant experience' (p. 8) . They must be innately given.[18]

There is no better way to argue against Chomsky's conclusions than to show, case by case, that the grammar rules involved are in fact known. Consider Principle of Structure-Dependency first. This principle states that all grammar rules, such as question-forming in English, is structure-dependent. The English question-forming rule is: to form a question one must prepose the word '*is*' in a declarative sentence of the type '*Something is* ___ ', where '*something*' can be expressed in a single word (e.g.

[17] I thank Frederick Newmeyer for prompting this paragraph.

[18] It is to be noted that 'innate' and 'conscious' are not exclusive. There are things which seem to be both innate and conscious. For example, the concept of a straight line. This concept is arguably innate, but we also consciously know what a straigtline is. One might also argue that the categories of noun and verb are also innate and conscious. But in this paper we are concerned with grammar rules and principles. It is difficult to see how a grammar rule can be both innate and conscious. So, as far as grammar rules are concerned, it seems to be true that if a rule is innate then it is not conscious, and vice versa. This note was prompted by comments from Frederick Newmeyer.

'*water*', '*he*'), or in a string of words which has its own structure (e.g. '*that stuff over there*', '*the man you saw in the park*'), but one must not, in most cases, just prepose the first '*is*' one finds in the declarative sentence. This rule is known to any competent speaker of English. All rules of grammar, such as (11–14) and the transformations presented in Section 2.3, are structure-dependent. But the rules are all known to the English speaker.

The Principle of Structure-Dependency points out a common feature of grammar rules, which is that they are all structure-dependent. But the speaker does not need to know this principle *per se* in order to judge the grammaticality of sentences, because he only needs to apply the grammar rules, which he knows. But it is nevertheless an interesting question why grammar rules are structure-dependent rather than structure-independent. I shall discuss this question in Section 5.3 below. But this question should not distract us from noticing the fact that grammar rules, such as the rule for question-forming, are known to the speaker already.

Next, consider the interpretation of (6) and (7), mentioned in Section 4.4.2 above. We know that there are the following two sentence types:

(11) Somebody is too ADJ to do something.
(12) Somebody is too ADJ to do something to.

We know that (6) is an instance of (11), together with, say, *John is too tired to walk home*. We also know that (7) is an instance of (12), together with, say, *The stick is too hard to break*.

We further know that (11) and (12) have different meanings. (11) means something like *he is so ADJ that he will not do it*; while (12) means something like *he is so ADJ that people will not do it to him*. Because (6) is an instance of (11), so it roughly means that John is so stubborn that he will not talk to Bill; because (7) is an instance of (12), so it roughly means that John is so stubborn that people will not talk to him.

So we know that (6) and (7) are interpreted differently. We know this, because we have the knowledge described in the preceding three paragraphs, and that knowledge is explicit, conscious.

To explain the problem posed by (9) and (10), we simply need to make clear how these two sentences are constructed on the basis of which sentence types. We first need to notice that the word 'expect' can appear in (at least) the following two sentence frames:

(13) *X* ___ *(wants, expects, etc.) to V him.*
(14) *X* ___ *(wants, expects, etc.) Y to V him.*

We further need to notice that *him* in these two frames has different interpretations, that is:

In (13) *him* and *X* do not co-refer, while in (14) they can.

Next, we need to note that sentence (9) is formed on the basis of (13), and that sentence (10) is constructed out of (15):

(15) I wonder [the men expected who to see them],

which utilises frame (14). Since (9) and (10) involve (13) and (14) respectively, so the word *them* is interpreted differently in (9) and (10). In this analysis, we need to know

and understand (i) the basic sentence frames, and (ii) how a sentence is constructed from them. But we do know these, so we do interpret (9) and (10) differently and correctly. No innate grammatical rules or principles are needed to explain this.[19]

5.3. *Is grammar not learnable?*

In Section 4.4 above, we saw that by presenting a number of linguistic phenomena Chomsky argued for innate linguistic principles and at the same time he argued against the idea of generalisation, analogy, etc. Let us have a close look at his argument.

Consider the example of question-forming, discussed in Section 4.4.1 earlier. There are two hypotheses: hypothesis 1, which is structure-independent, and hypothesis 2, which is structure-dependent. In some sense, hypothesis 1 is simpler than hypothesis 2. But the child 'unerringly' makes use of the structure-dependent rule in hypothesis 2. Now, how can this be explained? Chomsky considered an explanation in terms of 'communicative efficiency'. According to some empiricist philosophers such as Searle (1972), the essential function of language is communication. But Chomsky argued that Principle of Structure-Dependency simply cannot be explained in terms of communication, stating:

> let us try to account for it [i.e. Principle of Structure-Dependency] in terms of communication. I see no way of doing so. Surely this principle enters into the function of language; we might well study the ways in which in does. But a language could function for communication (or otherwise) just as well with structure-independent rules, so it would seem. For a mind differently constituted, structure-independent rules would be far superior, in that they require no abstract analysis of a sentence beyond words. ... To account for or somehow explain the structure of UG, or of particular grammars, on the basis of functional considerations is a pretty hopeless prospect [Chomsky (1975b), pp. 57–58; quoted earlier].

So, Chomsky concluded that 'The only explanation that has been proposed, to my knowledge, is ...: the principles belonging to universal grammar, which is an element of the "initial state"' [Chomsky (1975b), p. 174].

This argument by Chomsky creates deep unease. Much of what Chomsky said seems to be true. For example, that the child makes use of the structure-dependent rule rather than the structure-independent rule; and that for an alien's mind structure-independent rules might be more efficient than structure-dependent rules. Yet Chomsky's conclusion that Principle of Structure-Dependency is an innate linguistic

[19] Matthews (1998) also dissolves some of Chomsky's puzzling examples by means of certain constructions. He too argues that 'speakers learn individual constructions' and that 'to know a language is not to know an abstract set of principles that together constrain movement, binding, and so on' [Matthews (1998), p. 109]. And he also concludes that there is nothing in Chomsky's examples that 'cannot, in a commen sense view, be explained without ['Principles and Parameters Theory']' [Matthews (1998), pp. 112–113].

principle does not seem to be so convincing. So there seem to be some problems with Chomsky's argument. Putnam even said bluntly that Chomsky's argument is 'not good' [Putnam (1980), p. 287]. The reason Putnam gave was that the structure-dependent rule can be, he suggested, learned by some general learning mechanisms and need not be innately given.

