(draft)

Variable-free Grammar

Jan Koster
University of Groningen

1. Introduction: constraints on variables

Since the beginning of transformational grammar in the 1950s, its transformational rules were formulated with variables. Thus, in Chomsky (1957: 69) the rule of Wh-movement has a structural description as in (1a), with the two variables X and Y,

(1) X -- NP -- Y

where the NP is later on transformed into a Wh-phrase. Wh-movement was described as movement of the NP across the variable X to its left:

(2) NP -- X -- Y

Variables as X stood for arbitrary (possibly zero) portions of the affected structure. Since it was clear that such variables were not entirely arbitrary, much effort of early transformational grammar went into the formulation of "constraints on variables", as in Ross's classical dissertation of (1967).

In practically all "conditions on rules" –the focus of linguistic theorizing during the next few decades-- such variables were preserved, for instance in the formulation of Subjacency in Chomsky (1973). Also my own most recent formulation of such conditions, the Configuaritional Matrix (Koster 1987, 1999), maintains the traditional variables, as indicated by the dots:

(3) **The Configurational Matrix**

All (local) grammatical relations have the following form:

$$[_\beta \ldots\alpha\ldots\delta\ldots]$$

where:

a. α is an antecedent
b. δ is an element depending on α
c. β is some minimal domain
d. α and δ freely share their properties

According to this formulation, α and δ *can* be adjacent, for instance in head complement structures:

(4) $[_{VP}$ V DP]

(where α = V, δ = DP and β = VP). This expresses the idea that head and complement are adjacent in underlying structures (see Chomsky 1986a).

The same Configurational Matrix categorizes "filler-gap" constructions as those resulting from Wh-movement:

(5) $[_{CP}$ *what* did you see [*t*]]

In this structure, the value of β is CP and the antecedent (= α) *what* is not adjacent to the dependent element [t] (= δ). So, in order to see (4) and (5) as realizations of the same pattern, a formulation of (3) with the traditional variables seemed necessary.

What I would like to propose in this article is that variables can be eliminated and that, therefore, there is no problem as to what are the "constraints of variables" in the sense of Ross (1967) and subsequent work by others. Instead, I would like to claim that the proper reformulation of (3) is variable-free. In other words, I hypothesize that all core grammatical relations in all languages are characterized by the following formula:

(6) **Law of Grammar**

Grammatical core relations universally have the form:

$$[_\beta\ \alpha\ \ \delta\ ]$$

This formulation preserves the significant empirical generalizations of the Configurational Matrix, namely:

(7) a. α precedes δ
b. bi-uniqueness: one α for δ and one δ for α
c. bilocality (replaces c-command and locality)
d. recursion: both α and δ can be a β

Much of the empirical scope of the Configurational Matrix was illustrated in previous work (for instance, Koster 1987, 1999) and it is much in tune with the empirical generalizations made by Kayne (1994).

I will illustrate "precedence" (7a) in somewhat more detail in the next section, but here I will only briefly touch upon the other aspects, referring the reader to the relevant literature.

Recursion (7d) is the least controversial property, since it is generally agreed upon that syntactic structures are recursive.

Bilocality means that the locality conditions are the same for antecedent and dependent element. Standard locality principles (like Subjacency) define the minimal domain β in which a dependent element δ must find an antecedent α. In Koster (1987, 1999) it was concluded that c-command can be replaced by similar locality conditions defined on α rather than on δ.

Bi-uniqueness is a less well-known property of grammar, but in general one seems to find one-one relations between antecedents and dependent elements. This determines the binary-branching nature of phrase structure and also –I assume-- the fact that there can only be one Spec to a given head. Also the theta-criterion seems to follow from the bi-uniqueness property of (6).

2. Precedence

Discussing the Q-Universal of Baker (1970) and its formulation in Bresnan (1970), Chomsky (1973: 234) makes the following interesting comment:

> In particular, "*wh*-words"—the relativized constituents in relative clauses or questioned constituents in interrogatives—can be moved only to the left, such movement being permitted only when there is an initial COMP in the phrase to which the transformation is being applied.

It has been known, in other words, for at least 30 years that Wh-movement is to the left. Kayne (1994) generalized this conclusion to all movement operations in connection with his idea that phrase structure is universally head-initial. This has led to a view of grammar according to which *lexical* projections are always embedded in a shell of equally head-initial *functional* projections. Since constituents are moved to check features in the functional projections, movement can only be to the left.

