An historical perspective on syntactic trees

Dany Jaspers

Hogeschool voor Wetenschap en Kunst, campus VLEKHO

djaspers@vlekho.wenk.be

1. Introduction*

The history of the two basic rule types in generative syntax makes a most interesting tale. Rewriting rules and transformations - called *merge* and *move* in their current guise - have undergone several modifications over the years, both reflecting and guiding evolutions in the generative approach to syntax. In this article, we shall take stock of the specific properties of these precursors of *merge* and *move*, with special attention to similarities and differences. The picture that emerges shows a rapprochement between these two rule types. Phrase structure rules in particular have undergone a thorough change. The transformation rule, for its part, has in some sense come full circle. Its current format - though very different in its details from earlier characterisations and embedded in a wider context which is far more advanced - bears a surprising resemblance to that of the early days of generativism.

2. The origin and growth of syntactic trees

In 1914, Leonard Bloomfield (1887-1949) - probably inspired by Wilhelm Wundt (Seuren 1990:32) - introduced tree structures into the theory of grammar. Surely, the idea that words are organized into phrases was well-established in traditional and historical grammar. Referring to Richard Ogle, Chomsky (1986a) claims that it is an idea "that was clear in the 17th century Port-Royal grammar and that has antecedents in classical rhetorical theory, where "non-figurative style" is characterized as "untransposed," that is, with "words that are connected placed next to one another," in the formulation of a Sophist handbook of the Aristotelian period." (A more extensive discussion of how Chomsky perceives the Port Royal-style phrase structure system can be found in his 1966 *Cartesian Linguistics*). It was not until Bloomfield published *Language* in 1933, however, that the idea was explored with a fair degree of precision. Tree representations of linguistic structures in terms of Immediate Constituent Analysis feature very prominently in the book and though they have been modified and adapted to the particular concerns of different later frameworks, their success has remained constant.[1]

Ten years after Bloomfield first introduced tree structures, the Danish traditional grammarian Otto Jespersen published his *Philosophy of Grammar* (1924) in which he stated that besides those things in language which are formulaic (and often irregular) in character, there are also "free expressions", built on the basis of a "notion of (...) structure" (Jespersen 1924:19) which guides the speaker in "framing sentences of his own". These expressions always "show a regular formation."(Jespersen 1924:24). This idea

---

[1] The term 'tree' is used in this article as a cover term for Immediate Constituent analyses and pictorial representations thereof. It is a notable feature of *Language* that Bloomfield refrained from presenting any kind of arboreal graphs, attributed to a "gentlemanly reluctance to stoop to drawing figures" by Seuren (1998: 204).

of rule-governed freedom, of creativity of language use with a regular aspect attributable to a mental notion of structure, would suggest that the trees introduced by Bloomfield to represent such structure reflect a mental reality, an association which Bloomfield himself would certainly not subscribe to.[2]

Among Bloomfield's students and followers, some - including Kenneth L. Pike and Eugene A. Nida - did take the (anti-behaviourist) position that these constructs reflect a mental reality to be discovered through introspection. Others - including Charles F. Hockett, Rulon S. Wells and Zellig Harris - rejected such an intuitive approach along with its rather subjective criteria. In conformity with their structuralist and behaviourist views, they viewed tree structures as no more than handy instruments in their efforts to classify an ever-expanding compilation of language data. Given the enormity of the data base, one of their main concerns was *economy* of description, and tree representations were felt to be helpful tools in the pursuit of their descriptive objectives.

Zellig Harris' version of phrase structure, however, originally in his morpheme-to-utterance paper and then in *Methods in structural linguistics* (Harris 1951) was not quite the same as Bloomfield's, Wells' and others, and immediately foreshadows the later generative phrase structure system. Actually, it bears resemblance to what later came to be known as X'-theory in generative grammar. As stated in Chomsky (1964: 11), the "unordered set of rewriting rules" in his standard theory is a system which "is a generalization from Harris' morpheme-to-utterance statements, which constitute the nearest approach to an explicit generative grammar on this [= the syntactic] level" Chomsky (1964: 11, fn. 4). What is however absent from Harris' approach is the idea of generation, in the sense that grammars like his were typically "not conceived as generative grammars, but as descriptive statements about a given corpus" (Chomsky 1964: 11). His methods were supposed to be operational procedures for reducing a corpus to a more usable form. Though they are often referred to as "discovery procedures", Harris' philosophy of science was instrumentalist, so that there was nothing "real" to be discovered: one used whatever procedures seemed appropriate for the purpose at hand. This non-realist stance, combined with the absence of the notion of recursivity and a less accurate notion of the formal properties of both rewriting and transformational rules and rule systems, makes his methods different from later developments.