What is the difference between Chomsky's view and Putnam's on the fact that children invariably use the structure-dependent rule rather than the structure-independent rule? The difference lies in the different senses of the word 'innate'. This word has two senses, one *strong* and the other *weak*. In its strong sense, to say that an ability or a piece of behaviour is innate is to say that it is governed, determined, or caused by a corresponding principle, which is genetically encoded. In its weak sense, it is to say that it is determined, or caused by some innate mechanisms. To illustrate this difference, take the example of walking upright. The ability of walking upright is arguably innate (the word 'innate' here is ambiguous). One may explain this by hypothesising an innate principle of walking upright (the strong sense). One may also explain this in terms of the physical structure of the creature (the weak sense).

Now, Chomsky's view is a strong one. According to him, the child's structure-dependent-rule-using behaviour is governed by an innate linguistic principle, called Principle of Structure Dependency. In Chomsky's words, 'the child's mind ... contains the instruction: Construct a structure-dependent rule, ignoring all structure-independent rules' [Chomsky (1975b), p. 32, quoted earlier]. But Putnam's view on this is a weak one. According to Putnam, this behaviour of the child is to be explained in terms of certain general learning mechanisms, and there is no need to postulate an innate linguistic principle.

The reason why Chomsky's conclusion that there must be an innate linguistic principle, namely, Principle of Structure-dependency, does not feel so convincing is now clear. This is because that there are two ways of explaining the fact that children invariably use structure-dependent rules rather than structure-independent rules. Chomsky's explanation is one, Putnam's is the other. If Putnam's explanation has no plausibility at all, if one is simply not inclined to give that explanation a single thought, then one will not find Chomsky's explanation at all perplexing, and one will simply accept Chomsky's explanation without having any unease.

Yet, for Chomsky, the explanation in terms of general learning mechanisms has no plausibility, and is hopeless. So, for Chomsky the existence of innate linguistic principles is the only plausible explanation. Recall that in Chomsky's eye the idea of general learning mechanisms suffers a two-folded problem. On the one hand, the notion is too vague and has no substance. On the other hand, the idea simply cannot explain many linguistic phenomena.

Consider the example of question-forming in English. Why doesn't the child form the question shown in (3) on the analogy of (1)? How does the child know that he should not use such an analogy?

Or consider the interpretation of (6) and (7). Why doesn't the child interpret (7) on the analogy of (5)? Why doesn't he make the generalisation (8)? How does the child know that he should not use this analogy or make this generalisation? How does the

child know that he should instead make generalisations (11) and (12), and establish an analogy between (6) and (11), and between (7) and (12)?

Or consider the interpretation of (9) and (10). Why doesn't the child interpret (9) on the analogy of (10)? Why should he instead make generalisations (13) and (14)? How does the child know all this?

These questions are, undoubtedly, very sharp questions. Now, how can the proponents of general learning mechanisms (e.g. Putnam) answer these questions? They simply cannot, stated Chomsky:

> All that Putnam has so far assumed is that S_0^L [the genetically determined initial state for language learning], whatever it may be, contains only the general mechanisms for learning. Recall that he gives no hint as to what these are. To invoke an unspecified "general intelligence" or unspecified "multipurpose learning strategies" is no more illuminating than his reference, at one point, to divine intervention. We have no way of knowing what, if anything, Putnam has assumed. The point is worth stressing, since it illustrates a common fallacy in discussions of this sort. The use of words such as "general intelligence" does not constitute an empirical assumption unless these notions are somehow clarified [Chomsky (1980a), p. 320].

Chomsky further commented:

> If indeed Putnam could characterise "general intelligence" or "multipurpose learning strategies" in some manner, and indicate, however vaguely, how the structure-dependent property of syntactic rules follows from the assumption that innate structure is as characterised, I would be happy to consider the hypothesis that this property should be attributed to "general intelligence" rather than to S_0^L ... Furthermore, if it can be shown that all properties of S_0^L can be attributed to "general intelligence," once this mysterious notion is somehow clarified, I will cheerfully agree that there are no special properties of the language faculty. But Putnam offers not even the vaguest and most imprecise hints as to the nature of the "general intelligence" or "multi-purpose learning strategies" that he believes to exist. Therefore, his claim that some particular properties of S_0^L can be explained in terms of these notions cannot be assessed. It has the form of an empirical hypothesis, but not the content of one [Chomsky (1980a), p. 311].

One must admit that Chomsky's argument has tremendous force. Yes, the notion of 'general intelligence' or 'multi-learning mechanisms' is extremely vague. Yes, we know so little about the nature of general intelligence or multi-learning mechanisms. Yes, no one can (yet) give a substantial answer to the question why the child uses some analogies or make some generalisations but not others. So it seems that Chomsky is justified in concluding that the notions of 'general learning mechanism', 'generalisation', 'analogy', etc. cannot explain language acquisition, and that the linguistic knowledge which is exhibited in the child language ability is not learned at

all but is innately given. It also seems that the explanations I offered in Section 5.2 above are useless.

Alas, this argument, compelling as it is, has a deep error in it, which is by no means easy to detect. But once this error is pointed out, one cannot fail to see it. To explain what this error is, let me make a distinction between two kinds of knowledge: knowledge that such-and-such is the case, and knowledge about how one comes to have that knowledge. I shall refer to these types of knowledge as 'knowledge-that' and 'knowledge-how' respectively. To illustrate this distinction, consider the following example. Someone points to a chair and utters to me *That's a chair!*. I know that by *chair* he means the object he is pointing to. But how do I know this, after all when uttering that sentence, the person could be referring to the object, to the shape, to the material, to the colour, to the size, to the weight, or to the location, etc.? The fact is that I do not know how I come to know that *chair* means the object (and not the colour, the shape, etc.), but I do know that *chair* has that meaning. The latter is knowledge-that, and the former knowledge-how.

