

6. Prestructuralist and Structuralist Approaches to Syntax

1. Prestructuralist syntax
2. The poverty of structuralist syntax in Europe
3. Structuralist syntax in America
4. Conclusion
5. References (selected)

Abstract

The chapter describes in rough outline the development of the notion of syntax since its earliest beginnings in Classical Antiquity, starting with the late hellenistic Alexandrian linguist Apollonius Dyscolus, for whom syntax was mainly the study of the functions of the nominal cases in sentences. Since the modern European languages largely do without nominal case, the notion of syntax became nebulous, with the result that up to ca 1950 it was widely thought that syntax was not part of the language system but of speech, the creative use of language. This changed with American structuralism, especially with Leonard Bloomfield, who proposed a system of Immediate Constituent (IC-) Analysis, derived from Wilhelm Wundt, as the theoretical basis of syntax. This led to the question of the motivation of specific IC-analyses, which was answered, in American linguistics, mainly in two ways. One answer, given by the so-called God's truth school, was that one could somehow observe IC-structures by introspection. Another, given by the so-called hocus-pocus school, was that the correct assignment of IC-structures will result from the simplest and most compact description of the language as a whole in terms of IC-structures. This latter answer was better integrated into current philosophy of science and carried the day. It led directly to the development of Generative Grammar during the 1950s and 1960s. This whole development is treated in the broader terms of what is known as structuralism, that is, the general trend, starting in the late nineteenth century, to see mental, especially cognitive, processes as resulting from rule-governed formal rules and principles, much in the way physical nature is the result of mechanical laws: the mechanisation view of the physical world was extended to the realm of cognition. Furthermore, the notion of constituent (tree) structure, both as dependency trees (Tesnière) and as IC-trees (Wundt, Bloomfield), is considered from the point of view of formal algorithmic calculus as applied in arithmetic. In this sense, the introduction of constituent structure into the study of syntax has contributed to the formation of the notion of a formal, compositional semantics for sentences.

1. Prestructuralist syntax

The term *syntax*, in the sense of the theory of how to combine words into sentences goes back to the ancient Greek Stoic philosopher Chrysippus (third century BCE), but

it did not gain currency in language studies for another five centuries, till after the publication of *Peri Suintáxeōs* [On Syntax] (see Householder's translation with commentary of 1981) by the Greek linguist Apollonius Dyscolus, who lived and worked in Egyptian Alexandria during the second century CE. The Greek noun *súntaxis* means literally 'the act of placing together' and thus occurs in different senses, such as 'composition', 'arrangement', 'organisation', 'covenant', 'treatise' and even 'state constitution' (in Modern Greek it means 'pension, retirement pay').

Apollonius's book on syntax deals with a variety of topics, in particular the use and function of the definite article and of relative and anaphoric pronouns, verbal argument structure (in particular what verbs assign what cases to what nominal arguments with what meanings), participles, prepositional constructions and adverbial adjuncts. It does not discuss but merely assumes implicitly that regular sentences consist of a subject term and a Verb Phrase, the latter consisting of a finite verb form possibly accompanied by nominal argument terms and/or prepositional or adverbial constituents. It was in this form that syntax was taught at schools and universities through the ages till well into the twentieth century, especially in connection with the Classical languages, but also with regard to those modern European languages that possess a well-developed nominal case system, such as German or Russian.

A difficulty arose with regard to the modern West European languages, especially English, Dutch, French, Italian, Spanish and Portuguese, which largely lack a nominal case system. For those languages, the notion of syntax became attenuated to the point of becoming more or less extinct. Thus, by the end of the nineteenth century, grammar books and the teachers using them, at least in western Europe, were at a loss as regards the notion of syntax. The difficulty was not felt to the same degree with respect to the Central and East European languages, with their sometimes elaborate case systems, where one simply continued in the old Apollonian tradition. The predicament that existed for the West European languages was resolved during the second half of the twentieth century, especially in the context of transformational-generative grammar, when linguists became aware of the real complexities of syntactic structures and processes, especially, but far from exclusively, in the area of sentential, clausal and participial complementation. During the intervening period, say, from 1900 till 1960, linguists, with only a few notable exceptions, tended to be at a loss as to what is involved in the study of syntax.

2. The poverty of structuralist syntax in Europe

2.1. Mainstream European structuralism

What we now call structuralism was, in fact, a very general notion regarding the workings of the human mind, which came up in the human sciences during the second half of the nineteenth century. It is probably no oversimplification to say that structuralism is best defined as the attempt to apply the mechanical notion of machine to the human mind (see Seuren 1998: 141–144). Since the seventeenth century it had been commonly accepted that the physical world as a whole, and each individual living body in particular, could be seen as a system of interrelated mechanisms. At first, it was the clock work mechanisms developed during the seventeenth and eighteenth centuries that served as

the prototypical example of what was meant by mechanism, but during the nineteenth century the notion was primarily exemplified by the various kinds of transportation and production machines that came into general use in the industrialising countries of Europe.

It was thus the machine metaphor in the study of the human mind that came to define the notion of structuralism in a general sense. In Europe, the most articulate spokesman of this new philosophy was the Parisian critic, philosopher and historian Hippolyte Taine (1828–1893) who had an enormous influence on French cultural (especially literary) life during his own day and after. His influence on linguistics is not generally recognised, but, according to Aarsleff (1982: 356–371), Ferdinand de Saussure (1857–1913) derived his basic linguistic notions, such as the distinction between *langue* and *parole*, and that between diachrony and synchrony, and also his notions of *sign* and *valeur*, from Taine, with whose works he became familiar during his years in Paris previous to his appointment to a Geneva chair in 1891. Taine's thoughts on the nature of language and language use are found in his book *De l'intelligence*, first published in 1870 and reprinted many times until the beginning of the first world war in 1914. Though widely read during the *fin de siècle* in France, this book is now largely forgotten, unlike Saussure's posthumous *Cours de linguistique générale* (1916), which has been highly influential in European linguistic and literary structuralism from roughly 1935 onwards.

Other early proponents of structuralism in linguistics were the Poles Jan Baudouin de Courtenay (1845–1929) and Mikołaj Kruszewski (1851–1887), both of whom taught at Kazan in Russia. They were not inspired by Taine's very general and rather high-flown ideas, as Saussure was, but concentrated more directly on the facts of language, in particular the interpretation of speech sounds in terms of a type-level cognitive system of phonological units, which Kruszewski called phonemes. As a result, it was phonology that dominated European structuralist linguistics from the beginning until roughly 1960.

Apart from the focus on phonology, linguistic structuralism was characterised in an overall sense by an insistence on the autonomous nature of language and linguistics. Language began to be seen, though not yet in explicit terms, as an autonomous machine in the mind or brain, connected with other such machines but separated from them in that the language machine has its own internal structure and its own functional principles. Language thus began to be regarded as a kind of module in the mind, in the sense later made explicit in Fodor (1983) and other publications. Ideas about the nature or internal structure of the language module were, on the whole, rather restricted. Both in Europe and in America, the language module (Saussure's *langue*) was, for a long time, seen mainly as a collection of lexical items, idioms and a few patterns for the combination of smaller elements into larger wholes. The two continents differed in that in Europe the language module, or *langue*, was treated as an element in the conscious mind, accessible through introspection, whereby the notion of mind remained ill-defined. Across the Atlantic, by contrast, *behaviourism* reigned supreme until the early 1960s. There the language module was considered to be materially given in the brain as a set of cerebral connections established through associations of stimuli and responses having occurred in the personal history of each individual. The question of conscious access to the language system was considered unscientific and thus left out of consideration. Nowadays, of course, the dominant view is that the language system or module, though materially realised in the brain, is defined as a complex piece of brain software, inaccessible to

consciousness or introspection beyond the choice of propositional content, of lexical material and of values on sociolinguistic or interactional parameters.