With the advent of generative grammar, Jespersen's abovementioned traditional concerns, which had become anathema in Post-Bloomfieldian instrumentalist circles, were revived and Bloomfield's tree concept was analysed with more insistence than ever before. Indeed, from its earliest stages onwards, the main concern of the Chomskyan generative research programme has been the nature and the growth of knowledge of language in the individual, clearly a mentalist perspective. To come to terms with the nature of linguistic knowledge a rich and diverse set of technical mechanisms was called upon and defined with formal precision, including *phrase structure rules* and *transformations*. In view of the insight that severely limited experience suffices to acquire an extremely complex and rich system of knowledge of language these mechanisms were surmised to be part of the genetic component in the acquisition of knowledge of language. The format of the innate system of Universal Grammar (UG) proposed in the earliest generative work not only sanctioned these two types of rules, it was also shown that some version of them was indispensable for representing the complexities of categorial structure and for describing the way in which syntactic structures could be changed in the course of the derivation of a sentence. While clearly motivated by the abovementioned insight that the real task was not assigning tree structures to individual sentences but rather finding an account for the sum total of the language user's linguistic competence, for his ability to make creative "infinite use of finite means" (Humboldt 1836), the introduction of *sets of generative rules* was made possible in practice by new tools provided by "the theory of computation (recursive function theory, the theory of algorithms)" (Chomsky 1986a).[3] Their formal outlook was in

---

[2] In the fragment from *The Philosophy of Language* where the notion of structure is introduced, we read the following: "from innumerable sentences heard and understood he will abstract some notion of their structure". It is not clear from this quote whether Jespersen believes the learning process is mediated by an innate language faculty or purely the result of habit formation. What is abundantly clear from the rest of the text, however, is that the learner/speaker uses a system which gives him ample room for creativity and freedom (cf. Chomsky's CALU: creative aspect of language use). He is not "a mere slave to habits, but has to vary them to suit varying needs - and this in course of time may lead to new turns and new habits; in other words, to new grammatical forms and usages."(Jespersen 1924:29)

[3] As mentioned above, this shift in focus from trees to the structure of rules and rule sets generating them, in other

part suggested by work in what is now called "generative phonology, and in this domain "rewriting rules" of the form X→Y, where X is an expression "rewritten " as Y in the course of derivation, seems an appropriate device" (Chomsky & Lasnik 1993).

With the advantage of hindsight, it may be claimed that the decision in the fifties and early sixties to factor the complexity of syntactic structures into two *completely independent* components - phrase structure rules which form phrase-markers and transformational rules to express global properties of expressions - was at least partially incorrect. Though the specific implementation made the newly explored and doubtless real displacement property stand out very clearly, it was flawed by the total disjointness of the two rule types, which made them look so different as to seem totally unrelated. Particularly the choice to formalize phrase structure rules as top-down rewrite rules has had some unfortunate consequences, as it gave birth to the idea that all phrase structure generation occurs prior to lexical insertion and further transformational computation. This first of all resulted in a massive overlap between the categorial component and the lexicon in Standard Theory. These two components both specified the class of subcategorization frames in which lexical items may appear. Admittedly, this particular redundancy was soon observed and efforts to get rid of it were made. General, universal properties of phrase structure were abstracted from the rewrite rules and assigned to a component of UG called X'-theory, and those properties that had to be stated in the lexicon anyway were thereby eliminated from the phrase structure component. Secondly, however, the radically different format of rewrite rules and transformations also firmly established the internal interface level Deep-structure, a complete Phrase-marker formed by the application of the first-stage base rules, i.e. of phrase structure rules and lexical insertion. For a very long time, this level of representation remained unassailable, for a range of reasons, the most important of which are (a) the strong intuitive appeal of a deep vs surface structure model and (b) the emergence and influence of the Katz-Postal hypothesis, according to which all of the information necessary for the interpretation of the sentence has to be present before transformations apply.

The intuitive appeal of the deep vs. surface structure model, first of all, was very strong in the days of the Standard Theory.[4] It was felt that it reconciled generative grammar with a long tradition. Chomsky (1964:15) wrote:

> "The transformational model, so described, expresses a view of the structure of language which is not all new. In particular, we find the observation that the semantic content of a sentence is represented only in an unexpressed deep structure, based on elementary underlying strings, in the *Grammaire générale et raisonnée* of Port-Royal (1660)."

Actually, there is a link with an even older tradition: the following fragments, with their conception of a misleading surface masking an underlying unexpressed deep structure more streamlined for semantic interpretation have an unmistakable Platonic ring to them.

> "Notice that in the case of the transformational model, the symbols and structures that are manipulated, rewritten and transformed as a sentence is generated may bear no very direct relation to any of its concrete subparts (...)[at the surface; DJ]" (Chomsky 1964:16)
> "The important point is that according to this conception of grammatical structure, the categories and grammatical functions represented in the actual labelled bracketing of the temporally given string will, in general, not be those that determine the semantic interpretation of this string, though they will be directly related to its phonetic interpretation." (Chomsky 1964:15)

Some linguists - the *generative semanticists* - were very charmed by this philosophical backing and the alluring

---

words, the idea of generation, was the central change from Harris' to Chomsky's model. One might be inclined to think that this novelty derived from Zellig S. Harris' (1951:366-377) notion of "synthesizing" sentences, but this was for Harris in his *Methods* certainly not a central, but at most a derivative notion.

