

M. Gross

TAXONOMY IN SYNTAX

The term 'taxonomy' has a poor reputation in contemporary linguistics.

Whereas classification is still considered to be an indispensable, a respectable and respected enterprise in any science, this is not true in linguistics (neither in phonology nor in syntax).

I have asked myself about the reasons of this state of affairs and have come to the conclusion that it can be traced back to a wide variety of reasons but - strangely enough - all of them are exterior to linguistic science.

Most of the work performed at the Laboratoire d'Automatique Documentaire et Linguistique may - at first glance - appear as a taxonomy of the sentence types of French, and may, in this sense, reflect a possible theorizing attitude as it is found in a variety of true scientific activities.

An inquiry into syntax is exactly of the same nature as an inquiry into any set containing a huge number of objects. The number of objects, in the present case the number of sentences, makes it practically impossible for a human being to perceive their organization (structure) in a direct manner, or even to discover the existence of a possible organization. It is thus quite natural to group similar objects into classes and to examine the classes thus constituted. Since the number of classes is less than the number of objects, it may be hoped that such attempts will be suggestive of ways in which they are interconnected. Then, one may be able to infer an organization for the initial objects. This is certainly a commonplace

and any scientist would take such a reminder as an offense. Nevertheless, it has almost completely been forgotten in present day linguistics^{1/}

Thus, generative grammar opposes the construction of logical or mathematical models to descriptive approaches which are termed procedures. Many different models have been constructed; they are supposed to have predictive and explicative power. However, they are constructed on the basis of a rather limited set of empirical observations. Only count as linguistic facts those phenomena which may serve to validate or to invalidate a given model. Very often, the basis of such models consists in phenomena which do not have much in common. For example, both the passive and the coordinating conjunction are considered to be transformations. At the same time however, quite a few problems of importance are not even taken up in discussions, e.g. the determination of the lexical entries that correspond to a given form (see further below).

This frenzy of constructing models may only be due to the fact that people tend to take literally the famous truism which says that "language is a system in which everything depends on everything". This "axiom" seems to legalize the study of any interaction between any phenomena. In this approach - even if the linguistic facts themselves are authentic - facts are taken arbitrarily out of a population of phenomena whose size has never been estimated. Such an estimation is quite possible today. One is then forced to realize the gap which separates this infantile stage of linguistic research from the stage reached by exact sciences. In the sciences, specialists have succeeded in sorting out from the enormous mass of phenomena that have been studied systematically those that seem to be fruitful and general to-day. In spite of the fact that they too, tend to state that in Nature everything has its place (an extreme determinism, but which, in principle, is considered correct by minds imbued with formal logic), they were forced to eliminate from their object of study, a great number of marginal facts; even if these facts happened to be the most obvious ones.

The approach advocated by Chomsky and the users of his formalism (with or without modifications) suffers thus from the lack of empirical foundation which would permit the con-

struction of scientific theories (and not just of theories in the sense of human sciences). Thus, although it has been clearly demonstrated that structural relations depends heavily on the lexical entries in question, no attempt has been made so far to describe the distribution of these interrelations with respect to the various sets of lexical entries, though such an enterprise is feasible.

The transformation often referred to as RAISING is a rather characteristic example of this practice. For most linguists, there exist two types of RAISING. The first type can be illustrated by sembler - to seem:

Il semble que Max dorme
 'It seems that May is sleeping'
 Max semble dormir
 'Max seems to be sleeping'

The second type can be exemplified by s'imaginer-to imagine:

Max s' imagine que Luc est prêt
 'Max imagines that Luc is ready'
 Max s' imagine Luc (être) prêt

A number of publications^{2/} seem to be devoted solely to find out whether the examples with s'imaginer are linked by a rule which has the status of a transformation (taken in its restricted technical sense) or not. It would seem that there is no dispute about the transformational connection between the examples with sembler. However, no one has ever constructed a list of the verbs of type s'imaginer-to imagine and a list of the verbs of type sembler-to seem. We have at our disposal such a list for French: it turns out that there are three verbs of type sembler and more than six hundred verbs of type s'imaginer in French^{3/}. This indicates that the type of s'imaginer is of general character, in contrast with the type of sembler; and nothing permits us to predict such a situation. It is thus unjustified to put more emphasis on sembler-to seem in a grammar than on s'imaginer. Of course, the inverse would not be justified either. It goes without saying, however, that these observations have to prompt the question as to the significance of this distribution, since - as we know - lexicon and sentence structure are closely interrelated. The systematic investigation of the lexicon thus constitutes a means, perhaps the only one at our disposal at present, to understand language in a comprehensive way, i.e. to construct a

picture of language which has a certain generality. It is only this framework that permits us to discover a wide range of new phenomena which are central to language and possibly contrast them with marginal cases or exceptions. Generative grammar thus can only be considered as a limited procedure, in fact, as a pragmatic and somewhat arbitrary procedure, which to a certain extent is dependent of the fashion phenomena that it generates.