The point of the distinction between knowledge-that and knowledge-how is this. Not having knowledge-how does not mean that one does not have knowledge-that. For example, I do not have the knowledge how I know that *chair* means chair, but this does not erase my knowledge that *chair* means chair.

Since the distinction between knowledge-that and knowledge-how is important for this paper, I shall here try to clear a possible misunderstanding. It was after all Ryle (1946) who first distinguished knowledge-that from knowledge-how. Though using the same terms, Ryle's distinction is different from mine. Ryle considered certain 'concepts of intelligence', such as *clever, wise, prudent, skilful*, etc. He argued that knowing how to do something intelligently does not reduce to knowing a certain set of rules or maxims. For example, a stupid chess player, no matter how many rules or tactical maxims are imparted to him, still cannot play chess cleverly. On the other hand, a sharp student may argue well who has never heard of formal logic. In Ryle's doctrine, knowledge-that means knowledge of certain rules or maxims (such knowledge can be explicit or tacit), and knowledge-how is knowledge of how to do something intelligently. My 'knowledge-that' is also knowledge of certain rules or maxims, specifically knowledge of grammar rules, the only difference is that I regard this knowledge as being explicit. But my 'knowledge-how' is very different from Ryle's. In my case, it means knowledge of how certain pieces of knowledge are acquired. To iterate, in my case, the distinction between knowledge-that and knowledge-how is between (a) that knowing something is case and (b) that knowing how that knowledge is acquired. The two types of knowledge are independent of each other. We may have a certain piece of knowledge without knowing how (exactly) we came to have it, i.e. without knowing the detailed acquisition processes, algorithms, etc.

Now, the error in Chomsky's argument is this. He failed to distinguish between the knowledge-how and the knowledge-that involved in his examples. Surely, we do not yet know how the child comes to know the structure-dependent rule for forming yes-no questions in English; we do not yet know how the child comes to know not to interpret (9) on the analogy of (10), and not to make generalisation (8), etc. But this

does not mean that we do not know, or that we do not know that the child eventually knows, that the rule of forming yes-no questions in English is something like that expressed in Hypothesis 2, that sentence (6) and (7) are of different types, that (9) and (10) are constructed differently and hence the word *them* in the two sentences are also interpreted differently (see Section 5.2 above). In other words, although we do not have the knowledge-how involved in these examples, we do have the knowledge-that involved. This knowledge-that, is explicit and conscious. This knowledge *is* knowledge of grammar, and it is not 'lying well beyond consciousness' or 'innately given'.

It might be argued that Chomsky himself does draw a distinction between knowledge-that and knowledge-how. The reason is this. According to Chomsky, the speaker of a language knows whether an arbitrary sentence is grammatical or not (this ability is called *grammatical competence*); and this knowledge is knowledge-that. But the speaker usually cannot explain how he (the speaker) knows that, and this is knowledge-how. So it might be argued that it is precisely this knowledge-how that Chomsky tries to capture and use to explain grammatical competence. But in this distinction 'knowledge-that' is only knowledge of the grammaticality of *individual* sentences. I agree with Chomsky that the speaker has this knowledge-that. But I argue that the speaker also has knowledge of *rules* of grammar, such as (11)–(14), and that this knowledge is also knowledge-that. More importantly, I argue that such rules explain why certain sentences are grammatical or ungrammatical (see Section 5.2 above). It is this layer of knowledge-that, i.e. the knowledge of such rules as (11–14), that Chomsky has neglected. The reason why Chomsky neglected such rules is that he did not think that these rules had been, or could be, explained in terms analogy, generalisations, and so on (see Sections 3 and 4.6 above). Though the inadequacies Chomsky said of the notions of analogy and generalisations, etc. were quite true, he was not justified in dismissing the approach of explaining grammatical competence in terms of rules such as (11–14). The acquisition details of such rules are not yet known, but the rules themselves are known to every normal speaker.

As said earlier, Putnam regarded Chomsky's argument as being 'not good', and his reason was that examples such as the example of question-forming in English can be, he suggested, explained in terms of the notion of 'general intelligence' or 'multi-purpose learning mechanisms'. He (Putnam, 1980) went on to sketch how this can be done. But Putnam's counter-argument depended entirely on the credibility of that notion. As Chomsky pointed out, that notion is too vague and has little substance (so far), and hence has no credibility (at this moment in time). No wonder Chomsky remarked that 'I see no arguments here that bear even remotely on any position that I was maintaining or even offering as a speculation' [Chomsky (1980a), pp. 323–324]. But both Putnam and Chomsky failed to see that the argument does not actually depend on whether we can clarify the notion of 'general intelligence' or 'multi-purpose learning strategies', that is, whether we can provide any details of such things. Knowledge of grammar, as shown in Section 5.2 above, is explicit and conscious. This is true, even if we know nothing about how that knowledge comes about, whether by 'general intelligence', 'multi-purpose learning mechanisms', 'generalisation', 'analogy', or some other

things. What knowledge of grammar is and how we come to have it are two different questions. The confusion of the two questions was the fundamental error in Chomsky's linguistic thinking. And it is also the reason why the debate concerning this topic in linguistics and philosophy has not made much progress.

To sum up this subsection, Chomsky's conclusion that knowledge of grammar is not learned but innately given and unconscious was based on a fundamental error, which is that he failed to distinguish between knowledge of grammar and knowledge of how that knowledge is acquired. In other words, the error was the confusion of knowledge-that and knowledge-how. Grammar rules are known to us all, but we do not know (exactly) how we come to have that knowledge. But what knowledge of grammar is and how it is acquired are not the same question. We know what knowledge of grammar is [see Section 5.2 above; see Lin (1999) for more details], even if we know nothing about how that knowledge is acquired.