[4] The deep-surface distinction even 'struck a chord' outside the field of linguistics, witness Leonard Bernstein's 1972 Harvard Lectures on deep and surface structures in music. Such success, by the way, has often led to misconceptions of linguistic theory among nonlinguists. There has been "an unfortunate tendency to confuse syntactic deep structure - a level of derivation - with innate universal grammar - which helps determine the nature of rules applying to all levels of derivation. Thus one sometimes encounters statements like "Chomsky posits an abstract 'deep structure' in language which is universal and innate."" (Jackendoff 1983: 242 fn.7)

elegance of this kind of model with its strict association between misleading surface and sound on the one hand and logically purer depth, semantics and thought on the other: they equated deep structure with semantic representation and gave it the format of some version of the predicate calculus.

The Katz-Postal deep structure hypothesis, secondly, stated that deep structure alone is the input to the semantic interpretive component. Consequently, all of the information necessary for the interpretation of the sentence had to be represented in the underlying generalized phrase marker at the relevant level and transformations were thought not to be able to introduce meaning-bearing elements[5]. The strong KP hypothesis - that deep structure *alone* determined meaning - was quickly overturned on the basis of sentences with multiple quantifiers like the following (which had already been discussed and judged problematic for the hypothesis that transformations are meaning-invariant in Chomsky 1957:100-101).

(1) a.   Everyone in the room knows at least two languages.
     b.   At least two languages are known by everyone in the room.

If transformations do not change meaning, the prediction is that an active sentence and its passive counterpart are synonymous. Yet, the (normal) interpretations of the above sentences differ in that the order of interpretation of the quantifiers parallels the *surface* order of the quantifiers. One might try to save the hypothesis by postulating different deep structures for these sentences and by stipulating that the order of quantifiers in the underlying semantic representation must correspond to the order in surface structure. This is the turning generative semantics - or *semantic syntax*, in Seuren's (1974) terminology - took. Adherents of the Extended Standard Theory and its descendants, for their part, concluded from these facts that a relocation of semantic interpretation was in order and (gradually) shifted it to S-structure. Simultaneously, narrow restrictions were imposed on possible D-structures by constraining the format of transformations and thereby enforcing a much closer structural relationship between D-structures and S-structures than ever before.

In sum, the different format of rewrite rules and transformations has had a great -and rather disadvantageous- influence on the conception of the grammar model. It gave rise to D-structure and ushered in the Katz-Postal hypothesis, which was instrumental in the canonization of D-structure as a 'deep', 'special', and for a long time unassailable level of representation. On the positive side, it should be mentioned that the Katz-Postal hypothesis was in effect one of the first large-scale attempts to constrain the power of transformations, sizing the set down to its meaning-preserving singulary subset. It is precisely this kind of drive to constrain operations which has brought to light a number of similarities between phrase-structure rules and transformations enabling some kind of unification.

3.   Unification of phrase-structure rules and transformations

Efforts to constrain operations largely grew out of the realization that their considerable expressive power posed a problem. For a language to be learnable under the boundary conditions set by experience, the contribution of UG not only has to be rich enough to account for "the attested variety of languages" (Chomsky 1986a:51) but at the same time sufficiently constrained to make very few languages available to the language-learner, given the data. If UG is not sufficiently restrictive and specific, far too many possible grammars will be accessible on the basis of the experience that the learner is exposed to, clearly an unwanted outcome, since people in the real world do manage to arrive at the grammar of their particular language on the basis of limited data. So, while the theory of UG incorporating the rich descriptive mechanisms of the standard theory met the condition of descriptive adequacy to the extent that its rules correctly described the linguistic facts they were designed to deal with, it was at the same time lacking in explanatory adequacy. Chomsky has repeatedly noted that it "is the tension between these two tasks that

---

[5] This is the stronger form of the Katz-Postal hypothesis. Jackendoff (1972:7) gives the following definitions of the weak form KP1 and the strong form KP2 respectively:
>   KP1: "Semantic projection rules operate exclusively on underlying phrase-markers; hence transformations do not change meaning."
>   KP2: "All semantic information is represented in underlying structure."
KP2 is equivalent to KP1 only if projection rules cannot add meaning in addition to the semantic features contained in the lexical items in the sentence.

makes the field an intellectually interesting one" (Chomsky 1986a:52). Indeed, for the purpose of accurate description it often seems necessary to resort to new devices, whereas to make the theory explanatorily adequate the system of devices has to be restricted as much as possible.

3.1.   Construction-independence

On the basis of these considerations, the language- and construction-specific transformations as well as the phrase structure rule sets of the standard theory period were subjected to scrutiny, and the universal aspects in them were gradually reduced to a very simple format, supplemented with overarching universal principles imposing conditions on the representations formed by the rules. For transformations, the universal movement scheme was reduced to the construction-independent rule *move*, or even *affect* in the GB period. The process of filtering out its construction-particular features led to a partitioning into two basic types: A-movement - uniting passives, ergatives, raising - and A'-movement - including wh-interrogatives, relatives, *though*-movement[6], *tough*-movement. Phrase structure, for its part, was made category-independent and reduced to a general structural template called X'-theory. These developments have been instrumental in laying bare some core properties which were discovered independently for both components, but which turn out to be no more than manifestations of the same property in two rule systems of the computational system. Since these properties characterise both X'-theory and movement, it is plausible to assume that these two components have more in common than the original formalisation of their operations suggested, an issue that will be gone into in the sections 2.2. - 2.4.