There may be various reasons to elaborate taxonomy in syntax. First, one should think that a classification might serve as an experimental verification of transformational theories. These theories have always been constructed on the basis of very few examples. The large number of lexical entries could thus be used to vary transformational rules considerably. Our first idea was thus to verify the validity of the various rules for as many lexical entries as possible. However, it turned out very soon that this verification had entirely displaced the problems of syntax, and that the main questions raised by generative grammar had lost their priority with respect to problems which first appeared to be questions of detail only (for example, the assessment of the number of lexical entries corresponding to a given verb) and also with respect to problems brought about by the undeniable existence of an important number of non-transformational relations between sentences.

The problem of assessing the number of lexical entries that correspond to a given form is far from being simple. A rather elementary exemple may suffice to illustrate the type of questions which may crop up in this connection. Consider the following two sentences:

Max continue à travailler
 Max continue de travailler
 'Max continues to work'

The verb continuer can take two different prepositions. This syntactic difference may force us to hypothesize two verbs continuer: continuer à and continuer de. However, formal resemblance as well as our understanding of the meaning of these verbs run counter to such a hypothesis. The only way to describe syntactically the two sentences would be to set up a rule which would make the prepositions à and de equivalent, e.g. à = de (or de = à) in the context of continuer. But the justification

for incorporating such a rule into grammar has to be based on more reliable criteria than our intuition about the identity of the two verbs. Unfortunately however, it is rather difficult to find facts that would justify this rule, which is, at least in part, due to the fact that the contexts in which their complementarity can be observed are extremely rare, and they do not shed any light on the situation. Such problems arise for almost all verbs in French, with the following general question: should the literal use of a verb and its metaphorical use constitute a single lexical entry, or two lexical entries from the point of view of syntax (and semantics).

The same problem leads sometimes to conclusions of quite different types. Consider the following two sentences:

Ce comportement étonne Max
 'This behavior surprises Max'
 Max s'étonne de ce comportement auprès de Luc
 'Max is surprised at this behavior and tells
 it to Luc'

In contrast to the case of continuer, these two uses of étonner are semantically as well as syntactically different. On the one hand, the meanings of the two sentences are different: in the second sentence étonner has the meaning of a verb of communication, in the first sentence it has not. On the other hand, there is no possibility of deriving transformationally one sentence from the other because of the fact - among other things - that the complement phrase auprès de Luc is unacceptable in the first sentence:

Ce comportement étonne Max auprès de Luc
 This behavior surprises Max Prep Luc'

The observations made in this case may thus lead us to the conclusion, in a more convincing manner than in the case of continuer, that we are dealing here with two distinct verbs: one is étonner which is transitive, and the other s'étonner de which admits the complement phrase auprès de N. However, a closer examination of a more complete list of verbs in French (for example, the list of verbs of the conjugation manual by Bescherelle) reveals that pairs such as étonner - s'étonner de appear rather frequently in the lexicon and are thus not a quasi-exceptional phenomenon as the pair continuer à - continuer de was. A great number of verbs which may all be termed psychological admit the two constructions exemplified above

(e.g. affliger 'afflict', alarmer 'alarm', irriter 'irritate', désoler 'distress', réjouir 'rejoice' and several dozens more). In addition, the meaning difference observed between the two forms is always exactly the same. In this way, one establishes that there exists a relationship between the two constructions with étonner, but there is no (at least given the present stage of our knowledge) well-motivated formal means (or transformation) to connect these two forms, and thus trace back the two lexical entries of étonner to a single one.

Such examples carry to transformational generative grammar a criticism that has already been applied to distributional grammars. We have found many types of pairs of sentences whose members cannot be or should not be related to each other by transformations, but which nevertheless are clearly related to each other. The failure of transformational grammars to describe these relations constitutes an inadequacy that is equally serious as the inadequacy of distributional grammars to handle the passive. To be sure, Harris and Chomsky have proposed new rules which were able to remedy the defects of previous accounts. I, however, do not see at present what kind of non-transformational rules would account for the difficulties observed. I think, however, that the only way to arrive at a better understanding of these phenomena is to accumulate data as varied as possible, i.e. to construct a taxonomy as complete as possible of the phenomena at hand. It is important to realize that the transformations are far from being able to cover the whole domain of syntactic phenomena and that there are problems which thus far have not even been considered by most linguists for the simple reason that they have never thought that a taxonomic approach might be fruitful. Another point which shows that classification is necessary can best be shown by considering the way in which classifications are used in biology.