Before ending this subsection, let us reconsider the issue of accounting for Principle of Structure-Dependency, i.e. why people use structure-dependent rather than structure-independent rules, in terms of communication. Chomsky claimed that 'I see no way of doing so', because 'For a mind differently constituted, structure-dependent rules would be far superior'. But once we see that the structure-dependent rules are known (see Section 5.2 above), there is no difficulty in accounting for the principle in terms of communication. English speakers use the structure-dependent question-forming rule, because they know that this rule is the rule for forming questions in English, and they know that other English speakers also know that, and so on. English speakers all use this rule because they know that it is the way of communicating the relevant experience and thought. [See Lewis (1975) and Lin (1999) for more detailed discussions on this topic.] The structure-dependent rule for forming questions is our (humans') way of communication. A structure-independent rule might be more efficient for the aliens (if the rule is their way of forming questions and if they know this), but it would not be efficient for us humans, as we do not know such a rule. So, Chomsky was wrong is asserting that to account for or explain rules such as the structure-dependent question-forming rule on the basis of functional considerations is a 'pretty hopeless prospect'.

5.4. Is generative grammar a branch of natural science?

In Section 4.6.2 above, we saw that Chomsky has been likening generative linguistics to natural sciences. And his method of doing generative linguistics does indeed have a strong scientific flavour. He tried to arrive at more and more general rules, principles or laws of grammar. But the preceding two subsections suggested that grammars are known to the speakers all along, that there is nothing about grammar to be discovered. This is to say that there is, and there need to be, no science of grammar. Thus, I do not see how generative linguistics is a science. In the next section I shall propose a different linguistic research program, which is genuinely scientific.

6. Discovering 'discovery procedures'

We saw in Sections 4.4 and 4.6 above that, because Chomsky regarded grammar as being unknown, he argued that it cannot be discovered by some taxonomic, mechanical procedures. Instead, he proposed that linguistics should aim at providing evaluation procedures for choosing among possible grammars, not aim at searching for a discovery procedure for mechanically yielding the grammar of a language given primary linguistic data as input. There was a change of aspect. That change of aspect was the turning point of Chomsky's linguistic thinking. It was also the breaking point between generative linguistics and taxonomic linguistics. It was the starting point of the 'Chomskyan revolution'.

I shall propose here that we need **another change of aspect**. The new aspect is that grammar is known. We know what the grammatical rules of our language are. We know why a sentence is grammatical or not. Nothing about grammar is hidden from us. What is hidden from us is knowledge-how, i.e. knowledge how we come to know all those grammar rules.

In this new aspect, our task is no longer to discover the grammar of a language, for there is nothing there about grammar to discover: everything about grammar is known already. Rather, our task is to discover how we (or the child) come to know the rules of grammar. We can make the assumption that the child is equipped with certain discovery procedures, such that given primary linguistic data, the child applies the discovery procedures and arrives at the grammar of the language in question. Under this assumption, our task will be to discover what the discovery procedures are.

We might start by assuming that the discovery procedures are procedures of induction, generalisation and analogy. But as Chomsky rightly pointed out, these notions are vague, and seem to be empty. Chomsky (1975a) also accurately remarked that taxonomic procedures of classification, segmentation, etc. are 'an unusually refined, detailed, and sophisticated development of a theory of this general [empiricist] character' (p. 13), are 'among the most sophisticated and interesting efforts undertaken within a significant (i.e. nonvacuous) empiricist framework' (p. 36), are 'an instance of E, perhaps the most complex version that has yet been developed' [Chomsky (1975b), p. 138]. (See Section 4.6.1 above.) We thus can carry on from where taxonomic linguistics stopped.

Notice that when we formulate discovery procedures we are not stabbing at the dark. It is not that we do not know what are trying to discover. Rather, our task is clear. The input is primary linguistic data, and the output is the set of rules of grammar, which we know. The research is empirical, because the rules of grammar, which we know, serve as tests for the adequacy of the discovery procedures. If the discovery procedures explain the acquisition of the grammar rules, then they are descriptively adequate; otherwise they are not. There might be different sets of discovery procedures which are all descriptively adequate. In that case, there would have to be some kind of evaluation procedure for choosing among them — simplicity might well be an important measure. But this is common in scientific inquiries. Our research is rather like finding the procedures (mechanisms) for determining

that *red* means the colour, not the shape, size, etc. Both are standard scientific inquires.

In comparison, the task of taxonomic linguistics before 1957 was not so clear. It tried to discover both the discovery procedures and the grammars of languages. But since it was not realised that grammars are known already and therefore does not need to be discovered, the old taxonomic linguistics fell prey to Chomsky's attack easily. If grammars are still to be discovered, then how can one know whether the discovery procedures, which are also to be discovered, are adequate or not? There must be something against which the adequacy of the discovery procedures can be tested. But there was no such a thing in the old taxonomic linguistics. So, finding discovery procedures was indeed like stabbing at the dark. Chomsky rightly observed:

> Parenthetically, we might observe that step 3 [i.e. the determinations of what is learned] is missing in many formulations of psychological theory, much to their detriment. In fact, even the concept "what is learned" is missing in familiar "learning theories." Where it is missing, the basic question of "learning theory" cannot even be formulated [Chomsky (1975b), p. 16].

Without knowing what is learned, what would be the point of a 'learning theory'? Without a way of evaluating the adequacy of discovery procedures, the practice of devising discovery procedures would be arbitrary and would lose all its significance. This was an important reason why Chomsky proposed that the task of the linguist was, at least at that moment of research, to provide an evaluation procedure, and not a discovery procedure (see Section 4.1 above).

In this new aspect, discovery procedures are regarded as being psychologically real. This is exactly as Chomsky suggested (cf. 4.6.1 above). We try to find the properties of the initial state of the child's mind. Instead of thinking that the initial state consists of innate grammatical principles, we postulate that it consists of a set of discovery procedures.