3.2.   The bottom-to-top property

3.2.1.   Transformations

Vertical order plays a very important and pervasive role in the transformational component in the form of dominance and the slightly more complex structural relation whose definition crucially refers to dominance, viz. c-command. It is worth repeating and stressing that the c-command relation between two constituents is asymmetric in the sense that it is the moved constituent which has to c-command the dependent copy in the source position, and not vice versa. This description amounts to requiring that the internal structure of the transformational operation *move* have a bottom-to-top orientation. In *Reflections on Language*, Chomsky (1975:107-110) made this point explicitly with respect to NP-movement (though it can be generalized):

> "the permissible rules are rules of "upgrading" which move a noun phrase closer to the "root of the sentence," that is, to a less embedded position; the impermissible rules are rules of "downgrading," which increase the embeddedness of the noun phrase. We might stipulate, then, that upgrading rules are permitted, but not downgrading rules" (Chomsky 1975:107).

Referring to Robert Fiengo, Chomsky (1975:108) consequently reduced the upgrading conjecture to independent principles of anaphora. Anaphoric dependency relationships show the same asymmetry that can be observed in the domain of movement: licit anaphoric relationships are those where the dependent element, i.e. the anaphor, is c-commanded by its antecedent, and not vice versa.

---

[6] Constructions produced by this rule, are (i) and (ii) (from Jackendoff 1972:128):
      (i) Handsome though Melvin is, Dolores will never marry him
      (ii) Handsome though it is said by many people that Melvin is, Dolores will never marry him

Ross (1967a, Chapter 6) shows that this rule involves "crucial use of a variable; the phrases can be moved up from indefinitely deep embeddings, subject to Ross's constraints" (cf. (ii)).

Chomsky listed the following constructions as exceptions to the upgrading conjecture ([our italics, square brackets and trace; DJ]):

(2)   a.   *The city* [VP was destroyed by [**THE ENEMY**]]
        b.   *There* [VP is [**A BOOK**] on the table]
        c.   The candidates *t* [VP have [**EACH**] indicated their willingness to debate]

He notes that "there are familiar analyses - not without competitors, but at least quite plausible - that postulate that in these case a rule of downgrading applies." (Chomsky 1975:107) The latter moves the capitalized constituent into the VP from a higher position external to it. To solve the resulting problem in a. and b., namely the fact that downgrading leaves a trace behind which has the status of a free variable not within the scope of a binding noun phrase or quantifier and therefore in violation of independent principles of anaphora, there is a later rule whose operation erases the offending trace. In a., the relevant rule is NP-preposing; in b. *there*-insertion. (In c. the problem does not even arise, according to Chomsky, since *each* is not a referential expression and its trace is not a bound variable, so that there is no antecedent-anaphor relation and, consequently, anaphora principles cannot be violated). The reason why we bring up these cases is that few people today would still adhere to a downgrading/ top-to-bottom analysis for any of these cases. In other words, there has been a gradual evolution in the direction of a pure bottom-to-top or raising perspective on movement. In the GB framework, deriving passives from actives by NP-postposing and subsequent object NP-preposing is no longer feasible: aside from the problems caused by structure-building processes and addition of the formatives *be* and *by*, there is the fact that NP-postposing followed by NP-preposing creates a violation of the Theta Criterion; and also the fact that the *by*-phrases of certain passives contain NPs which cannot become the subject of an equivalent active sentence, etc. For *there*-sentences, analyses with a base-generated postcopular indefinite subject have been around at least since the early seventies. The first were Jenkins' (1972, 1975) cleft reduction and complex NP proposals, followed by Stowell's (1978) small clause analysis which conceives of the postcopular material in existentials as one complex small clause constituent functioning as the comple-ment of *be*, so that the indefinite subject is never in the VP-external subject position and no downgrading has to be called upon. Stowell's proposal was taken up in Burzio (1981), Safir (1982), Lumsden (1983), and became a serious contender to Milsark's (1976) downgrading analysis , which was defended in Chomsky (1975; 1981:85ff[7]) but abandoned in *Knowledge of Language* (Chomsky 1986a; §4.3). The general acceptance of the VP-internal subject hypothesis[8] in the late eighties gave the deathblow to downgrading *there*-insertion analyses. This VP-internal subject hypothesis was introduced for reasons independent of *there*-insertion (Kitagawa 1986, Kuroda 1988), one of them the development of an analysis for floated quantifiers (*all, both*) (Sportiche 1988). All in all, there seems to be a marked evolution in the direction of a computational mechanism which does not cater for downgrading rules but is uniformly and uniquely bottom-to-top oriented.