When a syntactic phenomenon depends on a lexical entry, for example on the main verb, the theoretical study of the phenomenon is certainly not indifferent to the choice of particular examples, since such examples are used to verify generalizable intuitions of acceptability. In other words, theoretical results will heavily depend on the choice of particular lexical entries and there is nothing surprising in that. But when selecting lexical entries, linguists could take advantage of the analogous procedures in standard use in biology. No doubt, all biological studies have - more or less

directly - made use of classifications. Thus, heredity, the structure of proteins and of nucleic acids, the trajectory of nervous influx are problems of all living beings and yet, they are never studied on species taken arbitrarily. Some species are privileged from the experimental point of view for various reasons: size, configuration, fast reproduction, easy breeding. In this way, scientists have chosen *drosophila melanogaster* or *neurospora* in a great number of studies on heredity, *escherichia coli* in genetics and *aplysia* in studies of the nervous system. All these choices have been motivated in detail with respect to the multitude of other possible *a priori* choices. The classification has played and continues to play a crucial role for the choice of the object of experiments. It is thanks to this taxonomy that it has been possible to formalize and to diffuse an enormous quantity of information which has been carefully verified and accumulated during centuries. On the basis of this information, it was then possible to track down the general phenomena. This has been a necessary condition for the constitution of biology as an exact science, and I think, the same is valid for the study of language. I am convinced that the existence of a systematic classification for sentence forms would permit linguists to select those particular lexical entries which lend themselves best to experiments in given areas rather than others, or which might help to construct or to verify relatively general theories. At present, the choice of lexical entries is made practically randomly, which has the consequence that the same experiment very often cannot be reproduced, and thus the theories have to undergo crucial revisions as time goes by.

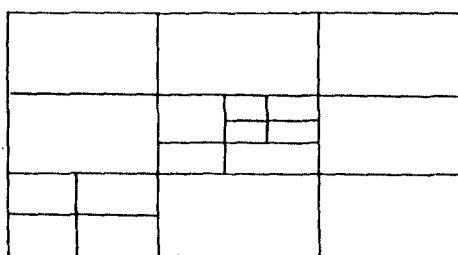
Let us discuss the notion of reproducibility of an experiment. This notion has never been mentioned in generative grammar, and the reason for this is not because this type of linguistics which tends to be scientific is not an experimental science. The notion of experiment used to justify a hypothesis is in fact latently present in the activity of constructing examples whose acceptability is then verified by the linguists. The absence of reference to this notion does not have the same reason as it has in 'hard sciences', where it is hardly mentioned for the simple reason that it constitutes an obvious condition of existence for any experiment. However, the colleagues of a specialist who announces a surprising result will immediately attempt to reproduce the same experiment or, at least, verify all the conditions and all its details. It would

seem that the majority of linguists are simply not aware of such rules that constitutes the basis of any scientific procedure. The reason for this is perhaps that it is so easy to find extralinguistic causes for the explanation of the variation in acceptability judgments, that linguists do not make any effort to vary the experimental conditions that insure the stability of the result. Thus, whenever there is disagreement as to the existence of a phenomenon or as to the justification of a hypothesis, it is always easy and intuitively quite natural to refer to dialectal or sociological variations (i.e. to various layers of language). One could even refer to diachronic variations, since quite often, to make use of literary language means using forms which have acquired general hence historically conditioned usage. It is this elementary intuition about the causes of variation that most linguists never try to overcome, although it can often be done, for example, by varying the lexical conditions of the experience in a systematic fashion, or by making use of privileged lexical entries. Of course, only a considerable accumulation of experiments could privilege certain lexical entries over others. Such elements could be distinguished, for example, by finding out that their use is more widespread than the use of others, and that they are largely independent of accidental variations. Another reason for the neglect by linguists of the reproducibility of their experiments seems to be due to the absence of the tradition already mentioned above: that of accumulating experiments in a systematic fashion. Thus, for example, if one is engaged in examining the lexical distribution of a transformation that depends on the main verb, one is often confronted with cases of dubious acceptability which are situated between cases that are quite acceptable and cases that are clearly unacceptable. This situation should bring to the fore and in a natural way the problem of the reproducibility of experiments and thus, of the empirical foundations of the linguists' hypothesis about the transformation he considers. This, however, never happens in generative grammar.

Let us come back to the classifications made by traditional grammarians. One should first note that these classifications consists of disjointed classes, i.e. of equivalence classes in the sense of algebra. In addition to the division into classes (which corresponds roughly to the classification of words according to the parts of speech), we also find a number of further subdivisions which can perhaps be grouped hierarchically. But such a system has not been made explicit.

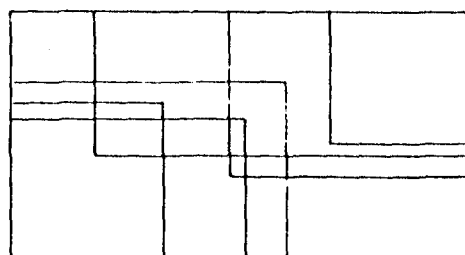
One could also think of dividing a set into non-disjoined classes, i.e. into subsets whose intersections are not empty. It is clear, however, that this second procedure would be impracticable even for the smallest classes. Whereas it is easy to grasp a partition of elements into disjointed classes (see fig.1) it is practically impossible to get a direct perception of the same set, if the classes are divided in a more complex way (see fig.2), i.e. when the formal relation that connects the elements is not that of equivalence. The problem is whether it is possible and interesting to define systems of classes which are more general. This problem is partly empirical, partly of a formal nature.