Remarkably, hardly any structuralist theory of syntax was developed in Europe, even though the need for such a theory was widely felt and many attempts were made at getting one off the ground. The two main problems were (a) the unclarity as regards the notion of syntactic structure and (b) the fact that no-one had a clear idea of the complexities of syntax and of the rule systems underlying syntactic constructions. In America, empirical access to syntax was achieved by the hypothesis that syntactic structure is definable as a hierarchy of immediate constituents, as will be explained in a moment. It was through this hypothesis that syntax began to open itself to linguists at large. In Europe, however, one had no grip on the problem of syntactic form, nor on the data constituting it. Notable exceptions were Bech (1955) for German (cf. Kiss, this volume) and Paardekooper (1955) for Dutch. These classic authors had a sharp eye for the baffling complexities of the syntactic structures in the languages concerned. Although they were skeptical as regards linguistic theory and even averse to it, they made vast numbers of the most acute observations and sometimes established regularities that later proved to have great explanatory power. Yet it took a long time before they were acknowledged and appreciated for what they were in the community of theoretical linguists.

In mainstream theoretical linguistics, the notion of syntax remained badly underdeveloped. Saussure, for example, took it that (1916: 30–31):

- 1) (...) parole is an individual act of free will and intelligence, in which we distinguish:
- 2) the combinations by means of which the speaking subject makes use of the language code in order to express his personal thought;
- 3) the psycho-physical mechanism enabling him to externalise these combinations.

What he meant by language code is explained in (1916: 172–173). Having said that *parole* is characterised by the freedom of combining words according to the thought to be expressed, he denies that all of syntax is, therefore, part of *parole*, because (a) there are fixed or idiomatic locutions, which are part of the lexicon and thus of *langue*, and (b) there are rules for combining morphemes and words into larger wholes. When these larger wholes are words, the corresponding morphological rules are unequivocally part of the language system (*langue*), but when they are word groups or phrases, then the dividing line between *langue* and *parole* is unclear. In short, Saussure's notion of the language system is strictly word-based, as a result of which he was unable to develop any serious notion of syntax. Interestingly, he was criticised for this by the American linguist Leonard Bloomfield in his review of the *Cours* (Bloomfield 1924).

Alan H. Gardiner (1879–1963) likewise failed to reserve a proper place for syntax in linguistic theory. His main publication on linguistic theory is Gardiner (1932), on the distinction between the language system (*langue*) and speech or language use (*parole*) (for discussion, see Seuren 1998: 171–177). As regards the nature and position of syntax, Gardiner had not made up his mind. On the one hand he maintained that the word is the unit of the language system, while the sentence is the unit of speech: “The smallest section or unit of speech is the sentence” (Gardiner 1932: 208). This naturally implies that syntax, or the art of combining words into well-formed sentences, does not belong to the language system but is a matter of free creative use of words. Yet he does not appear to have been quite happy with this position. In Gardiner (1932: 184) we read: “Thus there is such a thing as ‘sentence form’, and like all other linguistic forms, it is a

fact of language, not a fact of speech.” Gardiner failed to resolve this contradiction or ambiguity. But it is clear that, whatever place he wished to reserve for it, he saw syntax as a fairly trivial list of sentence patterns. (For a clear survey of the problem of how to define a sentence and hence how to define syntax, see Bühler 1934: 356–366.)

2.2. Dependency Grammar

The only form of theory-based structuralist syntax developed in Europe is *Dependency Grammar*, also called *Valency Grammar* (cf. Klotz, this volume), developed by the French linguist Lucien Tesnière (1893–1954), whose actual theory was not published until well after his death in Tesnière (1959). Tesnière distinguished between the main functor of a sentence, the verb, and its arguments (thus deviating from the classic Aristotelian Subject-Predicate division and following the modern trend of Predicate-Argument structure). He developed a method of structural analysis known as *dependency tree*, which allows one to compute the value of a function.

This is most easily demonstrated with the help of a couple of examples from simple arithmetic. In fig. 6.1.a, one sees the computation of the function $(5 \times 6) + 8$. This is an addition with the two arguments 5×6 and 8. Since the addition function requires two or more actual numbers as arguments, the number corresponding to (or the value of) 5×6 must be computed first. Therefore, one must start with computing the value of the multiplication function with the arguments 5 and 6. This value is 30. Then the value of the addition function with the arguments 30 and 8 is computed, resulting in the final value 38. In fig. 6.1.b, the highest function is the multiplication function \times with the arguments 5 and $(6 + 8)$. To compute this function, one first has to compute the value of $6 + 8$, that is, 14. Now the multiplication function with the arguments 5 and 14 can be set to work. The final resulting value is thus 70.

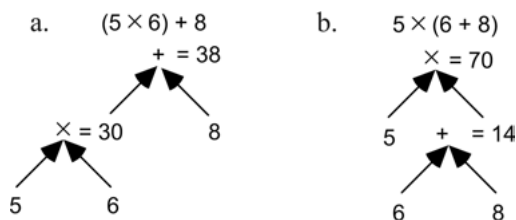


Fig. 6.1: $(5 \times 6) + 8$ and $5 \times (6 + 8)$ in terms of dependency trees

Tesnière applied this to sentence structure in the following way. A sentence like *The children ate sweets* is analysed as consisting of the functor word *ate* and the two argument terms *the children* and *sweets*, respectively. One may even subanalyse the phrase *the children* into the functor *the* with the single argument *children*, resulting in the referring Noun Phrase *the children*. The diagram is shown in fig. 6.2.

One may go further and say that the past tense *ate* is the result of a functor PAST with the single argument *the children eat sweets*, as in fig. 6.3.

One difference with arithmetic is that values are not automatically of the same category. In arithmetic, the arguments of an arithmetical function are numbers and every

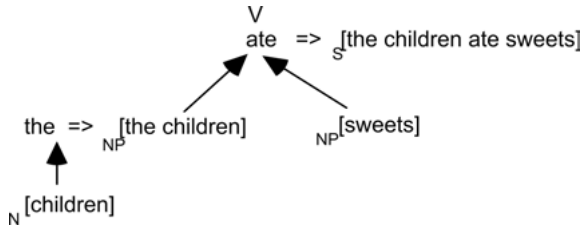
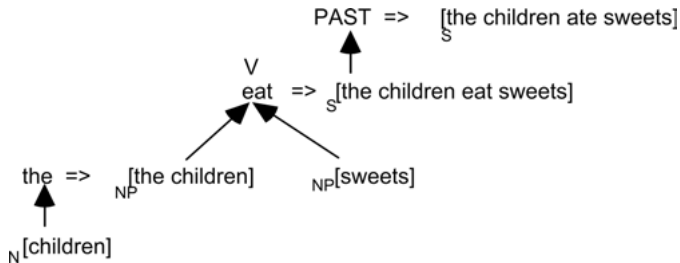
Fig. 6.2: Dependency analysis applied to *The children ate sweets*

Fig. 6.3: The same as fig. 6.2, with Past Tense added

value of a function is again a number. Not so in syntax, where each value must be assigned a category and functions are thus defined for arguments belonging to specific categories. Thus, the value category of the definite article *the* can be given as N(oun) P(hrase), and the value category of the verbal function *eat* as S(entence), just like the value category of the function PAST.

A difficulty lies in the morphosyntactic form of the resulting values. For example, to get *the children eat sweets* from the arguments *the children* and *sweets* as input to the function *eat*, specific rules of syntactic arrangement are needed, which are not given by the dependency tree.

In fact, dependency trees smack more of semantics than of syntax, in that the value category of each functor stands not so much for a well-defined set of syntactic properties as for a semantic value in a model, whereby the highest value category S can be interpreted as a truth value in a model. While this can be considered to have the advantage of integrating syntactic and semantic description, it must at the same time be admitted that the formal syntactic part of the equation is not entirely transparent. Nor do dependency trees easily lend themselves to the kind of tree-transformational manipulations that are a central element in Transformational-Generative Grammar. Even so, Dependency Grammar became popular in central and eastern Europe, especially in the Prague School of linguistics and also in the Moscow school of Meaning-Text Theory (Zholkovsky and Mel'čuk 1965; Mel'čuk and Pertsov 1987).

3. Structuralist syntax in America

3.1. Constituent structure in the early Bloomfield and Sapir

A different picture arises on the other side of the Atlantic Ocean. There we see a decisive breakthrough to structuralist syntax, owing to an insistence on the notion that not the word but the sentence is the main unit of the language system. The first name to be mentioned is that of Edward Sapir (1884–1939). As he was trained in anthropological linguistics, his general ideas about language were most directly inspired by his knowledge of American Indian languages.