In this new aspect, finding discovery procedures will no longer be 'dull and unsatisfying' [Chomsky (1975a), p. 25]. The old taxonomic linguistics was 'dull and unsatisfying' because even to find a very simple grammar rule one had to go through the procedures of segmentation, classification, making use of classes, sequences of classes, etc., while one could easily write down the rule in question based on one's intuitions (cf. Section 4.1 above). But we are no longer trying to discover grammar rules. We are rather trying to discover discovery procedures. The search is as exciting as finding the neuro-algorithms for processing certain information.

The grammar rules in a language are known to the speaker. Section 5.2 above illustrated what such rules are. There are two types of grammar rules. The first consists of certain simple sentence types, such as (11)–(14). They are rules specifying the ways in which certain words should be used. The second type of grammar rules are transformations, which construct complex sentences, such as (10), out of simpler ones, such as (14). Harris (1970) presented a large list of transformations, most of which are known already. What is the nature of such sentence types and transformations? Lin (1999) argues that they are *conventions*: they are conventionalised ways

of describing certain experiences, and expressing certain thoughts. Because they are known to the speakers of a community, they make communication among the members possible. Discussing the idea of convention and communication would take us too far afield, but see Lin (1999) for details.

We know the grammar rules of our language all along. But how do we come to know them? The answer we postulate is that we are endowed with certain discovery procedures. Taxonomic linguistics has given us an idea of what such procedures might be. The discovery procedures may well be procedures of segmentation, classification, etc. This suggests that the (human) brain has certain built-in algorithms of processing linguistic data and finding the distributional structure. The discovery procedures, or algorithms, may well explain the speaker's 'speaking habits' Harris and Jespersen talked about (cf. Section 2.5.2 above).

What lies in front us, then, is a genuine linguistic science. The goal is clear, and so is its method. The theory of discovery procedures is empirical, it can be proved or disproved by whether it explains the acquisition of the grammar rules, which we know. This science does not suffer from any of the methodological problems Chomsky identified in the old taxonomic linguistics. The objective in this new science is indeed to develop 'a general mechanical procedure [or procedures] for constructing grammars of the appropriate type from the raw data' [Chomsky (1964a), p. 245], but note that 'grammars of the appropriate type' are already known to us. We are of course 'very far' from this objective yet [Chomsky (1964a), p. 245], but that is why it is an *objective*. It is highly reasonable to believe that this procedure (or procedures) 'will be a major achievement which will, for example, shed great light on the human ability to learn language' [Chomsky (1964a), p. 245].

7. Conclusions and discussion

In this paper I presented Harris's transformational grammar. Harris's theory, we saw, was well motivated. He believed that there are distributional structures in a language, and he sought to develop certain mechanical procedures for finding what they are. But in practice, Harris's grammar rules were first formulated on the basis of intuition and then their distributional properties were discussed. This shows that the rules were not really 'discovered' using the procedures, rather they had already existed somewhere, in one's intuition or knowledge. Indeed, the rules Harris's presented in his many articles on transformational grammar are rather intuitive. What Harris (and other structural linguists) did not see is that grammar rules are known to speakers already. (Most grammatical rules Harris presented in his many writings are more or less known to the speaker, cf. Section 2.3 above. There is no need to 'discover' such rules.) Because of this, he (and other structural linguists) did not realise that his task should be just developing the discovery procedures, rather than developing the procedures and discovering grammars at the same time.

The dual task of developing mechanical procedures and discovering grammars using the procedures, which the taxonomic (or structural) linguists imposed on themselves, was the ultimate cause of the downfall of taxonomic (or structural)

linguistics. If grammars were not known already and were yet to be discovered, then how could one develop a general, mechanical procedure for discovering them? No scientist would ever think about such a possibility, as Chomsky accurately pointed out. The right thing to do would be to set up ways of evaluating different grammars (theories) which are all compatible with available data — this is standard practice in science. This was exactly what Chomsky proposed to do — replacing discovery procedures with evaluation procedures. If grammars were yet to be discovered, then Chomsky's method of doing linguistics would be entirely correct.

Chomsky's linguistic thinking was very logical, and tremendously appealing. Suppose that a grammar is not known and is yet to be discovered. Then we must treat a grammar as a theory. We must then have ways of choosing among competing grammars (theories). Notice that the child eventually arrives at the grammar of his language, so we must consider how he can do this. Since we do not even know what the grammar of the language is, it is just odd to consider general learning mechanisms, such as procedures of generalisation, analogy, segmentation, classification, etc., quite apart from whether such notions can explain language acquisition. The only possible explanation of child language acquisition must be that the child is endowed with a set of innate grammatical principles, which specify in advance the format of human grammars and which significantly limit the number of possible grammars. Since we treat a grammar as a theory, our research is thus scientific. As in any other scientific inquiry, we must try to discover more general principles from which the less general ones can be derived. Hence we must pursue the minimalist program.

Thus, if grammars were not known already, then I do not see why Chomsky's linguistics should be criticised, let alone refuted. Neither do I see how one could succeed in refuting Chomsky's linguistics. This explains why Chomsky's linguistics, after having been criticised and attacked for forty years, is going as strongly as ever.

However, Chomsky's linguistic thinking, though remarkably logical and immensely compelling, was based on a false premise, which is that grammars are not known already. But Chomsky, being a great thinker, did not simply assert that grammars are innate and not known. He argued time and time again, by presenting a series of perplexing linguistic examples, that the knowledge of grammar involved must be innately given. His argument has considerable force, but there was a deep error in it. The error is that Chomsky failed to distinguish between knowledge of grammar and knowledge about how that knowledge is acquired. In other words, he failed to distinguish between knowledge-that and knowledge-how. His examples can be explained using rules of grammar which are known, though how the rules are acquired is still not yet known in any detail. The fact is that, to repeat, we do not yet know, in any detail, how knowledge of grammar is acquired. But this is independent of the fact that we do know the grammar of our language.[20]

[20] Matthews (1998) argues that all Chomsky's UG is purported to explain can be explained in terms of common sense constructions (cf. Note 19 above). He suggests that believing in UG is simply 'a matter of faith' [Matthews (1993), p. 252; Matthews (1998), p. 113]. But treating the belief in UG in this way does not seem to be satisfactory. In this paper I have tried to show that the thinking leading to UG was rather logical, but there was, unfortuately, a deep error in that thinking. I hope that once this error is seen there will be no basis for believing in UG.