Recently, I have concluded that a major class of traditional rightward movements, the class of extrapositions, is in fact spurious (Koster 2000b). If this conclusion is correct, there are hardly any convincing examples left of rightward movement and we could considerably simplify the class of possible grammars.

Similar progress seems possible with respect to the question whether all languages have head-initial phrase structure. From Greenberg's (1966) survey of word order types, it could already be concluded that SOV, SVO and VSO are the main word orders found in the languages of his sample. Since then, VSO has been eliminated as a possible base structure for reasons of head-complement adjacency (VO or OV) and by empirical arguments that the languages in question have rules moving the verb across the subject.

So far, then, a plausible case could be made that VO and OV are the only remaining possible base orders. A theory like the one found in Kayne (1994) makes the further claim that VO (with initial V-head) is the only possibility permitted by Universal Grammar. This is also entailed by universal accounts of syntactic structure like (3) and (6) above.

The idea of universal head-intial phrase structure was confirmed by Zwart's successful reanalysis of Dutch as an underlying VO language (1993, 1994). Zwart's conclusions were in part derived independently of Kayne's work and were the more remarkable since it had been believed for more than 20 years that the level of OV order (preserved in Zwart's analysis as an intermediate level) was the deepest word order of Dutch (see Koster 1975). The new VO analysis of Dutch solves certain traditional word order problems (such as the interaction between verb cluster formation and "extraposition") and is the only analysis so far that explains why there is a connection between (the now derived) OV order and certain scrambling phenomena.

Unlike English, for instance, Dutch does not show verb-object adjacency and the following order is very natural:

(8) ...dat hij *het boek* gisteren waarschijnlijk *gelezen* heeft
that he the book yesterday probably read has
"that he probably read the book yesterday"

The object is separated from the verb by VP-external adverbials, which indicates that the object has been moved to the left (namely by Vanden Wyngaerd's (1989) rule of Object Shift).

Just stipulating that Dutch is OV does not explain this correlation between OV order and scrambling with respect to adverbials. In general, lack of head-complement adjacency is an indication that movement has taken place. Languages with OV order usually have no adjacency, while adjacency is usually better preserved in the VO languages (although strict adjacency can be broken by "heavy NP shift" or by verb movements, as in the Romance languages).

Recently, deeper insight has been obtained in how the word order of a VO language differs from a language like Dutch, which shows OV order in subordinate clauses. Both languages must have their objects checked in the relevant functional projections to the left of the VP, but whereas Dutch and German check their DPs individually, English checks them collectively by moving the whole VP to the positions for Case checking (Koster 1999, 2000a).

Apart from the empirical gains, this pattern takes away an important objection against earlier versions of the VO analysis of Dutch and German. The objection used to be that it was just stipulated that the Dutch and German objects were moved (with "strong" features to be checked), while in English the objects remained *in situ*, possibly followed by checking at LF ("weak" features). The strong-weak distinction did not seem to make sense in this context.

Under the new analysis, *all* checking is done by overt movement. The only difference is a difference in how the size of the checking phrases is defined (Pied Piping). Pied Piping differences are a very common form of language variation. Also in recent analyses of extraposition phenomena, Pied Piping played a very important role (Koster 2000b). As I will argue further in the next section, Pied Piping is by far the most important mechanism for "travelling" long distances in grammar.

The successful reanalysis of a traditional OV language like Dutch to a language conforming to the universal VO pattern entailed by (3) and (6) can hopefully in one way or another be applied to other traditional OV languages as well. Given the massive empirical support for the order of (3) and (6) in a wide range of construction types, the null hypothesis is that all languages are underlyingly VO. The underlying pattern is "distorted" by movements for checking, such as the V-movements in VSO languages and the object shifts in Dutch and German. Languages that check with the whole VP, like English, preserve the universal base order VO (unless further movements take place).

To get an idea of the generality of "precedence" of the antecedent α with respect to the dependent element δ (as in (3) and (6)), consider the following list (which could easily be expanded):

(9) a. movement
b. base structures
c. gapping
d. anaphora

Movement and base structure were already discussed: in movements, the moved element precedes the gap (the trace) and in base structures, the head precedes the compliment. But in many other construction types we find the same.