His most important publication in this regard is (Sapir 1921). This book is a curious mixture of traditional European thought on language and an emergent, more typically American, positivist attitude. He stresses the fact that (1921: 165) “Language exists only in so far as it is actually used – spoken and heard, written and read.”. The underlying language system is actually no more than a hypothesis naturally arising in the mind of anyone who starts reflecting on language. Yet this positivism is thrown to the wind when he speaks about the sentence (Sapir 1921: 33):

Radical (or grammatical) element and sentence – these are the primary *functional* units of speech, the former as an abstracted minimum, the latter as the esthetically satisfying embodiment of a unified thought. (...) The sentence is the logical counterpart of the complete thought only if it be felt as made up of the radical and grammatical elements that lurk in the recesses of its words.

Historically, it is interesting to see how Sapir is struggling with the concept of sentence, and how the notion of formal syntax is beginning to shape up, though still loaded with meaningless metaphors like “lurk in the recesses of words”. He refuses to accept that the notion of sentence still eludes him (Sapir 1921: 36):

We have already seen that the major functional unit of speech, the sentence, has, like the word, a psychological as well as a merely logical or abstracted existence. Its definition is not difficult. It is the linguistic expression of a proposition. It combines a subject of discourse with a statement in regard to this subject. Subject and ‘predicate’ may be combined in a single word, as in Latin *dico*; each may be expressed independently, as in the English equivalent, *I say*; each or either may be so qualified as to lead to complex propositions of many sorts.

While, ironically, he claims that it is not hard to define the notion of sentence, we see him fall back on purely psychological, nonpositivistic, criteria for the delimitation of this notion. And these psychological criteria are directly derived from the Aristotelian tradition, where indeed a proposition is the mental assignment of a property to an underlying, discourse-given entity. This notion was revived in the late nineteenth century by European authors like Wegener (1885), Lipps (1893) and Stout (1896), on whom Sapir relied and who distinguished a topic-comment structure in sentences-in-discourse. Yet the authors just mentioned made a clear distinction between a sentence’s topic-comment structure and its overt syntactic predicate-argument structure, which in most cases does not reflect the topic-comment structure, the latter being generally manifest through an

intonational overlay. This distinction between strictly syntactic and discourse-semantic (or information-structural) sentence structure was not made by Sapir, who conflated the two in an attempt to come to terms with the notion of sentence (Sapir 1921: 125–126):

It is well to remember that speech consists of a series of propositions. There must be something to talk about and something must be said about this subject of discourse once it is selected. This distinction is of such fundamental importance that the vast majority of languages have emphasized it by creating some sort of formal barrier between the two terms of the proposition. The subject of discourse is a noun. As the most common subject of discourse is either a person or a thing, the noun clusters about concrete concepts of that order. As the thing predicated of a subject is generally an activity in the widest sense of the word, a passage from one moment of existence to another, the form which has been set aside for the business of predicating, in other words, the verb, clusters about concepts of activity. No language wholly fails to distinguish noun and verb, though in particular cases the nature of the distinction may be an elusive one. It is different with the other parts of speech. Not one of them is imperatively required for the life of language.

Interestingly, this same conflation of topic-comment and syntactic structure is found in as late a publication as Hockett (1958: 201), where one reads:

The most general characterization of predicative constructions is suggested by the terms *topic* and *comment* for their ICs [i.e. immediate constituents]: the speaker announces a topic and then says something about it. Thus *John | ran away; That new book by Thomas Guernsey | I haven't read yet*. In English and the familiar languages of Europe, topics are usually also subjects, and comments are predicates: so in *John | ran away*. But this identification fails sometimes in colloquial English, regularly in certain special situations in formal English, and more generally in some non-European languages.

Sapir's immediate example may well have been Bloomfield's (1914) book *Language*, where we read (Bloomfield 1914: 61):

In the primary division of an experience into two parts, the one focused is called the *subject* and the one left for later attention the *predicate*; the relation between them is called *predication*. If, after this first division, either subject or predicate or both receive further analysis, the elements in each case first singled out are again called subjects and the elements in relation to them, *attributes*. The subject is always the present thing, the known thing, or the concrete thing, the predicate or attribute, its quality, action, or relation or the thing to which it is like. Thus in the sentence *Lean horses run fast* the subject is *lean horses* and the horses' action, *run fast*, is the predicate. Within the subject there is the further analysis into a subject *horses* and its attribute *lean*, expressing the horses' quality. In the predicate *fast* is an attribute of the subject *run*.

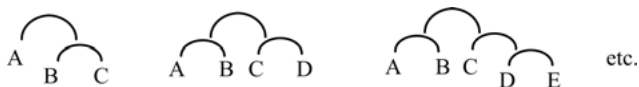
The importance of this text lies not so much in its conceptually confused notion of sentence structure as in the fact that a linguistic structure is seen as consisting of a number of elements, later called *constituents*. In the early years of the century it was not entirely clear what this could mean, but by the 1930s some linguists were discovering the great potential of this way of looking at linguistic structures. Nowadays we are accustomed to the idea that a construction is a combination of constituents, each of which belongs to one or more given classes and which jointly form a constituent that

again belongs to a given class. To us, a century after Bloomfield's (1914) book, it is a trivial insight that the highest possible grammatically relevant class is the class denoted by the symbol *Sentence* (S), but, historically speaking, it took some time for that insight to break through and become explicit.

It is not often realised in linguistic circles that Bloomfield took his notion of hierarchical constituent structure, each layer consisting of one or more immediate constituents (ICs) until the last layer which consists of ultimate constituents, from the German philosopher-psychologist Wilhelm Wundt (1832–1920), who actually drew IC-diagrams (see Percival 1976; Seuren 1998: 220–221). Wundt wrote (1880: 53–54; translation mine):

The simplest form of a *thought*, i.e. a self-contained apperceptive representational process, occurs when a total representation ('Gesamtvorstellung') falls into *two* parts that are connected with each other. This happens in the *simple judgement*. If we use the sign \wedge for apperceptive connections of successive representations, then $A \wedge B$ is the psychological symbol of the simple judgement.

As soon as the total representation, the splitting up of which results in a thought process, is separated into three or more single representations the judgement is no longer simple but *composite*. In a composite judgement the connection of the single parts is never uniform, in the sense that the form $A \wedge B$ would extend over a larger number of members, as in $A \wedge B \wedge C$ (...). On the contrary, these apperceptive connections always proceed in such a way that first, as with the simple thought, the total representation is separated into two single representations, upon which either or both of these can be subdivided into two further single representations, and so on. Herein lies the essential difference between apperceptive and associative connections. If we use the sign $-$ for the associative connection of successive representations, we see that an associative sequence $A-B-C-D$ (...) can contain any number of members. In contrast to this, the apperceptive thought process always proceeds in forms like the following:



This principle of duality or of binary connection has found its unmistakable expression in the categories of grammatical syntax. For all these categories always reduce to just *two* representations which are connected with each other. Thus we distinguish first the two main representations Subject and Predicate, which correspond with the first division of the thought. The Subject may be divided again into Noun and Attribute. The Predicate, when it is nominal, splits into the Copula and the Predicate proper, upon which the latter, like the Subject, may split into Noun and Attribute again. But if the Predicate is verbal it may split into Verb and Object, or into the Predicate proper and the supplementary Predicate.