Fig.1.



Division of a set into 18 disjoint subsets. A hierarchy may be established on the basis of the different sizes of the rectangles.

Fig.2.



Division of a set into nondisjoined classes where, for example, each subclass can be considered as a rectangle having two sides on the sides of the largest rectangle. However, other interpretations are equally possible.

The perceptual reasons for the difficulty in understanding this latter system of classes are rather evident. Therefore, all important classifications which have been constructed so far are based on disjointed classes. Notice that one way

to represent graphically the disjunction of classes is a classificatory tree where the characteristics determine branching. Such a representation is found in the classifications of zoology and botanics. The choice of such a representation in grammar would raise quite a few questions. Consider, for example, the word peu 'little, few'. This word appears in the nomenclatures with the following characteristics: determiner, article or prearticle as in the sentence

Peu d'amis sont venus
'Few friends came'

pronouns as in Peu sont venus
'Few came'

and adverb as in Max dort peu
'Max sleeps little'

For well-motivated and independent reasons, the three classes of peu: determiners, pronouns and adverbs form disjoined classes. Consequently, the analysis of peu in terms of the characteristics of these classes forces us to distinguish lexical entries for peu which only have in common the (rare) properties shared by the corresponding parts of speech. It is quite evident that such a solution is not satisfactory, though the grammarians who have adopted this system have proceeded in a logical manner, preserving the coherence of their system of disjoined classes.^{4/}

It is interesting to note that the examples discussed in generative grammar reflect a different position with respect to the organization of the lexicon. It goes without saying that lexical entries must be classified in generative grammar as well, but the formalism which is used has for a consequence that the classes obtained are not disjoined, but intersecting classes. In fact, lexical entries would be characterized by syntactic or semantic features generally considered to be binary as shown in the following example:

[	sincérité + N + Dct + Count + Abstract	]
---	--	---

On the basis of such lexical representations, it is impossible to discover any organization of the lexicon. Notice that this description of lexical entries by complex symbols has been imposed on generative grammar by the priority which is accorded in this grammar to the study of rules, more precisely to the study of the form of the rules and of their interactions. The main objective of the study of syntax has thus to do with systems of rules, the lexical entries play only a secondary role. The main reason for the absence of serious lexical studies in the framework of generative grammar has to be sought in this conception of syntax.

The studies performed at the L.A.D.L. are thus aimed to fill certain important gaps in the linguistic research of the last few years. They may also serve as a model for the classification of sentences in other languages. I am convinced that there is no other way to construct theories, if one wants to take this term in the sense of natural sciences and not in its wide-spread metaphorical sense often used in human and social sciences. Classifications constitute the only practicable and empirical basis for theories, to-day. On the basis of theories thus obtained, one could develop more advanced theories which would enable experimental verifications, and explanations analogously to the situation in the 'hard sciences'.

The classifications of simple sentences in French and Italian⁵ which have been worked out at the Laboratoire d'Automatique Documentaire et Linguistique can be considered as a modest step towards this goal. It would seem that the classification of syntactic forms has not found its Linneus as yet. The characteristics employed are not always satisfactory, but they are the only ones available; they have been extracted from the mass of reproducible properties of distributional and transformational syntax. The partition into tables, i.e. into disjointed classes, is primarily based on the number and the form of the complements of the verb. This subdivision as well as the detailed study of a number of other subdivisions has not revealed the general organization (structure) of the system of simple sentences, but it has, no doubt, shown in many ways that such an organization does exist. The insufficiency of syntactic criteria makes it necessary to use non-syntactic procedures as well. Several cases suggest that there are close links between

meaning and form. These links have emerged in various sections of the lexicon where the classification has been carried out on the basis of purely syntactic criteria. These links have thus encouraged the investigators to make use of semantic properties in the case where syntactic criteria were no longer available. This approach has, however, been restricted to certain cases since, in general, we accept the structuralist criticism of the use of semantic notions in grammatical description. In fact, semantic notions have not been operational in the great majority of cases. However, the dimensions of our study has made it possible to verify that semantic notions could be reproducible in certain cases. These notions could then be used as criteria for the partitioning of the lexicon.

The researchers have also taken into consideration characteristics that ease the practical handling of the classifications. The experience acquired over more than ten years has indicated that the manipulation of the elements of a class is not efficient unless the number of elements is kept between one and two hundred. The classification to be presented has to some extent been adapted to this purely extrinsic requirement. In some cases this has led to subdivisions which may seem arbitrary, but which are destined to facilitate the utilization of the tables. It has to be stressed, however, that the classification is but subsidiary, the goal of the research is theoretical. The accumulation of the data has been made necessary in order to shed light on the combinatorial structure of simple sentences, or, to put it differently, on the traditional problem of verb rection, which is in itself a fairly complex question. This enterprise is exceptional in transformational grammar. There are several reasons which have led to this situation. Some of them may be sought in the evolution of traditional grammar where classification is largely restricted to the lexicographers' practice. Others may be due to the criticism of structural grammars advanced by Chomsky.^{6/} It seems worthwhile to analyze all these reasons in the course of a presentation of a taxonomical system.