It is by no means easy to see that we know the grammar of our language, that it is not hidden from us. Linguistics in the past forty years, or perhaps even since the time it was born, has proven this point. Certain analogies are probably what have prevented us from seeing that we know the grammar of our language all along. For example, we tend to think that knowledge of grammar is like knowledge of riding a bicycle, or is like knowledge of physics. We are tempted to think, due to such analogies, that grammar is hidden from us, that knowledge of grammar is either implicit or innate. But what we are tempted to think is not necessarily true. Instead of *thinking* that knowledge of grammar must be hidden, we should *look* at the relevant cases to see whether innate or implicit linguistic knowledge must be assumed. In Wittgenstein's words, 'Don't think, but look'.[21] The temptation to think that grammars are hidden from us and are yet to be discovered is tremendously strong, but we must fight against this great temptation. In tricky cases such as the ones discussed in Section 4.4, we must remind ourselves of the grammar rules which we have known all along, and we must arrange the known facts in such a way as to dispel any doubts (cf. Section 5.2 and also Section 5.3).

That grammars are known to us is the new aspect we need in linguistics. In this new aspect, the task for linguistics is not to discover grammars, for grammars are known to us already. The task is rather to discover how we, or the child, come to know the grammar, or the rules of the grammar, of our language. Seen from this angle, the taxonomic procedures of segmentation, classification, etc. are very plausible. There may well be such procedures in the child's brain, which enable him to pick up the distributional structure in the data presented to him, to come to know the rules of the grammar. Such procedures may well explain, say, why the child eventually comes to have the structure-dependent question-forming rule and not the structure-independent rule, because it may well be that the procedures *determine* that the child will have the former rule and not the latter. In other words, it may well be that the child cannot help but to have the structure-dependent rule and not the structure-independent rule. So the procedures may well explain why the child is able to learn a language so easily.

In this new aspect, what lies in front us is an exciting linguistic science. It is like the old taxonomic linguistics, but with a much clearer goal and a sounder philosophy. Many important questions can be seen in a new light. Firstly, why is there a huge difference between human language and animal language (if animals do have languages)? The difference may well lie in the discovery procedures: animals either do not have discovery procedures at all, or their discovery procedures are not as sophisticated as ours. Once we have enough details about our discovery procedures, we will be able to provide a substantial explanation for the difference between humans and animals (as far as language is concerned).

[21] It would be very natural for someone to argue: 'We do look at linguistic data! We look at data not only from one language, but also from a variety of languages. And that is how we came up with so many abstract principles!' My reply would be: 'Yes, you do look at linguistic data. But you have been looking from a wrong perspective. You need a change of aspect'.

Secondly, how did the human brain come to have discovery procedures? Were discovery procedures the direct result of evolution, or were they a by-product of something else which had been evolved? A lot of research will be needed in order to answer these questions. But they are sensible questions. At least, they seem to be as sensible as questions concerning the evolution of innate linguistic principles (Pinker and Bloom, 1990; Pinker, 1994).

Thirdly, are discovery procedures language-specific or they are used in other cognitive tasks as well? This question cannot be answered a priori: future research will be able to answer it. But it is reasonable to think that discovery procedures are not language-specific. What discovery procedures do is, in essence, to find certain regularities, or patterns, in the primary linguistic data. And finding regularities and patterns is rather common in other cognitive tasks.

Fourthly, what is the relationship between this linguistic science and today's computer technology? It seems that the former will be greatly helped by the latter. With the great computing power computers have, we can simulate discovery procedures (algorithms) on computers quickly and easily. Indeed, there have been computer systems which can classify words into classes on the basis of their distribution.

Fifthly, how is the creativity of language to be explained? This is not difficult to answer in this new framework. The speaker of a language learns, using discovery procedures (and probably some other things, to be discussed in the following paragraph), the grammars rules of his language. These rules are finite in number, but they can generate an infinite number of sentences. Harris's transformational grammar shows how this is done (see Section 2.5.3 above). See also Jespersen (1924, 1937) and Lin (1998). For the nature of grammar rules, see Lin (1999).

Lastly, are discovery procedures sufficient for the child to arrive at the grammar of the language? Again, we cannot, at this moment, answer this question precisely. But it is reasonable to suppose that they are not sufficient. The child may need to have some innate notions (or concepts), such as the notions of general/specific signification, of definiteness, of comparatives, of negation, of numbers, of times, of sexes, of animateness, etc. (Jespersen, 1924; Strawson, 1959, 1974). The child may also need some mechanisms of reinforcement, no matter how vague the concept 'reinforcement' is at present (Skinner, 1957). This is because it is doubtful that the child can learn a language or acquire its grammar simply by listening to a radio. On these issues, see Lin (1998).

Newmeyer (1996b) argued that there was a 'Chomskyan revolution'. The present investigation confirms Newmeyer's view: yes, there was a revolution; but it was primarily a revolution in linguistic thinking, that is, a change of aspect. The 'Chomskyan revolution' is still going on, but the storm has long passed. We are presented with a period of tranquillity for serious reflections (Botha, 1989; Newmeyer, 1986, 1996a,b; Matthews, 1993; Harris, R.A., 1993; Seuren, 1998; Lin, 1998, 1999). No more heated exchanges, no more hasty decisions, no more passionate assertions or rejections — just quiet reasoning. It is serious reflections and quiet reasoning that the present paper is based on. This work has concentrated mainly on the early development of Chomsky's linguistic thinking (chiefly from 1950s to

1970s). [For an account and analysis of Chomsky's more recent writings on generative linguistics and its philosophy, see Lin (1999)]. But as shown in Section 4 above, Chomsky's research program has been remarkably consistent. In the words of Newmeyer et al. (1996, p. 79), Chomsky's program is 'one of foundational stability and continuity'. Once Chomsky's early linguistic thinking is well understood, the more recent development in his linguistic theory then becomes much easier to comprehend.