Bloomfield (1914) is largely based on Wundt's work, and there can be no doubt that the passage just quoted from Bloomfield (1914: 61) was directly inspired by Wundt's notion of IC-analysis. Yet Bloomfield never actually drew a diagram, not even in his *magnum opus* (1933). Nor did Sapir in his (1921) book, where an elaborate IC-analysis is given of the one-word Paiute Noun Phrase corresponding to 'they who are going to sit and cut up with a knife a black buffalo'. Sapir describes this structure entirely in English prose (Sapir 1921: 31–32):

One example will do for thousands, one complex type for hundreds of possible types. I select it from Paiute, the language of the Indians of the arid plateaus of southwestern Utah. The word *wii-to-kuchum-punku-rügani-yugwi-va-ntü-m(ü)* is of unusual length even for its own language, but it is no psychological monster for all that. It means ‘they who are going to sit and cut up with a knife a black cow (or bull)’, or, in the order of the Indian elements, ‘knife-black-buffalo-pet-cut up-sit(plur.)-future-participle-animate-plur’. The formula for this word, in accordance with our symbolism, would be (F) + (E) + C + d + A + B + (g) + (h) + (i) + (0). It is the plural of the future participle of a compound verb ‘to sit and cut up’ – A + B. The elements (g) – which denotes futurity – (h) – a participial suffix – and (i) – indicating the animate plural – are grammatical elements which convey nothing when detached. The formula (0) is intended to imply that the finished word conveys, in addition to what is definitely expressed, a further relational idea, that of subjectivity; in other words, the form can only be used as the subject of a sentence, not in an objective or other syntactic relation. The radical element A (‘to cut up’), before entering into combination with the coordinate element B (‘to sit’), is itself compounded with two nominal elements or element-groups – an instrumentally used stem (F) (‘knife’), which may be freely used as the radical element of noun forms but cannot be employed as an absolute noun in its given form, and an objectively used group – (E) + C + d (‘black cow or bull’). This group in turn consists of an adjectival radical element (E) (‘black’), which cannot be independently employed (...), and the compound noun C + d (‘buffalo-pet’). The radical element C properly means ‘buffalo’, but the element d, properly an independently occurring noun meaning ‘horse’ (...), is regularly used as a quasi subordinate element indicating that the animal denoted by the stem to which it is affixed is owned by a human being. It will be observed that the whole complex (F) + (E) + C + d + A + B is functionally no more than a verbal base, corresponding to the *sing-* of an English form like *singing*; that this complex remains verbal in force on the addition of the temporal element (g) – this (g), by the way, must not be understood as appended to B alone, but to the whole basic complex as a unit – and that the elements (h) + (i) + (0) transform the verbal expression into a formally well-defined noun.

Yet even though this elaborate description corresponds directly to the left-branching IC-structure of fig. 6.4, Sapir does not draw the corresponding diagram, which is presented here (as is the custom nowadays) as an upside-down ‘tree’ structure, with the root at the top and the branches expanding downward.

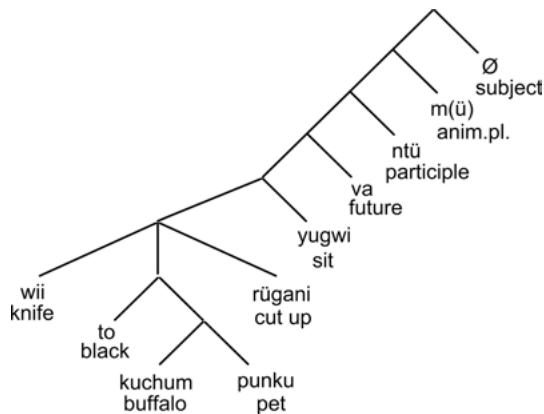


Fig. 6.4: The Paiute word *wii-to-kuchum-punku-rügani-yugwi-va-ntü-m(ü)* as a tree diagram according to Sapir (1921: 31–32)

roughly 1955, shrank from actually drawing tree diagrams will remain a mystery unless one realises that the drawing of diagrams was traditionally abhorred by scholars who worked in the human sciences and would not stoop to drawing figures, that being contrary to their strictly nonmathematical way of thinking. It took a long time for this attitude to be eliminated.

In fact (see Seuren 1998: 221–221), the first actually drawn tree diagram in the linguistic literature (with the root at the bottom end) is not found until Nida (1946: 87), shown in fig. 6.5, close to the end of the period we now identify as that of structuralist linguistics. Tree diagrams did not become common until after 1960.

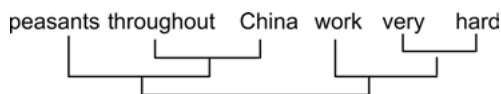


Fig. 6.5: Nida's (1946) analysis of *Peasants throughout China work very hard*

3.2. Bloomfield

The structuralist theory of syntax came to full fruition in the work of the American linguist Leonard Bloomfield (1887–1949), especially in his (1933) book *Language*. Bloomfield, whose Austrian paternal grandparents had immigrated into the United States in 1868, began as a student of Germanic philology, but he soon extended his interests to the general theory of language, taking in his stride a few American Indian languages and also Tagalog, the main language of the Philippines. In 1921 he won the chair of German and Linguistics at Ohio State University in Columbus. There he was influenced, and soon won over, by the psychologist Albert P. Weiss, who was an ardent follower of the new school of behaviourism in psychology which had started during the first world war.

In actual fact, behaviourism was much more than just a new school in psychology. It was an ideology presenting itself in the context of a positivist approach to whatever is taken to be connected with the mind and hence with society. Strictly speaking, behaviourism held that (a) human beings are fully material: the mind is a fabrication based on a phenomenological delusion and on religion, (b) the main problem of psychology is to find a causal connection between stimuli impinging upon a human or animal organism and those forms of behaviour that cannot be explained by direct physical causation (such as bleeding after a cut). The simplest possible hypothesis was taken to be the assumption of a physiological transfer mechanism in the brain (conditioning), in virtue of which an existing behavioural reaction corresponding to a stimulus X, such as salivation on the sighting of food, can be prompted also by a stimulus Y provided X and Y have co-occurred with sufficient frequency in the personal history of an individual. Thus, when the sighting of food has been accompanied often enough by the sounding of a bell, salivation will occur upon the sounding of the bell even when no food is sighted. The behaviourists believed that this simplest possible hypothesis, perhaps extended with a few refinements, was adequate for the explanation of human and animal behaviour. They left no room for any genetically fixed specific predisposition for the processing of incom-

ing stimuli, let alone for any spontaneous self-starting cognitive processes. For them, nervous systems were general purpose machines of a purely passive and associative type. During its heyday, behaviourism was extremely influential across the human sciences and even influenced US government policies. This heyday ended abruptly in the early 1960s, when the insight broke through that the explanation of human behaviour, in all its complexity, requires a far richer hypothesis than a behaviourist stimulus-response mechanism.

After his conversion to behaviourism, it became Bloomfield's aim to establish a theory of language according to behaviourist principles. The results of this resolve are found in his (1933) *Language*. Yet, although behaviourism is presented with a great deal of emphasis especially in the opening chapters of this book, its effects remain largely limited to Bloomfield's not altogether successful attempt at setting up a behaviourist semantics, according to which the meaning of a linguistic form consists in its behaviourist association with a set of physical stimuli. Other than that, fortunately, there is hardly any behaviourism to be detected in the book. The main significance of Bloomfield's express emphasis on behaviourism lies in the fact that it demonstrates his deeply rooted urge to turn linguistics into an autonomous science in the contemporary sense of the word.

His theory of syntax, in particular, has remained entirely free from behaviourist blemishes (contrary to the behaviourist psychologist Skinner's unsuccessful and unprofessional 1957 attempt at grounding a behaviourist theory of syntax). It is squarely based on the notion of IC-analysis described above, which he had derived from his study of Wundt, to whom, as will be agreed, no behaviourist tendencies can be ascribed. For Bloomfield, all linguistic products (sentences, phrases, words, morphemes) have an internal structure which is describable in terms of a layered hierarchy of constituents, as shown in the figures 6.4 and 6.5 above. He writes (1933: 161–162):

A linguistic form which bears no partial phonetic-semantic resemblance to any other form is a *simple* form or *morpheme*. (...) From all this it appears that every complex form is entirely made up, so far as its phonetically definable constituents are concerned, of morphemes. The number of these *ultimate constituents* may run very high. The form *Poor John ran away* contains five morphemes: *poor*, *John*, *ran*, *a*..., and *way*. However, the structure of complex forms is by no means as simple as this; we could not understand the forms of a language if we merely reduced all the complex forms to their ultimate constituents. Any English-speaking person who concerns himself with this matter, is sure to tell us that the *immediate constituents* of *Poor John ran away* are the two forms *poor John* and *ran away*; that each of these is, in turn, a complex form; that the immediate constituents of *ran away* are *ran*, a morpheme, and *away*, a complex form, whose constituents are the morphemes *a* and *way*; and that the constituents of *poor John* are the morphemes *poor* and *John*. Only in this way will a proper analysis (that is, one which takes account of the meanings) lead to the ultimately constituent morphemes. (...) The total stock of morphemes in a language is its *lexicon*.