It is rather likely that already the grammarians of antiquity correctly recognized the problem of studying the syntax of a given language, and that, first of all, they realized the necessity of classifications. In the course of

the history of grammar, there are certainly periods when the difficulties connected with classification have been overestimated and other periods where they have been underestimated. This can be gathered from Jean-Claude Chevalier's study^{7/} which discusses relatively recent periods of grammatical research, but also from the grammatical practice of the last few decades, when conceptual revolutions^{8/} have completely renewed the arsenal of methods of linguistic analysis.

Let us imagine a possible genesis of grammatical studies along the following lines. The first linguists realized the extraordinary richness of forms that discourses exhibit, but at the same time they also discovered that there existed a great number of similarities and dissimilarities between various discourses. From the given data, they would thus extract and elaborate the characteristics of these similarities and dissimilarities. It is likely that the first criterion of similarity (apart from mere repetition of words) to have been discovered is the parts of speech. To give to words general names such as substantive, verb or article means to constitute word classes. Such classes enable us to acquire a general picture of the lexicon of a language, since the lexical elements of this language fall then into a small number of types. This, in turn, makes the following operation possible: the names of parts of speech can be applied to sentences which, then, yield sentence forms. This operation, which is a type of morphism, preserves the syntactic structure of the discourse. For example, the operation

<u>garçon</u> -boy	→	<u>N</u>
<u>fil</u> le-girl	→	<u>N</u>
<u>ville</u> -city	→	<u>N</u>
<u>regarde</u> -look	→	<u>V</u>
<u>le</u> -the	→	<u>Art</u>
<u>la</u> -the	→	<u>Art</u>

would assign the unique type

Art N V Art N

to the set of six sentences:

'the boy looks at the girl'
 le garçon regarde la fille
 'the boy looks at the boy'

le garçon regarde le garçon
 'the boy looks at the city'
 le garçon regarde la ville
 'the girl looks at the girl'
 la fille regarde la fille
 'the girl looks at the city'
 la fille regarde la ville
 'the girl looks at the boy'
 la fille regarde le garçon

One sees the economy realized when one performs operations on sequences of parts of speech instead of sequences of words.

The procedure that I have just sketched is essentially that of Bloomfieldian structuralism, and I have thus condensed into one single stage a long succession of stages and discoveries spread over at least two thousand years.

Truly enough, given any set of related sentences, there is an abundance of criteria that a priori would be relevant for the classification of these sentences: a multitude of intuitions with respect to the form or the meaning^{9/} of sentences offer themselves simultaneously, and it is far from easy to find out which ones are interesting, i.e. which ones may contribute to a better understanding of language. The history of grammatical concepts is a steady development toward rational criteria, often interrupted by periods of regression, i.e. periods characterized by discussions about academic refinements of grammatical concepts whose operational meaning had been completely forgotten.

Traditional grammar as developed between the end of 18 th century and the beginning of 20 th century has made use not only of parts of speech but also of other criteria - so-called functional criteria - in classification. Such functional criteria are the notions of subject, (direct or indirect) object, adverbial phrases (time, place, manner, etc.), main clauses subordinate clauses, etc. These notions, which are neither independent of the parts of speech nor independent of each other, have been utilized in grammatical descriptions and still are, but the conditions under which they can be used have not been made explicit by their users. This circumstance is, no doubt, one of the main sources of the various obscuri-

ties and contradictions found in traditional grammatical studies.

Some concepts of traditional grammar may suffice to illustrate the nature of the criteria used in the description of sentences. Consider, for example, the traditional notion of indirect object. Already, a rather rudimentary analysis reveals that there is a rather complex dependency between this notion and other grammatical notions. The notion of object is primarily a semantic notion. At the same time, however, it also carries syntactic information, since one always speaks of an object complement (direct or indirect) but never of a subject complement. The notions of subject and complement are syntactic notions that exclude each other. It then appears that an object must always be a complement. This observation has far-reaching consequences. Thus, in the sentence

'Max likes this hat'
Max aime ce chapeau

Max must be considered the subject, and chapeau the object of the sentence. In the synonymous sentence, however