Consider gapping. It is a local phenomenon fully meeting the conditions as stated in (3) and (6) (see Koster 1987). The antecedent always precedes the gap:

(10) John *saw* Mary and Peter --- Sue

The gap (interpreted as *saw*) cannot precede the verb:

(11) *John --- Mary and Peter *saw* Sue

The same can be said about anaphora: in underlying structures the antecedent always precedes the anaphor:

(12) a. We want *John* to describe *himself*
b. *We want *himself* to describe *John*

As soon as we see the opposite order, in which an anaphor precedes the antecent, we know that some reordering has taken place:

(13) [To describe *himself*]$_i$ *John* does not really want t_i

The anaphor *himself* precedes the antecedent *John* thanks to the preposing of the infinitival complement.

All in all, precedence of α with respect to δ seems so wide-spread that we can safely conjecture that the order as stated in (3) and (6) is universal and that all deviances (like OV) involve some form of reordering, with or without Pied Piping.

What we have not solved yet is how we come from a formulation with a variable (3) (indicated by the dots) to the variable-free and optimally elegant formulation (6) (without any dots). It is to this problem that we will turn next. As in the case of extraposition and the word order differences between English and Dutch, Pied Piping happens to be the key.

3. The elimination of variables

Recall that variables seemed to be necessary to make both movement and base structures fit the Configurational Matrix. Head-complement structures showed strict adjacency of α and δ (14a), but Wh-movement usually shows a certain distance between moved element and trace (14b):

(14) a. [$_{VP}$ V DP]

b. [$_{CP}$ Wh$_i$ .... t_i] (cf. example (5))

Both are instances of (3) and both are in accordance with the properties listed in (7). However, the variable (dots) is necessary for (14b) and not for (14a) with its strict adjacency, indicating that the unification is not complete.

In order to see how we can establish full unification, we have to consider for a while how long distances are bridged in grammar. A standard way to connect elements over longer distances is the operation Move (as, for instance, it has been applied in (14b)). However, Move has always been suspect in that it creates outputs of the same type as those of the base rules (now seen as Merge). This is, of course, what was known as structure-preservingness (Emonds 1976). Chomsky (1995: 318) eliminates the structure-preserving hypothesis and says that it is in fact unformulable in the minimalist framework.

This might also indicate that something is not quite right with the formulation of the minimalist framework because the original, empirical problem remains, namely that Move produces structures of the same kind as Merge (see also Kitahara 1997).

Another reason why Move is suspect, on which I will focus here, is that the distances it bridges are also bridged by Merge. If you built up a CP with a Wh-phrase in its COMP, you start the merging process with, for instance, the V and its object. Successive applications of Merge automatically lead to the CP and its Spec (COMP). In other words, something seems to be redundant.

Interestingly, it is implicitly assumed that there is a third mechanism to bridge long distances, namely Pied Piping. Pied Piping carries certain features beyond its minimal phrase:

(15) $[_{PP}$ With $[_{DP}$ the brother $[_{PP}$ of $[_{DP}$ *which* girl$]]]]_i$ did you talk t_i

The fronted phrase is a Wh-phrase moved to check the features of the [+wh] head of the CP. In order to move the phrase in question, Wh-phrases must be defined somehow. Pied Piping is interesting because much larger phrases are moved than the minimally necessary Wh-phrase: *which* in the most deeply embedded DP in (15). It bridges a fairly long distance in (15), namely from the most deeply embedded DP to the most inclusive PP (the actual checking phrase).

How are Wh-phrases and their size defined? Unfortunately, this matter has been left largely implicit. There has always been a lot of informal reference to "percolation" and there have even been explicit definitions of percolation paths in a slightly different context (the g-projections of Kayne 1983). However, a systematic and explicit account of percolation phenomena is still largely a matter of future research. In fact, recent research indicates that Pied Piping is a much more common phenomenon than realized so far (see Koster 1999, 2000a, b).

In this article, however, I will mainly focus on the fact that Pied Piping (percolation) is a third way to bridge long distances, adding to the redundancy already implied by the coexistence of Move and Merge.