Clearly, this description of the sentence *Poor John ran away* corresponds to the diagram in fig. 6.6, which, however, he does not actually draw, thus following in Sapir's footsteps.

Yet he does say explicitly that each constituent deserves a name or label denoting the grammatical category, or "form class", to which it belongs (Bloomfield 1933: 165):

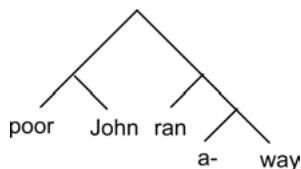


Fig. 6.6: Bloomfield's (1933) analysis of *Poor John ran away*

The meaning of a complex form depends in part upon the selection of the constituent forms. Thus, *drink milk* and *watch John* name actions, and, as we have just seen, are infinitive expressions, but *fresh milk* and *poor John* name objects and are substantive expressions. The second constituents, *milk* and *John*, are the same; the difference depends upon the selection of the first constituent. By virtue of this difference, the forms *drink* and *watch* belong to one English form class (that of transitive verbs), and the forms *fresh* and *poor* to another (that of adjectives).

His notion of form class is still not quite clear and in any case incomplete, in that he fails to present a list of possible form classes or, as we now prefer to say, grammatical categories, arguing (Bloomfield 1933: 165): “The features of selection are often highly arbitrary and whimsical. We combine *prince*, *author*, *sculptor* with the suffix *-ess* in *princess*, *authoress*, *sculptress* (in this last case with phonetic modification of [r,] to [r]), but not *king*, *singer*, *painter*. By virtue of this habit, the former words belong to a form-class from which the latter words are excluded.” But one may conclude that he considers that proper linguistic IC-diagrams, other than those presented by Sapir and Nida, should label their nodes by assigning them the grammatical category to which they belong. Bloomfield thus implicitly initiated the notion of labelled tree, now current in all forms of grammatical analysis, even though, as has been said, he never drew a tree diagram in any of his publications. If we anticipate the nomenclature that became current in the immediate post-Bloomfieldian period, we may rewrite fig.6.6 as the labelled tree shown in fig. 6.7 – even though no labels for the alleged morphemes *a-* and *-way* ever came into general use, and even though Bloomfield himself would analyse the finite verb form *ran* in terms of the underlying structure PAST + RUN:

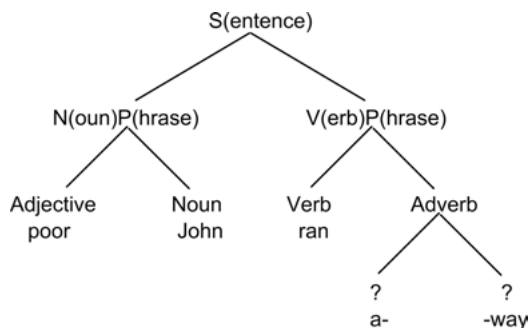


Fig. 6.7: The same as fig. 6.6, with labels attached to the nodes

One should realise that the type of constituent tree shown in fig. 6.7 can likewise be used for the computation of arithmetical functions. As a parallel to fig. 6.1 we may consider fig. 6.8, where likewise the functions $(5 \times 6) + 8$ and $5 \times (6 + 8)$ are computed:

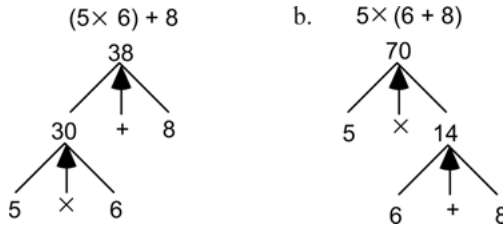


Fig. 6.8: $(5 \times 6) + 8$ and $5 \times (6 + 8)$ in terms of constituent trees

Here, the functor is itself a constituent of the superordinate constituent. It is necessary, for the arithmetical calculus to work, that the functor constituent be identified as such, so that it can be distinguished from its argument terms.

One problem with the type of analysis exemplified in fig. 6.7 is that it makes no distinction between morphological and syntactic structures. Bloomfield does have a definition of the notion sentence, which makes good sense in his perspective (Bloomfield 1933: 170): “An utterance may consist of more than one sentence. This is the case when the utterance contains several linguistic forms which are not by any meaningful conventional grammatical arrangement (that is, by any construction) united into a larger form.” And he defines the word as a “minimal free form” (Bloomfield 1933: 178). But what exactly distinguishes morpheme-to-word from word-to-sentence constructions remains unclear, despite Bloomfield’s insistence that they differ radically (Bloomfield 1933: 183–184):

In languages which use bound forms, the word has great structural importance because the constructions in which free forms appear in phrases differ very decidedly from the constructions in which free or bound forms appear in words. Accordingly, the grammar of these languages consists of two parts, called *syntax*, and *morphology*. (...) There has been considerable debate as to the usefulness of this division, and as to the scope of the two headings. In languages that have bound forms, the constructions in which bound forms play a part differ radically from the constructions in which all the immediate constituents are free forms.

The best he can do is in (Bloomfield 1933: 207), but no clear criterion is given:

In general, morphologic constructions are more elaborate than those of syntax. The features of modification and modulation are more numerous and often irregular—that is, confined to particular constituents or combinations. The order of the constituents is almost always rigidly fixed, permitting of no such connotative variants as *John ran away* : *Away ran John*. Features of selection minutely and often whimsically limit the constituents that may be united into a complex form.

One notes, incidentally, that Bloomfield still appears to have no eye for the real complexities and mechanisms of syntax. It wasn’t until a few decades later that syntactic phenomena became visible to linguistic observers.

Honesty forces us to admit that modern linguistic theories have equally failed to provide a criterion for the distinction between syntax and morphology, partly due to an overall failure to provide a clear definition of the notion word. All we do nowadays is classify categories (form classes) as being syntactic or morphological. Phrase and word classes are called syntactic, but stems and affixes are morphological. Differences be-

tween syntactic and morphological constructions are often commented upon, but only in an incidental way and without much in the way of general criteria.

A further serious problem with Bloomfield's IC-analysis lies in the fact that he fails to indicate on what grounds one particular possible IC-analysis is to be preferred over another. Why, for example, should we analyse *Poor John ran away* as consisting of an NP followed by a VP, and not as a tripartite structure of the kind shown in fig. 6.9?

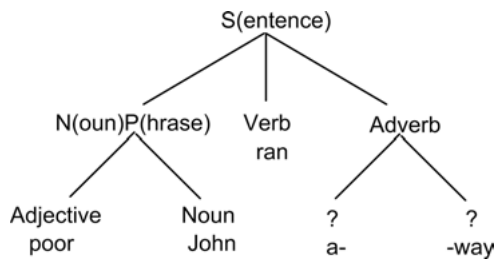


Fig. 6.9: Alternative IC-analysis for *Poor John ran away*

Bloomfield himself had no principled answer to this question, or if he had one, he merely adumbrated it. In cases where he still feels unable to encompass the notion of a complete description, as in the case of syntax, he falls back on intuitive, introspective criteria. Thus, as we saw in the larger quote given earlier from Bloomfield (1933: 161), he writes: “Any English-speaking person who concerns himself with this matter, is sure to tell us that the *immediate constituents* of *Poor John ran away* are the two forms *poor John* and *ran away*.” But when he is concerned with the restricted paradigms of morphophonemic alternation, or, in his own terms, “phonetic modification”, where the notion of complete description is easier to grasp, he applies the criterion of greatest overall simplicity to the problem of how to select a “basic alternant” (Bloomfield 1933: 164): “[W]e try, of course, to make the selection of a basic alternant so as to get, in the long run, the *simplest description of the facts*.” [emphasis mine]. Or Bloomfield (1933: 211–212):

We have not yet described, in terms of phonetic modification, the kinship of the three alternants [-ez, -z, -s] of the bound form that appears in English plural nouns. It is evident that three entirely different statements are possible, according to our choice of one or the other of the three forms as our starting point. Our aim is to get, in the long run, the *simplest possible set of statements* that will describe the facts of the English language. [emphasis mine]

What Bloomfield does not do is apply the criterion of greatest overall simplicity to the choice of immediate and remote constituents in syntactic constructions. That step was taken a decade and a half later by the American linguists Rulon Wells (1947) and especially Zellig Harris (1951).