'This hat pleases Max'
Ce chapeau plait à Max

the subject of the sentence is chapeau and the object of the sentence is Max. At the same time, it is evident that chapeau and Max play the same semantic role in the two sentences. Consequently, the notions of subject and object which are fundamentally semantic cannot be used to compare and to connect these sentences. The definitions of these notions do not enable us to state this relationship in a simple way, because, the notion of direct object is defined as the complement of a transitive verb and the notion of transitive verb is based on the notion of direct object (the two definitions need not be circular, however). Moreover, in these definitions, mostly in the definition of the notion of verb, one has often recourse to the notion of action and one would say that the object is exposed to this action. But then, the terms of "action" and that of "to be exposed to" cannot be defined: they are not technical terms as are object or transitive verb, but rather words of ordinary language taken in their everyday sense. It is well-known that these categories raise a number of serious problems, and that they

have been at the center of discussion since Aristotle, we also know that they are not satisfactory. The fact that one takes into consideration verbs of states in addition to verbs of action does not solve the difficulties. Nor does the distinction between affected and effected object (object of result). All this means nothing more than refinement based on new semantic relations between verb and object. At the same time however, with these notions, further undesirable extralinguistic factors have been introduced into the definitions. Notice furthermore that the notion of indirect (object) is also hard to grasp in spite of the fact that it is a notion of form and not one of meaning. In French, it is tied to the presence of a preposition. But there is no operational definition of the notion of preposition. Even if there is a general agreement on the prepositional status of a small number of words such as à, de, pour, sur, etc., there are quite a few prepositional constructions that can easily be analyzed in terms of other parts of speech (e.g. afin de in order to, à l'instar de, étant donné que -given that, vu que-considering that) and which show that there is a continuum between the category of preposition and the categories of noun phrase and verb phrase. In all cases where form is concerned, it is always possible to give precise definitions which may sometimes be arbitrary as, for example, in the form of word lists which eliminate all reference to non reproducible intuitions.

The notion of indirect object is symptomatic for the difficulties which one encounters if one wants to apply the concepts of traditional grammar to the systematic description of a language. Though the examples that are given by the majority of authors to illustrate the application of their definitions look rather straightforward, one has to realize very soon that this application cannot be reproduced in a coherent way on a larger scale in (elementary or secondary) teaching or in scientific research.

Structural grammar has contributed to the clarification and purification of many of our linguistic concepts. In particular, semantic factors occur no longer in the definitions of grammatical terms, they are more or less restricted to the purely combinatorial properties of words. This has been a considerable achievement: it is now possible to describe all sorts of linguistic phenomena in a coherent and uniform way. The methods used by structuralists were based on the formalism

developed in phonology and which was considered to be both necessary and sufficient for linguistic description. However, it turned out that this formalism could not be extended to the description of syntax in a straightforward way. Indeed, this formalism is too simple since it is based solely on the operations of concatenation of words or phrases complemented by conditions with respect to the possibilities of transition from one word to another, or from one phrase to another. The operation of concatenation assembles words into syntactically relevant units. These are the constituents or phrases. Thus, for example, an article, a noun and an adjective are concatenated to form a noun phrase. Or, a noun phrase and a verb phrase are concatenated to form a sentence.^{10/} In this way, structural analysis enables us to determine a hierarchical organization among the words and phrases that constitute the sentence. The basic unit of this organization is the word (or the morpheme, etc.). It has thus become possible to construct a typology of sentences, which was not possible - at least not in a univocal way - in traditional grammar. A simple example illustrates the structuralist procedure. Consider

'Max sees the little boy'
 Max voit le petit garçon
 'The girl has seen the boy'
 La fille a vu le garçon

The analysis in terms of parts of speech categories yields the following two sequences:

N proper V Art Adj N
 Art N Aux Past participle Art N

Since these sequences are different, one has to classify the two sentences as being different. If, however, one introduces a hierarchical organization and, in particular, the notion of noun phrase NP, then equivalence relations such as

NP = N_{proper}
 NP = Art N
 NP = Art Adj N

will enable us to modify the above representations which become:

NP V NP
 NP Aux Past participle NP

These sequences are already closer to each other. If we now proceed in the same vein and add to our grammar the relation of equivalence

V = Aux Past participle

we have identical structures for our two sentences, which is no doubt the result desired by the grammarian.

One can thus interpret the whole activity of structural analysis as an enterprise which aims at the construction of a classification of sentences. To be sure, grammarians do not seem to have looked at their activity from this point of view. It is, however, necessary to regard the grammarian's activity in this way, if we want to follow Chomsky in his criticism.

It is important to note that the example discussed above may be extended in two ways, which may appear to be quite distinct. On the one hand, we have the possibility of defining the groupings of several constituents into one constituent, i.e. the procedures which enable us to analyze a given sentence in terms of simpler elements. This approach has been the core of theoretical structuralism. On the other hand, one could also make use of constituent structure in order to investigate the connections between sentences. The sentences could then be taken either from a corpus, or they could be systematically constructed out of lexical elements. The result obtained would be a classification of sentences. So far, no linguist had undertaken such an enterprise.