More concretely, I would like to propose that Pied Piping phenomena can be accounted for by a slight extension of the operation Merge. In doing so, we arrive at (6) which can be interpreted as a full unification of Merge and the Configurational Matrix. The resulting theory will have only one mechanism to bridge long distances instead of three, namely percolation in accordance with (6). If this is correct, (6) accounts for the properties of both phrase structure (Merge) and chains (Move), but also for Gapping and Pied Piping phenomena. This unification is possible by combining (6) with a set of filters, which are defined strictly in terms of the local notions of (6) itself. This eliminates the variables of earlier transformations and conditions on rules.

In order to see how Merge can be extended to also cover Pied Piping, movement phenomena, gapping and all other phenomena covered by the Configurational Matrix, we have to have a closer look at how Merge is defined in Chomsky (1995, ch. 4). Merge applies to two objects, α and β, creating a new object K (*op.cit*. p. 243):

(16) $K = \{\gamma, \{\alpha, \beta\}\}$, where α, β are objects and γ is the label of K

Note that, apart from linear order, (16) defines objects that are already very close to being instances of (6), because the β of (6) can be interpreted as the label (γ) of an operation merging α and δ in (6).

The problematic part of Merge and its bare phrase structure interpretation concerns the following options for γ listed by Chomsky (*op.cit*. p. 244):

(17) a. the intersection of α and β
b. the union of α and β
c. one or the other of α, β

Chomsky rightly rejects (17a) and (17b), but from that it does not follow that (17c) is correct as Chomsky concludes, the reason being that (17) is too narrow a range of options. According to Chomsky, only α or β can be the label, so that they *project* as the head of K. Thus, with α as label, K is interpreted as follows:

(18) $K = \{\alpha, \{\alpha, \beta\}\}$

Chomsky further concludes that no additional elements enter into projections (p. 245).

This can only be correct, however, if we strictly limit ourselves to what is traditionally seen as the projection of a head. From a broader perspective, projection is just a subcase of Pied Piping: the mechanism that percolates features up to more inclusive categories. As soon as we realize this, it is clear that (17) is too narrow a range of options for upward percolation. A logical possibility not considered by Chomsky is that the label γ in (16) and (17) is a *subset* of the union of α and β (17b).

The core of my unification proposal is just this, namely that *the label of Merge is a subset of the union of α and β*. Which subset is a matter of strictly local filters. If we limit ourselves to projection in the narrow sense, we can only agree with Chomsky, but very often Merge transfers additional properties to the label. Consider a simple case of Pied Piping:

(19) $[_{PP}$ $[_{P}$ with] $[_{NP}$ *whom*]]

In this example, the original objects α and β are *with* and *whom*. Under Merge, a new object K is created with label *with* (indicated by the PP in (19) for ease of exposition). Thus, only the head projects, in accordance with Chomsky's proposal.

However, something more seems to be transferred to the label, namely the Wh-properties of *whom*: the whole PP qualifies as a Wh-phrase for feature checking. In other words, not only the head projects its features but, at least partially, also the complement sometimes. The mechanism looks exactly the same: strictly local transfer of properties, i.e., to the immediately dominating node. It is all Pied Piping and the differences are a matter of filters: Wh-features potentially percolate further up than head features. Head features percolate as long as a head projection is merged with a non-head. As soon as a new lexical head appears, this new head projects rather than the old one.

Wh-features, in contrast, percolate beyond minimal head projections, as shown by (15) and (19). Thus, if a Wh-phrase is merged with a new lexical head, its features may still percolate, as long as the new lexical head is of a certain type. In Dutch or English, for instance, N and P heads permit further percolation (as in (15) and (19)), while a new V and its functional projections block further percolation (in Dutch, but not always in German). The exact nature of percolation filters is far from simple and will be left for further research here. In general, I agree with Chomsky (1995: 264) that constraints on Pied Piping are not all that different from the more traditional conditions on movement. CPs, for instance, are almost always barriers for Pied Piping.

However, as mentioned above, my proposal rejects the variables of earlier conditions on rules and seeks to formulate the constraints in a strictly local way, as conditions on percolation involving no other elements than two adjacent terms and their immediately dominating category (as in (6)).

I will now show how Move can be reduced to the same mechanism, under elimination of the traditional variables. What we learned form the percolation of Wh-features is that features of a non-head can be percolated. What can be done with Wh-features can also be done with gaps, as was in fact already proposed by Gazdar (1981). Critical assessment of Gazdar's work focused on his claims about the relevance of having context-free grammars for natural languages. Assuming that Chomsky was right in rejecting the relevance of this notion for the learnability problem, we nevertheless see a reason in the present context to return to Gazdar's formalization of gap percolation, which has in one form or another become normal in the variant of generative grammar known as HPSG (see for instance Bouma *et al*. 1999).