---

[7] Later on in his *Lectures on Government and Binding* (1981:260-261), Chomsky's position evolves in the direction of postverbal base-generation of the indefinite contentive subject, however. In his analysis of PRO-drop in ergatives, he accepts Burzio's conclusion that ergative verbs in Burzio's sense "do not assign (or participate in assigning) a θ-role to subject position, but only to object position" and that given the θ-criterion, which stipulates that D-structure is a representation of GF-θ, it follows that the contentive subject in sentences with ergatives is "base-generated in the object position of the VP". Next, he states that there is an analogy between the rules of PRO-insertion and *there*-insertion, which leads to the (implicit) conclusion that at least in *there*-sentences with ergative verbs, the postverbal NP cannot be the result of NP-downgrading from VP-external subject position. In a note (p. 283:fn. 37) Chomsky adds that "Following Stowell (1978), Burzio argues that all cases of *there*-insertion are base-generated in the form (i):
        (i) [NP e] be NP..."
but he does not specify whether he is in agreement with this generalized base-generation hypothesis for *there*-sentences.

[8] The assumption that a thematic subject originates from a position internal to VP is in essence one of the subcases of Stowell's (1983) *Subjects across Categories* -hypothesis. Since Stowell's proposal antedates the proliferation of functional categories, however, he still considers S a major syntactic category in his article and treats the subject of a standard clause accordingly as base-generated and theta-marked in [NP, S (=IP)]. Consequently, the occurrence of VP-internal subjects is restricted to a very limited set of configurations, including such sentences as
        (i)   Mary had [VP her brother [open the door]]
        (ii)   We all feared [VP John [killed by the enemy]]

Within the generative semantics tradition, other proposals - invoking lowering rather than raising - were made, including transformations like Neg-placement, T-causative, T-modal, but the so-called 'explosion' or 'proliferation' of functional categories combined with the assumption that there can be phonologically empty formal feature movement in covert syntax, makes it possible to incorporate the insights of the generative semantics lowering proposals in such a way that rather than having lowering combined with deletion (and destruction) of a higher clause, the lower, visible constituent is Merge-generated in the position where one sees it, while, depending on the construction, either the whole category or its formal features move up to a higher (possibly empty) functional head position that also determines the scope of the operator in question. The main difference with the generative semantics analyses are (i) that no lowering technology has to be added to the theory for these cases, and (ii) that the higher scope-determining head-position is not destroyed, so that the two terms of the relationship remain intact. Empirically speaking, this bipolar conception without destruction of the higher pole seems to be a good alternative, as there is linguistic evidence that the higher scope-determining pole can show up at the surface in the form of a scope-marker. To briefly illustrate this point, we consider the following negatives in French and in substandard Dutch:

(3) a. je *ne* le savais *pas*
         I NEG it knew NOT
     b. 'k *(en)* wist et *nie*
         I (NEG) knew it NOT

In French, *pas* is introduced into the derivation as the Spec of a NegP in the periphery of the VP and in a Spec-Head relationship with a NEG-feature in the head of the NegP. The latter feature is projected and moved up to the head AgrS of the clause over which it has scope, where it is checked. The NEG feature is spelled out as *ne*. In the Dutch dialect example, the NEG feature moves into the equivalent scope-determining head position AgrS (and further into C, together with the other verb features that move there for verb second reasons); the checked NEG feature in the landing-site C can but need not be spelled out as *en*. In the latter case, there are no visible signs at the surface that there is anything more in the representation than the negative word *nie*. Yet, it would be very strange if the first NEG pole were not present at all: semantically we get exactly the same interpretation as in the construction with visible *en*, with exactly the same scope for the negative element, viz. over the whole clause. It seems then that a unified account with a featurally specified NEG in C-position and optional phonological realization of that feature is to be preferred over an analysis which treats the two options as structurally different while semantically identical.

3.2.2. Phrase-structure rules

The gradual disintegration and reorientation of the phrase structure rule system, for its part, occurred in the context of the abovementioned efforts to restrict the force of transformations. In *Remarks on Nominalization* (1970), Chomsky argued that transformations cannot perform derivational morphology, i.c. relate a verb and its nominalization. Yet, there are undeniable similarities between NPs headed by nominalizations and sentences built around a corresponding V, viz. the fact that grammatical relations are very similar in them and that the nominalization and the verb often have the same subcategorization statements. The solution he proposed to express the sharing of properties between the two - and cross-categorial regularities more generally - was to decompose lexical categories into two types of syntactic primitives: (1) a system of *syntactic features* (comparable to the use of distinctive features in phonology to express relations between segments): N = [+N,-V]; A = [+N,+V]; V = [-N,+V]; P = [-N,-V]; and (2) different *types* or levels: X^0, X', X''. This approach solved a problem first pointed out by John Lyons (1968:331), viz. that the Standard Theory of phrase structure rules was too unconstrained and allowed the generation of all kinds of wild phrases. All it said was that PS-rules have to be context-free rules of the form X → Y - Z; that they apply in a batch before the transformations; that they are unordered with respect to each other. Consider the following PS-rules and Lyons' (1968) reaction to them:

(4) a.   S → NP INFL VP

b. VP → V NP
c. VP → V
d. NP → Det N

> "The rules given above fail to formalize the fact that there is an essential, language-independent relationship between N and NP and between V and VP. As far as the formalization of phrase-structure grammars is concerned, it is a matter of 'accidental' coincidence that linguists will include in their grammars of different languages rules which always expand NP into a string of symbols containing N and rules which always expand VP into a string of symbols containing V. In other words, phrase structure grammars fail to formalize the fact that NP and VP are not merely mnemonically-convenient symbols, but stand for sentence-constituents which are *necessarily* nominal and verbal, respectively, because they have N and V as an obligatory major constituent." (Lyons 1968:331)

The latter property - that every major phrase is headed by a major lexical category which has the same categorial status as the phrase itself - goes by the well-known name of *endocentricity*. The category-neutral rewriting rules proposed in Chomsky (1970) and known as X'-theory overcome the 'wild phrases'-problem raised by Lyons and capture the traditional insight that major phrases are endocentric quite elegantly:

(5) a. X" → Specifier X'
    b. X' → X Complement

In his dissertation, Tim Stowell (1981:67) proposed to broaden the scope of this rule set to the functional category INFL, and in *Barriers* (1986b), Chomsky proposed that the X'-format be considered applicable to major/lexical categories (L-cats: N, V, A, P) and minor/functional categories (F-cats: INFL and C) alike. Jean-Yves Pollock's (1989) Split-INFL hypothesis, which introduced Tense and Agreement as separate new F-cats, and the further proliferation of F-cats in the late eighties have not altered that view. All of these new heads are assumed to conform to the same universal structural template. Though the most important evolution that Lyons' observation and Chomsky's category-neutral X'-template bears witness to is still masked by the fact that the new rule set is stated in the same form as the earlier rewriting rules, new terms such as 'head' and 'projection' leave no doubt as to which process has been seen in motion, viz. the gradual shift from a 'top-to-bottom' (henceforth ↓) system of rewriting rules to a 'bottom-to-top' (henceforth ↑) derivational construction of a phrase around a core lexical or functional category. That EST and GB have long held on to a mixed perspective with both top-down ↓-rewriting rules and from-the-bottom-up ↑-projection of syntactic features is evidenced by rule set 5. above (which can still be found in *Barriers* (Chomsky 1986b)) and is also clearly enunciated in the following quote:

> "First, every category of type [= bar-level; DJ] *i+1* has as an obligatory major constituent a category of type *i*: the latter is called the *head* of the former[9]. For example, the head of V', a type 1 category, is V, a type 0 category. Second, if *A* is the head of *B*, *A* and *B* have the same syntactic features. For example, both V and V' have the verbal feature matrix [+V, -N]. (...) Chomsky's Base Schema Hypothesis is a universal set of phrase structure rule schemata satisfying the first generalization and a convention for projecting the syntactic features of categories onto the categories of which they are the heads, which yields the second generalization." (Bresnan 1977:291).

3.2.3. The generalized transformation GT

With the advent of the generalized transformation GT in MPLT, the mixed perspective on tree-construction is given up[10]: syntactic feature projection and the transformational operation that forms structures now have the same upward orientation: ↑-projection and ↑GT.
Chomsky (1995) defines the core structure-building operation GT as a substitution operation "that takes a phrase marker K[1] and inserts it in a designated empty position Ø in a phrase marker K, forming the new phrase marker K*, which satisfies X-bar theory". He adds: "It targets K and substitutes K[1] for Ø in K." In

---

[9] This definition deviates from the more generally adopted one which restricts the term *head* to terminal elements, ie categories of type 0.
[10] At least, for substitution operations.

Bare Phrase Structure, there is an important change: Ø is no longer made use of. From the viewpoint of economy, all you have is the lexical items and their features. That is what can be drawn upon in the concatenative syntactic procedure GT (which is also a subpart of GT-Move) and any other object, whether a 'subliminal' part Ø of GT or anything else, is an addendum which it is better to do without if possible.[11] The definition of *move* in *Bare Phrase Structure* is therefore: "Given the phrase marker with terms K and , *move* targets K, raises , and merges with K to form the new category with the constituents , K."[12] This new perspective - which defines *merge* (a refined version of the operation GT) as a merger of two constituents, and *move* as an operation each application of which properly includes a *merge* application - establishes a firm link between the two operations. To put it differently, *move* consists of (i) Raising/Displacement of a constituent and (ii) *merge* of with the target K of the movement operation. This analysis sheds new light on an extremely important and influential contribution to the theory of syntax which was one of the first to tighten the relationship between phrase structure rules and transformations, viz. Joseph E. Emonds' (1976) theory of Root, Structure-Preserving and Local Transformations[13].

3.3. Deriving the structure preservingness hypothesis

The structure preservingness hypothesis states that structure-preserving transformations can move a category only to a position that is generated independently as a base expansion by the phrase structure rules. This idea would, however, not come out radically different if it did not require that the landing-sites *are* already pre-generated. Its spirit remains entirely intact if one requires that when a category moves, the movement operation can only let it merge with a target such that a structure is created that is compatible with the exigencies of the rule Merge. Interestingly, under the reformulation of *move* as a combination of displacement/raising and Merge, the Structure Preserving Hypothesis follows automatically. Since each application of *move* properly includes an application of *merge* at the target, the idea that landing sites are positions which could be generated by *merge* alone, follows by definition.