3.3. The “God’s truth” linguists (Pike)

Naturally, the question of the motivation of syntactic IC-analyses was a central concern in the schools of structuralist linguistics that developed in the wake of Bloomfield’s

teaching. It was, in fact, this very question that caused a radical split among Bloomfield's followers. One school followed the European trend, holding that IC-structures are psychologically real and that they are to be discovered on grounds of introspection: one somehow feels what the right structure is, for any given sentence or phrase, when one closes one's eyes and thinks very deeply about it. To get the right results it helps a great deal when the person doing the deep thinking is himself a linguist or a psychologist, since only linguists and psychologists have the "highly cultivated refinement in the description of one's own thought experiences" described and rejected as a criterion by the German psychologist Bühler (1934: 254): "Man darf den Befund nicht für alle Zeit an die Bedingung einer hochgezüchteten Feinheit des Beschreibens eigener Denkerlebnisse knüpfen, sondern muß danach streben, ihn auch weniger subtilen Augen zugänglich zu machen und noch mehr: es gilt ihn *objektiv* zu verifizieren." [One cannot ground the result for all time on the condition of a highly cultivated refinement in the description of one's own thought experiences. On the contrary, one must attempt to make it accessible also to less subtle eyes, and, what is more, one must verify it *objectively*.]

Householder (1952: 260), dubbed this school of linguistic thought the school of God's truth linguists, for two good reasons. First, for these linguists, the truth about syntactic structure was, in a sense, a question of divine or subdivine revelation, best restricted to the happy few who called themselves specialists. A second good reason for the name *God's truth linguists* was the fact that the group in question consisted mainly of protestant missionaries, who, for religious reasons, were staunchly opposed to anything that reeked of positivism or behaviourism yet were anxious to be "modern" and "scientific" but only in a strictly Christian sense. The main figures were Kenneth L. Pike (1912–2000) and Eugene A. Nida (1914–2011), both active members of the *Summer Institute of Linguistics* (SIL), founded in 1934 by one W. Cameron Townsend, a devoted but linguistically naive missionary in Mexico. SIL's main aim was to have the bible translated into preliterate languages. Pike had studied with Sapir but he was also strongly influenced by Bloomfield, who was considered the chief scientific linguist of the period. Only Pike developed a theory of syntax. Nida worked on morphology but soon withdrew into bible translation.

Pike's theory of linguistic, including syntactic, structure is found in Pike (1967), first circulated in 1954. In this book, Pike presents his theory of *Tagmemics* – the term being derived from Bloomfield's rather artificial and not very well defined term *tagmeme* (Bloomfield 1933: 166), which never caught on. He starts out by stressing that language is just one aspect of overall human behaviour and that human behaviour is, in general, meaningful and cast into specific structural moulds, called *behaviouremes*. Typical behaviouremes are a church service, a football game or a breakfast. Each behavioureme consists of "emic" slots (the term *emic* is derived from the terms *phoneme* and *morpheme*, and, of course, Bloomfield's *tagmeme*), filled by "etic" material (the term *etic* being derived from *phonetic*). The breakfast behavioureme in Pike's own family is described (Pike 1954: 59–60 – section 5.2) as being preceded by pre-breakfast preparations, etically filled in by the playing of Tschaikovsky's Fifth Symphony on the family record player to get the children up, washed and dressed, followed by the calling of the children to come and sit down at the table. Then breakfast can begin. It consists of a slot for the saying of grace, followed by a slot for the drinking of fruit juice (etically filled by some drinkable substance deserving that name). Then there is a composite slot

for the main meal, which includes a subplot for cereals as well as one for toast. Simultaneously there is an overall slot for conversation, filled by actual etic speech material.

In parallel fashion, speech material is divided up in emic segments called *sentences*, each of which is again divided up in emic sentence parts, each emic element being filled by etic material. The etic fillers are always the ultimate, material realisations of the emic slots. Some emic constituents are complex in that they consist of subconstituents. The emic level of description presupposes that the analysing linguist be conversant with the language (or culture) at hand: without actual knowledge of the language (or the culture), all the linguist can do is record (phon)etic material. The functional system required for emic structure is the mental property of each speaker, and it is the linguist's task to discover that inner language system by introspection. Tagmemics is thus *realist* in the sense that there is taken to be an actual emic object of investigation to be described or at least approximated by means of introspection resulting in linguistic analysis and description.

Later, this type of linguistic description was called *taxonomic* or *item-arrangement style*, in the words of Hockett (1954). Both labels are apt, because tagmemics does indeed take it for granted that morphosyntactic structures result from a taxonomy of possible morphosyntactic patterns, each pattern consisting of slots to be filled by etic material. No room was left for the thought that some such "patterns" may owe their existence to an underlying rule system for the composition of structures and that it might make sense to assume hypothetical underlying structures to be transformed according to tree-transformational processes into surface structures assigned to actually appearing utterances. Following Bloomfield, *portmanteau* etic fillers were considered in morphology, filling two or more underlying morphological slots at the time (such as etic *ate* for emic PAST + *eat*), but nothing of the sort was envisaged for syntax. No proposals were made, in this school of linguistic thought, to assume, for example, an underlying $s[s[\text{that John be ill}]_{VP}[\text{be likely}]]$ (or: *that John is ill is likely*) for a surface structure like $s[[_{NP}[\text{John}]_{VP}[\text{be likely}_{VP}[\text{to be ill}]]]]$ (or: *John is likely to be ill*).

In spite of the severe limitations of this type of morphosyntactic description, hundreds of grammars were produced in this vein, mainly in the context of the Summer Institute of Linguistics. Strangely, despite the fact that it has been shown a thousand times over during the past half century that this type of taxonomic description is unable to account for the syntactic complexities of natural languages, new schools have come up lately (in particular so-called Cognitive Linguistics) which indulge in precisely this type of work, complete with appeals to introspection, simply ignoring the arguments and mountains of facts, especially regarding sentential and clausal complementation, that have been adduced over the years showing its fatal inadequacy. One cannot but conclude that linguistics is still far from having achieved the status of a mature science.

3.4. The "hocus-pocus" linguists (Harris)

It was again Householder (1952: 260) who introduced the term *hocus-pocus linguists* to refer to the other main post-Bloomfieldian school which came into being during the 1940s. His reason for choosing this name was the fact that the school in question showed a predilection for abstract formulaic notation and formal methods of description and

analysis, strongly influenced by the rather esoteric foundational studies carried out, during the first half of the twentieth century, in the areas of mathematics and logic (see Scholz and Pullum 2007).

The school was based on the notion of overall description of a language in terms of Bloomfieldian constituency trees. The best overall description of a language was considered to be the one that covered all the facts with the help of the smallest number of IC-construction types. Minimalisation of the descriptive apparatus was the word, in accordance with the Ockhamist simplicity criterion: “postulated entities are not to be multiplied beyond necessity”. The best IC-analysis in any given case is thus the one that fits into an overall description of the language that is both adequate and minimalist.

The main protagonists of this school were the Yale philosopher Rulon S. Wells (1919–2008), whose influential article (1947) was widely read, and the Philadelphia linguist Zellig S. Harris (1909–1992), who became the leader of this movement after the publication of his (1951) book *Methods in Structuralist Linguistics*, foreshadowed by Harris (1946). Both Wells and Harris proposed a method whereby substitution classes should be established for each constituent in a construction in such a way that a maximally simple overall picture would emerge, whereby it was not excluded that two or more candidates would achieve equal scores on the simplicity scale (Harris 1951: 2). But whereas Wells simply proposed the method, Harris gave it a behaviourist grounding and worked it out in minute detail for phonology, morphology and syntax. Harris, moreover, took a great interest in the foundational studies in mathematics and logic that had been and were taking place in those years and was deeply influenced by them (see Scholz and Pullum 2007) and, unlike Pike and his followers, he did not eschew abstract analyses and symbolic notations. He taught extensively about this way of doing syntax and kept publishing till well into the 1970s.