Constituent analysis has first been criticized by Harris and later on by Chomsky. The main point of this criticism is that this form of analysis does not allow for the description of certain grammatical relations. As I have already mentioned, it is not possible for example to describe an active sentence and its corresponding passive as being equivalent, except in a rather uninteresting way, namely by stating that both sentences can be decomposed into a noun phrase and a verb phrase. The description of this relation lacks for a purely formal reason. By definition, constituent analysis cannot account for changes of word order; consequently for the permutation of subject and object which takes place in active-passive pairs. Transformational grammar is an outgrowth of this criticism: it introduces - in addition to the rules of constituent structure -

the more powerful rules of permutation and deletion.

In the course of his critical discussions of constituent analysis, Chomsky calls this type of description taxonomical. It is not quite clear why Chomsky applied this term here. No structural linguist ever considered the construction of a classification of the sentences of a given language as part of his program. The only program that has been made relatively explicit and which has also been qualified by Chomsky as being taxonomic is that by Harris.^{11/} In spite of the fact that Harris mentions word lists, it is not possible to consider his program as a taxonomy of sentences or of sentence types of a given language. In fact, there is nothing in common between the organization of the combinatorial constraints in phonology and syntax as propounded by Harris, and the great taxonomies constructed up to the present, e.g., the taxonomies in biology and botanics. What Harris proposes is in fact a formal grammar of the type studied by Chomsky later on. This grammar consists of various types of rules and the difference between this grammar and generative grammars - as Chomsky himself admits - is merely terminological.^{12/} Consequently, the position taken by Chomsky becomes quite clear: what he criticizes has nothing to do with taxonomy, what he has shown in a brilliant manner is the inability of constituent grammar to connect with each other forms clearly related. His discussion has no relation whatsoever to a possible taxonomic attitude in linguistics, it relates to this attitude only under the implausible interpretation that it was also the attitude of the structuralists. A justified criticism would be to point out that there are no satisfactory concepts for a given systematic classification, but such a classification had not been envisaged by anybody so far. Therefore, it is not at all clear why the term 'taxonomy' has been brought into the discussion.

No doubt, it is a consequence of this confusion (brought about by the inappropriate use of the term 'taxonomy') that taxonomic linguistics has been equated with structural linguistics. Following Chomsky, quite a few scholars have then attributed to taxonomy all the shortcomings of structural grammar. (This portion of recent history of linguistics gives us an interesting example of the birth of a dogma, a dogma whose remarkable success has discredited a number of interesting studies, and at the same time assured the ephemeral success of a number of essays which lacked any general value).

Nevertheless, the taxonomic approach, i.e. the construction of classifications, is a possible approach and it is a priori completely independent of the nature of the descriptive concepts used. It is only at the level where classificatory criteria are worked out - they may be either distributional or transformational - that criticisms might be justified, but again there is no reason to criticize a priori the taxonomic approach, i.e. the construction of any classification. It would seem that such a criticism could only question the adequacy of an already existing classification by showing that it has certain shortcomings: either because the objects to be classified have inadequately been chosen or because the choice of criteria has been inadequate. The example of the passive may be symptomatic in this sense, since traditional grammar has already shown that there are thousands of elements which undergo this operation. One can therefore be sure that any classification of sentences which does not take into account the passive must be defective.

My analysis of the situation is thus this: the negative attitude taken by a great number of linguists as a consequence of Chomsky's observations is due to a terminological confusion Chomsky^{13/} wrote: "...let us coin the term taxonomic phonemics' ...". But the verb to coin applied to a word or to an expression implies that one invents this word or this expression. In other words, the term is arbitrary and there is no reason to connect it to the ordinary sense of the word taxonomic. That is, our discussion may be pointless. The present state of linguistic research, however, does not allow this interpretation. It is a fact that linguists no longer construct inventories and when they tackle a new phenomenon or when they formulate a rule, they never perform any study of the distribution of the phenomenon (or of the rule) with respect to the lexicon, even restricted. Such studies would amount to establishing classifications, since in general observations vary according to lexical elements. The existing situation, true enough, may have more ideological than scientific reasons. There is no doubt whatsoever that the construction of theories or models - even in the most oversimplified manner - has a much higher prestige in the eyes of most linguists than the construction of classifications, i.e. than the making of lists which require much effort and subtle work. It is certainly the predictive character of theories which explains this difference of prestige. Unfortunately, it is not generally known

that classifications may be equally predictive as theories. (From this point of view, it is quite clear that man's anxiety in front of his future is the cause of his being fascinated in an equal way by the scholar who predicts phenomena and by the fortune-teller who predicts joys and catastrophes in the family).