Such eventual compliance with *merge* requirements when *move* applies is the type of structure-preservingness that a radically derivational approach considers applicable to landing-sites. It is to be noted, that under such an approach the term preservation is slightly misleading, in that the operation does not really preserve existing structure, but builds new structure in accordance with Merge. Pre-generating an empty position - which is what follows from the idea that landing-sites are already pre-generated - is incompatible with strict derivationality and creates all kinds of problems. If - on plausible minimalist (and compositional) assumptions - all there is to work with is lexical items with their feature sets, pregenerating an empty position and its bar-level superstructure is impossible, because with the terminal position empty, it is impossible to project a designated bar-level category, say the N' needed as a landing-site for NP-movement (passive, raising) - from that empty category. Projecting a set of nominal formal features can only be done when those features are available, i.e. when the landing site in question is filled with a nominal lexical item containing the formal features required for projection. Pre-base-generating an N and projecting an N' *ex nihilo* is an instance of anticipatory generatio spontanea, which raises the question how one determines why the pre-generated category is N and not V or C or anything else and what the content of the notion "projection" could possibly be in this case. It seems logical to assume that creating something out of nothing is even more of a tour de force than the transmutation of categories.[14]

---

[11] Note that the definition of GT quoted in the text is problematic: it states that a phrase marker K[1] is inserted in a designated empty position Ø *in* a target phrase marker K. The problem with that definition is that embedding the designated empty category inside the target makes it impossible to exclude a second application of GT which targets exactly the same target as the first time and adds yet another constituent in K. Further down the page where this definition is given, Chomsky sets that problem right when he introduces the extension requirement. Under this qualification the designated empty position is added *outside* the targeted phrase marker, so that when the procedure is reapplied, the newly introduced designated empty position can never end up as a sister to the previous one.

[12] This implies that the substitution operation is on the way out. Substitution is: "a term borrowed from earlier theory that is now somewhat misleading, though we continue to use it." (Chomsky 1994:13)

[13] Eventually, root transformations were discovered to be structure-preserving ones too.

[14] Transmutation of categories is a term Koster (1987:67) uses in his analysis of the following topicalization case from Dutch:

These considerations lend further support to the idea that D-structure as a designated and pre-generated structure preceding lexical insertion and transformational operations is not a very plausible construct.

3.4. Recursion

Rules of the computational system of human language are not only characterised by construction-independence and the bottom-to-top property, but also by recursion. This capacity for building members of a category from members of the same category, is at the very heart of the generative procedure. Allowing a finite rule system or a part of it to reapply to its own output, entails that the procedure can be applied indefinitely to produce indefinitely long sentences. In view of the close relation between phrase structure rules and transformations we are trying to document, it is interesting to note that in the course of the history of generativism the recursive property has been assigned to both components. In *Syntactic Structures*, it was assigned to the generalized transformations; in the Standard Theory as expounded in *Aspects of the Theory of Syntax* (1965), however, Chomsky made them a feature of the base component, "in particular, of the rules that introduce the initial symbol S in designated positions in strings of category symbols." (Chomsky 1965:137) In other words, only S-recursion was allowed and it was taken care of by the phrase structure rules of the base component. Chomsky (1965: 225 fn.11) pointed out that the shift was not unproblematic as it left unexplained such phenomena as adjectival recursion, iteration of adverbials "and various kinds of parenthetic elements". The reasons why generalized transformations were thrown out and recursion reassigned are explained in Chomsky (1965:132-134): they are to do with the fact that there were "no known cases of ordering among generalized embedding transformations" (although such ordering was permitted by the theory of T-markers), nor "singular transformations that must apply to the matrix sentence before a sentence transform is embedded in it." Given the different outlook of the current model, the main objections against generalized transformations (and parallel generation of tree chunks, of partial phrase-markers) have disappeared, so that the claim that recursion is a property of *move* as much as of *merge* is no longer objectionable. Both *merge* and *move* are operations which can be called upon in the course of the derivation to apply and reapply as the need arises. To our knowledge, no competence constraint has ever been proposed limiting the number of (re)applications of these operations (nor on the possible number of links in a chain).

4. Conclusion

Efforts to constrain transformational operations and to reduce overlap between the lexicon and the phrase structure component triggered a process of gradual unification of transformations and rewriting rules. Both of them are currently considered to be construction-independent, structure-preserving, recursive rule types, operating in a strictly bottom-to-top manner. Though the return to a closer association between them brings back certain ideas floating around in the early days of generativism, the new picture is much more insightful and detailed than any earlier conceptions of the two rule types. While several questions that bedevilled the generative pioneers are still not fully resolved today, we are on much more solid theoretical ground to confront them and have doubtless learned a great deal about the architecture of the human language faculty along the road.

References

Bloomfield L., *Language*. New York 1933.
Bresnan J., "Transformations and categories in syntax". In: Butts R., Hintikka J. (Eds.), *Logic, Methodology and Philosophy of Science*. Dordrecht 1977: 283-304.