It was Harris also, who soon discovered, around 1950, that grammars will gain enormously in simplicity if one postulates a structurally regular and simple system of production rules for underlying sentence structures, supplemented with transformational rules turning the underlying or deep structures into surface structures. The true originator of what was later to be known as Transformational Generative Grammar was, therefore, Zellig S. Harris. Yet this new transformationalist approach was not incorporated into his (1951), which he had finished writing in 1947, as appears from the Preface. It gradually developed through his later publications (1952, 1957, 1965), where he also relaxed his initial strictly positivist and strictly behaviourist convictions. The transition to the transformational paradigm is not discussed in the present article, because it is felt that it marks the end of the structuralist era in American linguistics, which is characterised by taxonomic methods of classification, by an adherence to behaviourism and by a rejection of the notion of explanation in science.

Harris (1951) is a very dull book, yet it is carried by genuine conviction and real inspiration from beginning to end. It starts by saying, entirely in a positivist vein, that linguistic data consist in the sounds actually produced by people going about their business. What people understand or comprehend is not considered data because comprehension is not physically graspable. (Later, Harris gave up his strict positivism and became much more lenient with regard to semantic data.) The linguist should thus set out collecting sound samples from people in streets, trains, planes and buildings and then lock himself up with his piles of sound recordings. He should then try to sort out what minimal inventory of recurrent sound bits can be combined into units called phonemes.

This is to be done on the basis of the *distribution* of the sound bits in their various sound environments. If it turns out, as Harris said it would, that certain types of sound bit are restricted to certain environments and that these environments do not overlap, then a phoneme X can be set up as an abstract unit of description such that X has different and well-defined manifestations, called *allophones*, in each environment listed. The environment-conditioned allophones are said to be in *complementary distribution* (a notion taken from Swadesh 1934, who had introduced it in the context of the determination of phonemes). Elements in complementary distribution are open to classification as one single higher-level unit of description, in this case a phoneme. Thus we read (Harris 1951: 25):

As the first step towards obtaining phonemes, this procedure represents the continuous flow of a unique occurrence of speech as a succession of segmental elements, each representing some feature of a unique speech sound. The points of division of these segments are arbitrary here, since we have as yet no way of enabling the analyst to make the cuts at precisely those points in the flow of speech which will later be represented by inter-phonemic divisions. Later procedures will change these segmentations until their boundaries coincide with those of the eventual phonemes.

Essentially the same method can be applied to set up higher-level descriptive units. Combinations of phonemes will be seen to form *allomorphs*, to be combined into higher-level units called *morphemes*. Morphemes again occur in constructions allowing for the setting up of ever higher-level descriptive categories until one arrives at the highest-level descriptive unit, the sentence. The book takes the reader, step by repetitive step, through all the levels proposed up till sentence level.

The notion of distribution is essential for Harris. It is considered virtually the only criterion for setting up tentative analyses, which are further tested on grounds of simplicity and generality of structure assignments (Harris 1951: 6): "The only preliminary step that is essential to this science is the restriction to distribution as determining the relevance of the inquiry.". And again (Harris 1951: 8): "It may be noted that distributional procedures do more than offer a rigorous alternative to meaning considerations and the like. Distributional procedures, once established, permit, with no extra trouble, the definite treatment of those marginal cases which meaning considerations leave indeterminate or open to conflicting opinion."

The only concession to "meaning" is made for cases where informants are asked to repeat an utterance or to say whether two sound bits are the same or different (Harris 1951: 7): "In principle, meaning need be involved only to the extent of determining what is repetition. If we know that *life* and *rife* are not entirely repetitions of each other, we will then discover that they differ in distribution (and hence in 'meaning')."

The book is thus a massive *discovery procedure*, ideally based on a large corpus of sound recordings, for a maximally compact statement of all possible constructions in a language at the different ascending levels of phonemes, morphemes, words, phrases and, finally, the sentence. The aim is to establish an axiomatised discovery procedure for the simplest possible grammar or grammars of a language.

A grammar set up in this way is not necessarily considered to be actually present in some form in the speaker's brain but is primarily taken to be merely a way of managing, or, as Harris says (1951: 3) "arranging" the data, whereby the possible reality of an object of description is either left open or denied. Harris was thus not a realist but,

entirely in the spirit of his day, an *instrumentalist* in the sense of that term current in the philosophy of science (Nagel 1961) and succinctly described in Botha (1968: 89):

Instrumentalists regard theories as instruments, tools for calculation or computational devices for the organising of data and ordering of laws. According to this view theories are used for drawing inferences and making predictions and their statements cannot be characterised as either true or false.(...) Instrumentalists do not always agree on the question whether theoretical concepts refer to entities in reality.

A telling passage, in this respect, is Harris (1951: 3):

The greatest use of such explicit structural descriptions will be in the cataloguing of language structures, and in the comparing of structural types. These descriptions will, however, be also important for historical linguistics and dialect geography; for the relation of language to culture and personality, and to phonetics and semantics; and for the comparison of language structure with the systems of logic.

Here, linguistic descriptions are said to be useful primarily for the setting up of taxonomies and in a secondary but ill-defined sense also for historical linguistics, dialect geography, phonetics and semantics, logic and what not, but where no mention is made of explanation or enhancement of insight.

Other than in Pike's tagmemics, which led to hundreds of actually written and published grammars, Harris's method did not lure any linguists into the activity of grammar-writing. The reason is obvious: what linguist will condemn himself or herself to such drudgery! Harris, of course, realised that. For him, the method of grammar-discovery he set out in his book was more an intellectual exercise than a practical proposal.

But then, how should a linguist who aspires to write an actual grammar of a language actually go about? Towards the end of the book, Harris proposes an answer. He first recapitulates his main purpose in writing the book as follows (Harris 1951: 366–368):

The over-all purpose of work in descriptive linguistics is to obtain a compact one-one representation of the stock of utterances in the corpus. Since the representation of an utterance or its parts is based on a comparison of utterances, it is really a representation of distinctions. (...)

The basic operations are those of segmentation and classification. Segmentation is carried out at limits determined by the independence of the resulting segments in terms of some particular criterion. (...) Classification is used to group together elements which substitute for or are complementary to one another. (...)

If we were analyzing a corpus without any interest in its relevance for the whole language, we could list all the environments of each tentative segment in all utterances of the corpus, and on this basis decide the segmentation in each utterance. Usually, however, we are interested in analyzing such a corpus as will serve as a sample of the language.

Then he observes that the higher the constituent level, with “sentence” as the highest of all, the smaller the inventory of categories but the larger the set of possible fillers (Harris 1951: 369–370): “This leads ultimately to sets of few elements having complex definitions but as nearly as possible random occurrence in respect to each other, replacing the original sets of many elements having simple definitions but complexly restricted distribution.” Then, at last, light breaks through (Harris 1951: 372–373):

The work of analysis leads right up to the statements which enable anyone to synthesize or predict utterances in the language. These statements form a deductive system with axiomatically defined initial elements and with theorems concerning the relations among them. The final theorems would indicate the structure of the utterances of the language in terms of the preceding parts of the system.

The style of his writing is abstruse, convoluted and plodding, but the message is clear. When we read “statements” as “rules” and “theorems” as “sentences”, as indeed intended by Harris, we have, in essence, the concept of *generative grammar*: Not (yet) of transformational generative grammar, because what is proposed amounts to a system of rules rewriting a single given symbol as a succession of new symbols (so-called *phrase-structure* rules), without any operations performed on strings of symbols. But it did not take Harris long to realise that much greater descriptive succinctness could be achieved when one posits certain simple sentence structures as “basic” and derives more complex sentence from these by means of transformational rules, as has been said.

A similar thought had already been expressed in Harris’s (1946) article *From morpheme to utterance*, where he opposed the bottom-up discovery procedure (from morphemes to sentences) to the top-down synthetic procedure now known as “generative”. There we read (Harris 1946: 178):

The [discovery] procedure outlined here could be paralleled by a series of substitutions beginning with the whole utterance and working down, instead of beginning with single morphemes and working up. In that case we would have to find formal criteria for breaking the utterance down at successive stages. This is essentially the difficult problem of determining the immediate constituents of an utterance. It is not clear that there exists any general method for successively determining the immediate constituents, when we begin with a whole utterance and work down. In any case, it would appear that the formation of substitution classes presents fewer theoretical difficulties if we begin with morphemes and work up.