All in all, the Chomskyan revolution started with a change of aspect. That aspect was that a grammar is unknown and should therefore be treated as a scientific theory. That change of aspect gave rise to several transformations of Harris's conception of transformational grammar. Discovery procedures were replaced by evaluation procedures. Transformations, which were seen as a kind of realities by Harris, were regarded by Chomsky as mere abstract elements in a theory. Chomsky's transformational theory was then changed into P&P, which then developed into the minimalist program. But Chomsky's change of aspect was faulty, because grammars are known to us all along. The fundamental error in Chomsky's linguistic thinking is the confusion of knowledge of grammar and knowledge of how that knowledge is acquired. That grammars are known to us all along is the new aspect we must get used to. In this new aspect, all the above transformations of Harris's transformational grammar, natural though they might have been, were unnecessary. This new aspect gives rise to a linguistic science, which is very much like the old taxonomic linguistics but is on a sounder scientific and philosophical footing. The once ridiculed and abandoned discovery procedures, seen in this new aspect, may indeed discover the deep secrets of human language and the human mind. It might look as if that we are moving backwards, to the old taxonomic linguistics. But that would only be partially true. We are, in a sense, back to the old spot, but with a new understanding, a new way of seeing things, and a new state of mind. This we would not have achieved without Chomsky's rich, deep, intensive, and extremely engaging work on generative linguistics.

Acknowledgements

I would like to thank Peter F. Strawson, James Higginbotham, Frederick J. Newmeyer, Gordon P. Baker, Pieter A.M. Seuren for comments and suggestions. I am also grateful to Peter H. Matthews for sending me two of his most recent articles, which are very relevant to this paper. But none of these scholars are responsible for the view expressed here.

References

Botha, R.P., 1989. Challenging Chomsky: The Generative Garden Game. Blackwell, Oxford.

Chomsky, N., 1956. Three models for the description of language. I. R. E. Transactions on Information Theory (IT-2), 113–124.

Chomsky, N., 1957. Syntactic Structures. Mouton & Co, The Hague.
Chomsky, N., 1959. Review of Joseph H. Greenberg, Essays in Linguistics. Word 15, 202–218.
Chomsky, N., 1961a. On the notion of "rule of grammar". In: Jakobson, R. (Ed.), Structures of Language and Its Mathematical Aspects. American Mathematical Society, Providence, 6–24. [Reprinted in Fodor and Katz (Eds.) (1964), 119–136].
Chomsky, N., 1961b. Some methodological remarks on generative grammar. Word 17, 219–239.
Chomsky, N., 1964a. A transformational approach to syntax. In Fodor and Katz (Eds.) (1964), 211–245. Originally published in A.A. Hill (Ed.), Third Texas Conference on Problems of Linguistic Analysis in English (9–12 May 1958). University of Texas, Austin, 1962, 6–24.
Chomsky, N., 1964b. Current Issues in Linguistic Theory. Mouton & Co., The Hague. [First presented at the Ninth International Congress of Linguistics, Cambridge, MA, 1962].
Chomsky, N., 1964c. The logical basis of linguistic theory. In: Lunt, H. (Ed.), Proceedings of the Ninth International Congress of Linguistics [1962]. Mouton, The Hague, pp. 914–978.
Chomsky, N., 1965. Aspects of the Theory of Syntax. MIT Press, Cambridge, MA.
Chomsky, N., 1966. Cartesian Linguistics. Harper & Row, New York.
Chomsky, N., 1972. Language and Mind (enlarged Edition). Harcourt Brace Jovanovich, New York. (First edition of Chs. 1–3, 1968.)
Chomsky, N., 1973. Conditions on transformation. In: Anderson, S.R., Kiparsky, P. (Eds.), A Festschrift for Morris Halle. Rinehart & Winston, New York, pp. 232–286. [Reprinted in Chomsky (1977a), 81–160].
Chomsky, N., 1975a. The Logical Structure of Linguistic Theory. Plenum Press, New York, 1975. [Typescript described as dating from 1955; the Introduction, 1–53, written in 1975].
Chomsky, N., 1975b. Reflections on Language. Pantheon Book, New York.
Chomsky, N., 1976. Conditions on rules grammar. Linguistic Analysis 2, 303–351. [Reprinted in Chomsky (1977a), 163–210].
Chomsky, N., 1977a. Essays on Form and Interpretation. North-Holland, New York.
Chomsky, N., 1977b. On wh-movement. In: Culicover, P.W., Wasow, T., Akmajian, A. (Eds.), Formal Syntax. Academic Press, New York, pp. 71–132.
Chomsky, N., 1980a. Discussion of Putnam's comments. In Piattelli-Palmarini, M. (Ed.) (1980), 310–324.
Chomsky, N., 1980b. Rules and Representation. Blackwell, Oxford.
Chomsky, N., 1981. Lectures on Government and Binding. Foris, Dordrecht.
Chomsky, N., 1986. Knowledge of Language: Its Nature, Origin, and Use. Praeger, London.
Chomsky, N., 1988. Language and Problems of Knowledge: The Managua Lectures. MIT Press, Cambridge, MA.
Chomsky, N., 1995a. Language and nature. Mind 104, 1–61.
Chomsky, N., 1995b. The Minimalist Program. MIT Press, London.
Chomsky, N., 1998. The Minimalist Inquiry: the Framework. MIT Occasional Papers in Linguistics 15.
Fodor, J.A., Katz, J.J. (Eds.), 1964. The Structure of Language; Readings in the Philosophy of Language. Prentice-Hall, Englewood Cliffs.
Haegeman, L., 1994. Introduction to Government and Binding Theory, 2nd Edition. Blackwell, Oxford.
Harris, R.A., 1993. The Linguistic Wars. Oxford University Press, Oxford.
Harris, Z.S., 1944. Simultaneous components in phonology. Reprinted in Harris (1970), pp. 1–31.
Harris, Z.S., 1946. From morpheme to utterance. Reprinted in Harris (1970), pp. 100–125.
Harris, Z.S., 1954. Distributional structure. Reprinted in Harris (1970), pp. 775–794.
Harris, Z.S., 1955. From phoneme to morpheme. Reprinted in Harris (1970), pp. 32–67.
Harris, Z.S., 1956. Introduction to transformations. Reprinted in Harris (1970), pp. 383–389.
Harris, Z.S., 1957. Co-occurrence and transformation in linguistic structure. (Presidential address, Linguistic Society of America, 1955.) Reprinted in Harris (1970), pp. 390–457.
Harris, Z.S., 1959. Linguistic transformations for information retrieval. Reprinted in Harris (1970), pp. 458–471.