(i)   [$_{AP}$ Tevreden over zichzelf$_i$]$_j$ [$_{CP}$ *dat*$_j$ [is hij$_i$ nooit t$_j$ geweest]]
        (satisfied with himself that$_j$ has he never t$_j$ been)
        Satisfied with himself, he has never been

The so-called d-word *dat* in the operator position of the embedded clause is an NP, the topicalized phrase *tevreden over zichzelf* an AP. This structure, Koster claims, is strong counterevidence against a Vergnaud-type analysis for topicalization. Vergnaud-raising would amount to the transmutation of categories, i.c. of NP into AP, which Koster denounces by stating that "the transmutation of water into wine is more credible". If anything, projecting a category from , i.e. from nothing, is an even stronger feat than transmutation.

Burzio L., *Intransitive verbs and Italian auxiliaries*. PhD diss., Cambridge/Mass 1981.

Chomsky N., *Syntactic structures*. The Hague 1957.

Chomsky N., *Current Issues in Linguistic Theory*. The Hague 1964.

Chomsky N., *Aspects Of The Theory Of Syntax*. Cambridge/Mass. 1965.

Chomsky N., *Cartesian linguistics*. New York/London 1966.

Chomsky N., "Remarks On Nominalization". In: Jacobs R.A., Rosenbaum P.S., *English Transformational Grammar*, 1970: 184-221.

Chomsky N., *Reflections on Language*. London 1975.

Chomsky N., *Lectures on Government and Binding*. Dordrecht 1981.

Chomsky N., *Knowledge Of Language: Its Nature, Origins And Use*. New York 1986a.

Chomsky N., *Barriers*. Cambridge/Mass.1986b.

Chomsky N., *A Minimalist Program For Linguistic Theory (MPLT)*. In: Hale K., Keyser J. (Eds.), *The View From Building 20 - Essays In Linguistics In Honor Of Sylvain Bromberger*. Cambridge/Mass. 1993.

Chomsky N., *Bare Phrase Structure*. MIT occasional papers in linguistics 5, Cambridge/Mass. 1994.

Chomsky N., *The Minimalist Program*. Cambridge/Mass. 1995.

Chomsky N., Lasnik H., "The theory of principles and parameters". In: Jacobs, J., von Stechow, A., Sternefeld, W., Venneman, T. (Eds.), *Syntax: An International Handbook Of Contemporary Research*. Berlin 1993.

Emonds, J.E., *A Transformational Approach To English Syntax: Root, Structure-Preserving, And Local Transformations*, New York 1976.

Harris Z., *Methods In Structural Linguistics*. Chicago 1951.

Humboldt W. von, *Über die Verschiedenheit des Menschlichen Sprachbaues*. Berlin 1836.

Jackendoff R.S., *Semantic Interpretation In Generative Grammar*. Cambridge/Mass. 1972.

Jackendoff R.S., *Semantics And Cognition*. Cambridge/Mass. 1983.

Jenkins L., *Modality in English syntax*. PhD diss., Cambridge/Mass. 1972.

Jenkins L., *The English Existential*. Tübingen 1975.

Jespersen O., *The Philosophy Of Grammar*. London 1924.

Kitagawa Y., *Subjects In Japanese And English*. PhD diss., Amherst 1986.

Koster J., *Domains And Dynasties. The Radical Autonomy Of Syntax*. Dordrecht 1987.

Kuroda S.-Y., "Whether We Agree Or Not: A Comparative Syntax Of English And Japanese". In: Poser, W. (Ed.), *Papers From The Second International Workshop On Japanese Syntax*. Stanford 1988.

Lumsden, *Syntactic And Semantic Properties Of Existential Sentences*. Phd diss., York 1983.

Lyons J., *Introduction To Theoretical Linguistics*. Cambridge 1968.

Milsark G., *Existential Sentences In English*. PhD diss., Cambridge/Mass. 1976.

Pollock J.-Y., "Verb Movement, Universal Grammar And The Structure Of IP". In: *Linguistic inquiry 20* (1989) 365-424.

Ross J.R., *Constraints On Variables In Syntax*. PhD diss., Cambridge/Mass. 1967.

Safir K., *Syntactic Chains And The Definiteness Effect*. PhD diss., Cambridge/Mass. 1982.

Seuren P.A.M., *Filosofie van de taalwetenschappen*. Leiden 1990.

Seuren P.A.M., *Semantic Syntax*. Oxford 1996.

Seuren P.A.M. (Ed.), *Semantic Syntax*. Oxford 1974.

Seuren P.A.M., *Western Linguistics. An Historical Introduction*. Oxford 1998.

Sportiche D., "A Theory Of Floating Quantifiers And Its Corollaries For Constituent Structure". In: *Linguistic inquiry 19* (1988): 425-449.

Stowell T., "What Was There Before There Was There". In: *Papers from the fourteenth regional meeting of the Chicago Linguistic Society* 1978.

Stowell T., *Origins Of Phrase Structure*. PhD diss., Cambridge/Mass. 1981.

Stowell T., "Subjects Across Categories". In: *The Linguistic Review 2* (1983) 285-312.

Webelhuth G. (ed.), *Government And Binding Theory And The Minimalist Program: Principles And Parameters In Syntactic Theory*. Oxford 1995.