All these remarks belong, however, to the sociology of the linguists' world and they should not prevent a specialist from classifying linguistic objects. But there may be more serious reasons, which are not quite novel, i.e. which are less exposed to the changes of fashion, and which may explain why no large scale syntactic classification has been carried out so far. Whereas a wide variety of morphological inventories have been compiled (cf. the various types of dictionaries), there are no dictionaries of sentences or of syntactic types which would be complete in the sense as dictionaries of words are complete. The reason for this may very well lie in the impression that syntactic forms exhibit a richness that goes much beyond the richness encountered in morphophonology. At least two reasons contribute to this qualitative difference of magnitude. On the one hand, the number of basic elements in phonology is limited (a few dozens of phonemes), but the number of basic elements in syntax is very large (of the order of several ten thousands words). On the other hand, whereas there seems to be an upper limit on the length of words in Indo-European (some dozens of phonemes), there is no upper limit on the length of sentences. This difference in cardinality of the set of words and the set of sentences is certainly a serious difficulty. Notice, incidentally, that this difference can be assessed in more precise terms.^{14/} However, we have seen that the use of parts of speech enables us to restrict our attention to the study of syntactic forms; this move diminishes considerably the effect of the combinatorial explosion hinted at above. On the level of syntactic forms, an enterprise aiming at the enumeration of sentence types is no longer unrealistic. Up to now, however, no linguist has ventured such an enterprise. Perhaps, this stems from the high complexity shown by the classificatory criteria of traditional grammar. Above, we have discussed these criteria. I think that the remarkable development of the descriptive notions in the formal syntax of distributional grammar enables us now to carry out a systematic classification of syntactic forms.

The simplification of descriptive notions is almost en-

tirely the merit of Harris. It has made possible for Chomsky to express the most profound philosophical point of view ever uttered on language: between the mentalistic and purely empirical approaches which both have led to an impasse, Chomsky was able to define a delicate equilibrium which goes much beyond the rather limited set of facts on which Chomsky based his observations and permits us to grasp the fundamental aspects of the relationship between language and mind, and to establish an admirable synthesis by means of formal procedures. It is equally clear that this simplicity of grammatical concepts also permits us to establish a classification, a classification whose existence is a fundamental question, since it is one of the rare efficient means by which the enormous range of theoretical possibilities can be restricted. At present, scholars are engaged in the elaboration of linguistic theories aiming at a global view of language. These theories will most likely not suffer as much as the preceding theories did, from the lack of empirical backing.

Notes

1. The only notable exception is that of Z.S. Harris (Structural linguistics, Chicago: University of Chicago Press, 1951) who looked upon the problem of syntax very early in this way, and who worked out for it an algebraic translation.
2. The important work by Paul Postal (On Raising, Cambridge, Mass. M.I.T. Press, 1974) provides an overall view of the discussions on Raising.
3. For many verbs the construction derived from the sentential complement is not attested in a direct manner:

Je dis que cet homme est capable de tout
 'I say that this man is capable of (doing)
 anything'
 Je dis cet homme être capable de tout

This last form is however indispensable in order to account for the forms accepted as in

C'est cet homme que je dis être capable
 de tout
 L'homme que je dis être capable de tout...

4. Since the definition of the parts of speech is essentially distributional, we have in the case of peu an example which is analogous to that of the passive in the sense that it shows the advantages of a transformational solution. Such a solution would reduce the three different uses to a single one.

5. A. Elia, in Rapports de recherche du L.A.D.L. n°7, 1977; Lingvisticae Investigationes 1978-2.

6. Syntactic Structures, The Hague: Mouton, 1957.

7. Histoire de la syntaxe, Genève. Droz, 1968.

8. Essentially, the combinatorial methods, whether structural or transformational.

9. In fact, semantic and syntactic criteria have not developed much as time went by. To be sure, all of them might appear to be adequate on intuitive grounds but only empirical verification makes it possible to assess their real pertinence.

10. There exists another method which is equivalent to the one just mentioned and which consists in analyzing or decomposing sentences or other constituents into smaller constituents.

11. Structural Linguistics, appendix to §20.3.

12. Transformational Approach to Syntax, A.A.Hill, ed., Proceedings of the Third Texas Conference on Problems of Linguistic Analysis in English, Austin, The University of Texas Press, 1962, pp. 124-158. Also in J.Katz and J.Fodor eds., The Structure of Language, Englewood Cliffs: Prentice-Hall, 1964.

13. Current Issues In Linguistic Theory. In The Structure of Language, op.cit., p.91.

14. Cf. M.Gross, Méthodes en syntaxe, Paris, Herman, 1975. I have shown, for example, that the cardinality of the number of sentences containing 20 words is of the order of 10^{50} in French.

Sujet						Auxiliaire avoir	Auxiliaire être	N ₀ V	Compléments indirects										Extraposition										N ₀ V de ce que P																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																					
N _{hum}	Ou P	le fait Ou P	V ¹ Ω	V ² Ω					N _{hum}	ppv	N _{hum}	le fait Ou P	à N _q	dans N _q	de N _q	ppv	que P	que Psubj	de V ¹ Ω	de V ^{1c} Ω	Infinitives																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																													
																					T _c = "passé"	T _c = "futur"	V _c = devoir	V _c = pouvoir	V _c = savoir	N ₀ V de ce que P																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																								
-	+	+	-	-	paraître	+	-	+	Loc	-	-	+	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-