Harris, Z.S., 1963. Immediate-constituent formulation of English syntax. Reprinted in Harris (1970), pp. 131–138.
Harris, Z.S., 1964a. Transformations in linguistic structure. Reprinted in Harris (1970), pp. 472–481.
Harris, Z.S., 1964b. The elementary transformations. Reprinted in Harris (1970), pp. 482–532.
Harris, Z.S., 1965. Transformation theory. Reprinted in Harris (1970), pp. 533–577.
Harris, Z.S., 1966. Algebraic operations in linguistic structure. Reprinted in Harris (1970), 603–611.
Harris, Z.S., 1970. Papers in Structural and Transformational Linguistics. D. Reidel, Dordrecht.
Harris, Z.S., 1988. Language and Information. John Wiley, New York.
Harris, Z.S., 1991. A Theory of Language and Information: A Mathematical Approach. Clarendon Press, Oxford.
Hockett, C.F., 1952. A formal statement of morphemic analysis. Studies in Linguistics 10, 27–39.
Jeckendoff, R., 1977. X-bar Syntax: A Study of Phrase Structure. Cambridge University Press, Cambridge.
Jespersen, O., 1924. The Philosophy of Grammar. Allen and Unwin, London. [Reprinted in 1992 by University of Chicago Press, Chicago].
Jespersen, O., 1937. Analytic Syntax. Allen and Unwin, London. [Reprinted in 1984 by University of Chicago Press, Chicago].
Katz, J.J., Postal, P.M., 1964. An Integrated Theory of Linguistic Descriptions. MIT Press, Cambridge, MA.
Lewis, D., 1975. Language and languages. In: Gunderson, K. (Ed.), Language, Mind, and Knowledge. University of Minnesota Press, Minneapolis.
Lin, F.Y., 1998. Jespersen, Skinner and Chomsky on language. Somerville College, Oxford, (manuscript).
Lin, F.Y., 1999. Chomsky on the "ordinary language" view of language. Synthese 120, 151–192.
Matthews, P.H., 1993. Grammatical Theory in the United States from Bloomfield to Chomsky. Cambridge University Press, Cambridge.
Matthews, P.H., 1998. Should we believe in UG? In: Janse, M. (Ed.), Productivity and Creativity: Studies in General and Descriptive Linguistics in Honour of E. M. Uhlenbeck. Mouton de Gruyter, Berlin, pp. 105–113.
Matthews, P.H., 1999. Obituary - Zellig Sabbettai Harris. Language 75, 112–119.
Newmeyer, F.J., 1986. Linguistic Theory in America, 2nd Edition. Academic Press, New York.
Newmeyer, F.J., 1996a. Generative Linguistics: A Historical Perspective. Routledge, London.
Newmeyer, F.J., 1996b. Has there been a "Chomskyan revolution" in linguistics? Language 62, 1–19. [Reprinted in Newmeyer (1996a), pp. 23–38.]
Newmeyer, F.J., Anderson, S.R., Chung, S., McCawley, J., 1996. Chomsky's 1962 programme for linguistics: a retrospective. In Newmeyer (1996a), pp. 66–79.
Piattelli-Palmarini, M. (Ed.), 1980. Language and Learning: The Debate Between Jean Piaget and Noam Chomsky. Routlege & Kegan Paul, London.
Pinker, S., Bloom, P., 1990. Natural language and natural selection. Behavioural and Brain Sciences 13, 707–727.
Pinker, S., 1994. The Language Instinct: The New Science of Language and Mind. Allen Lane, London.
Postal, P.M., 1962. On the limitation of context-free phrase-structure description. Quarterly Progress Report 64, Research Laboratory of Electronics, MIT, 231–238.
Postal, P.M., 1964a. Constituent Structure: A Study of Contemporary Models of Syntactic Description. Mouton & Co, The Hague.
Postal, P.M., 1964b. Limitations of phrase structure grammars. In: Fodor and Katz (1964).
Putnam, H., 1980. What is innate and why: comments on the debate. In: Piattelli-Palmarini (Ed.) (1980), pp. 287–309.
Rizzi, L., 1982. Issues in Italian Syntax. Foris, Dordrecht.
Ross, J.R., 1967. Constraints on Variables in Syntax. PhD dissertation, MIT.
Ryle, G., 1946. Knowing how and knowing that. Proceedings of the Aristotelian Society, XLVI.

Searle, 1972. Chomsky's Revolution in Linguistics. New York Review of Books, 29 June.
Seuren, P.A.M., 1998. Western Linguistics: A Historical Introduction. Blackwell, Oxford.
Skinner, B.F., 1957. Verbal Behavior. Methuen, London.
Strawson, P.F., 1959. Individuals: An Essay in Descriptive Metaphysics. Routledge, London.
Strawson, P.F., 1974. Subject and Predicate in Logic and Grammar. Methuen, London.
Wells, R.S., 1947. Immediate constituents. Language 23, 321–343.