In hindsight, we may say that what is still lacking, both in Harris (1946) and in Harris (1951), is the notion of a hypothesis regarding the IC-analysis of sentences and subsequent testing, by means of an *evaluation procedure*, for greatest generality and thus greatest simplicity. Rulon Wells proposed such an evaluation procedure in 1947, but this, apparently, made little impression on Harris.

Even so, one clearly discerns the influence of foundational studies in mathematics and logic on linguistic work of this nature, and especially on Harris, who was a keen follower of these developments. This influence was pointed out and further elaborated in Harwood (1955). Scholz and Pullum (2007) point to the great influence of the work done by the mathematician Emil Post during the 1920s in the theory of algorithms as purely formal systems for the generation of strings of symbols (sentences). They also express their amazement at the fact that Post is never quoted by the linguists in question (especially Noam Chomsky), even though they owe a direct debt to him. (The same was already observed by the Dutch philosopher-logician Evert Beth in his Academy address of 1963.).

4. Conclusion

Perhaps to the surprise of beginning students of syntax, we must conclude that it was not until quite recently that the notion of syntax began to have body and soul. Prestruc-

turalist or Apollonian syntax was heavily bound up with nominal case systems and the question of what cases are governed by what verbs with what meaning. During the period of structuralist linguistics, at least to the extent that it was theory-driven, linguists were desperately trying to get a clear notion of syntactic facts, to which the linguistic world had, on the whole, been blind. In the words of Harris (1946: 161): “(...) many grammars have carried little or no syntactic description.” Much as one may feel that the main school of transformational generative linguistics, started by Zellig S. Harris during the 1950s but quickly taken over by Noam Chomsky at MIT, soon lost its bearings and went haywire, it would seem that a sober and balanced assessment will lead to the conclusion that it was in the context of transformational linguistics that, eventually, the Apollonian view of syntax was replaced by the now current, much richer, idea of what syntax and syntactic structures amount to. What we call “structuralist syntax” was, in effect, but a stepping stone to the modern, more developed notion. To what extent this modern notion will prove fruitful and adequate, is a question which it would be premature to expect an answer to.

5. References (selected)

Aarsleff, Hans

1982 *From Locke to Saussure. Essays on the Study of Language and Intellectual History.* London: Athlone.

Bech, Gunnar

1955 *Studien über das Deutsche Verbum Infinitum, Band I.* Det Kongelige Danske videnskaberne selskab. Historisk-Filosofiske Meddelelser, 35:2 (1955) and 36:6 (1957). Copenhagen: Munksgaard. [2nd unrevised edition 1983 Niemeyer, Tübingen (Linguistische Arbeiten 139)]

Beth, Evert W.

1963 Konstanten van het wiskundige denken [Constants in mathematical thinking]. In: *Mededelingen der Koninklijke Nederlandse Akademie van Wetenschappen. Afdeling Letterkunde*, 231–256. (Nieuwe Reeks 26.7) Amsterdam: North-Holland.

Bloomfield, Leonard

1914 *An Introduction to the Study of Language.* New York: Henry Holt. [Photostatic reprint 1983, Amsterdam/Philadelphia: Benjamins]

Bloomfield, Leonard

1924 Review of *Cours de linguistique générale* by Ferdinand de Saussure. *Modern Language Journal* 8: 317–319.

Bloomfield, Leonard

1933 *Language.* New York: Henry Holt.

Botha, Rudolph P.

1968 *The Function of the Lexicon in Transformational Generative Grammar.* The Hague/Paris: Mouton.

Bühler, Karl

1934 *Sprachtheorie. Die Darstellungsfunktion der Sprache.* Jena: Fischer.

De Saussure, Ferdinand

1916 *Cours de linguistique générale.* Edited by Charles Bally and Albert Sechehaye, with the collaboration of Albert Riedlinger. Paris/Lausanne: Payot.

Fodor, Jerry A.

1983 *The Modularity of Mind.* Cambridge, Mass.: MIT Press.

- Harris, Zellig S.
1946 From morpheme to utterance. *Language* 22.3: 161–183.
- Harris, Zellig S.
1951 *Methods in Structural Linguistics*. Chicago: The University of Chicago Press.
- Harris, Zellig S.
1952 Discourse analysis. *Language* 28.1: 1–30.
- Harris, Zellig S.
1957 Co-occurrence and transformation in linguistic structure. *Language* 33.2: 283–340.
- Harris, Zellig S.
1965 Transformational theory. *Language* 41.3: 363–401.
- Harwood, F. W.
1955 Axiomatic syntax. The construction and evaluation of a syntactic calculus. *Language* 31.4: 409–413.
- Hockett, Charles F.
1954 Two models of grammatical description. *Word* 10: 210–231.
- Hockett, Charles F.
1958 *A Course in Modern Linguistics*. New York: Macmillan.
- Householder, Fred W.
1952 Review of Zellig S. Harris (1951). *International Journal of American Linguistics* 18: 260–268.
- Householder, Fred W.
1981 *The Syntax of Apollonius Dyscolus. Translated, and with commentary by –*. Amsterdam/Philadelphia: Benjamins.
- Lipps, Theodor
1893 *Grundzüge der Logik*. Leipzig: Dürr.
- Mel'čuk, Igor A., and Nikolaj V. Pertsov
1987 *Surface Syntax of English. A Formal Model within the Meaning-Text Framework*. Amsterdam/Philadelphia: Benjamins.
- Nagel, Ernest
1961 *The Structure of Science. Problems in the Logic of Scientific Explanation*. London: Routledge and Kegan Paul.
- Nida, Eugene A.
1946 *Morphology. The Descriptive Analysis of Words*. Ann Arbor: The University of Michigan Press.
- Paardekooper, Piet Cornelis
1955 *Syntaxis, Spraakkunst en Taalkunde* [Syntax, Grammar and Linguistics]. Den Bosch: Malmberg.
- Percival, W. Keith
1976 On the historical source of immediate constituent analysis. In: James D. McCawley (ed.), *Notes from the Linguistic Underground*, 229–242. (Syntax and Semantics 7.) New York/San Francisco/London: Academic Press.
- Pike, Kenneth L.
1967 [1954] *Language in Relation to a Unified Theory of the Structure of Human Behavior*. The Hague: Mouton. [First put into circulation by Summer Institute of Linguistics, Glendale, California, 1954]
- Sapir, Edward
1921 *Language. An Introduction to the Study of Speech*, New York: Harcourt, Brace and Cy.
- Scholz, Barbara C., and Geoffrey K. Pullum
2007 Tracking the origins of transformational generative grammar. (Review article of Marcus Tomalin, *Linguistics and the Formal Sciences: The Origins of Generative Grammar*. Cambridge, Cambridge University Press, 2006) *Journal of Linguistics* 43.3: 701–723.

- Seuren, Pieter A. M.
1998 *Western Linguistics. An Historical Introduction*. Oxford: Blackwell.
- Skinner, Burrhus F.
1957 *Verbal Behavior*. New York: Appleton-Century-Crofts.
- Stout, George F.
1896 *Analytical Psychology*. 2 vols. London: Macmillan / New York: Sonnenschein.
- Swadesh, Morris
1934 The phonemic principle. *Language* 10.1: 117–129.
- Tesnière, Lucien
1959 *Éléments de syntaxe structurale*. Paris: Klincksieck.
- Wegener, Philipp
1885 *Untersuchungen über die Grundfragen des Sprachlebens*. Halle: Niemeyer.
- Wells, Rulon S.
1947 Immediate constituents. *Language* 23.1: 81–117.
- Wundt, Wilhelm
1880 *Logik. Eine Untersuchung der Prinzipien der Erkenntnis und der Methoden Wissenschaftlicher Forschung*. Stuttgart: Enke.
- Zholkovsky, Aleksandr K., and Igor A. Mel'čuk
1965 O vozmožnom metode i instrumentax semantičeskogo sinteza [On a possible method and some tools for semantic synthesis]. *Naučno-texničeskaja Informacija* [Scientific and Technological Information] 6: 23–28.

Pieter A. M. Seuren, Nijmegen (Netherlands)