According to Gazdar, the presence of a gap could be indicated by a slashed category and transferred to the successive categories higher up. Thus, an NP gap (a "trace" in standard generative grammar) could be indicated by /NP (in NP/NP) and /NP could be inherited by the next category up, etc.:

(20) a. [NP *Who*] [$_{IP/NP}$ did you [$_{VP/NP}$ see NP/NP]]]

The presence of the gap is signalled on the successivley more inclusive categories VP and IP, as indicated by the slash notation.

From the current point of view, this is nothing other than Pied Piping again, i.e., certain properties of a category are transferred to successively more inclusive categories, just as in the case of the formation of Wh-phrases. Thus, we might say that Pied Piping for Wh-features creates *Wh-phrases*, whereas Pied Piping for gaps creates *Gap phrases*. As before, the percolation of gap features is not unrestricted. In the unmarked case, it does not extend beyond minimal lexical projections and their functional extensions (NP, PP, AP, CP; see Koster 1987 for details).

In other words, the traditional island conditions can be seen as filters on the percolation mechanism (Pied Piping) for gaps. Unlike in the earlier island conditions, the percolation and filtering mechanism can be formulated without variables. Each percolation "decision" is strictly local and can be entirely limited to the contexts defined by (6). In Dutch, for instance, PPs are islands (Van Riemsdijk 1978), which means that the following structure (an instantiation of (6)) is not well-formed and has to be filtered out (*met* means "with"):

(21) *[$_{PP/NP}$ met NP/NP]

If gap phrases can be defined in exactly the same way as Wh-phrases (but with slightly different filters), we can fully eliminate variables from the Configurational Matrix (3) and reformulate it as (6). A situation like (22a) (= 14b), for instance, would never be considered, but instead we would only have configurations as in (22b):

(22) a. $[_{CP}$ Wh_i .... t_i]

b. $[_{CP}$ [Wh-phrase] [Gap phrase]]

Thanks to percolation of the gap features, satisfaction (of the gap by the Wh-filler) can be determined at a strictly local basis, i.e. by only considering *adjacent* terms, just as in the case of head-complement structures (cf. 14a).

4. Conclusion

Originally, transformations and conditions on rules were formulated with variables, to account for the fact that rules connected variable portions of structure. Many conditions on rules could be characterized as locality conditions. These locality conditions showed a lot of overlap and redundancy and at the same time indicated that portions of trees affected by rules are not really variable but very local in some sense.

The redundancy was maintained in recent accounts by the coexistence of both Move and Merge and a third –more implicit—mechanism to bridge distances, namely feature percolation (Pied Piping). Also my original formulation of the Configurational Matrix (3) preserves the variables of earlier theories, thereby masking the fact that the core relations of grammar are strictly local.

As recent work seems to indicate, the most neglected mechanism, Pied Piping, is in fact the key to a solution. The puzzle of redundancy and diverse and variable locality can be solved by building Pied Piping into Merge. This could be done in a very simple way, namely be slightly expanding the class of projecting elements from only head features to a set that selectively also allows Wh-features and gap features. In this way, Merge could be completely unified with the Configurational Matrix (3), thereby inheriting the full empirical coverage of the latter (see Koster 1987). Thus, the resulting Law of Grammar (6) defines the nature of structures as diverse as “X-bar structures”, chains created by “move alpha” and gapping constructions. Particularly the operation Move is completely redundant under the proposed theory, while Chomsky’s original intuition of the strict cycle is maintained in some sense in the strictly local and successive application of Merge.

The proposed Law of Grammar does not only differ from earlier versions of Merge by its allowance of a richer set of percolated features, it also stipulates a strict linear order of antecedent and dependent element. Merge in its standard formulation abstracts away from this linear order. It is always possible that linear order is imposed by external demands (phonology, parsing, etc.; see for instance Neeleman and Van de Koot 1999) but there can be little doubt that it is an overwhelming empirical reality. Whether it must be eliminated as a fundamental part of the Law of Grammar remains to be seen.

Bibliography

Baker, Carl Leroy. 1970. Notes on the Description of English Questions: The Role of an Abstract Question Morpheme. *Foundations of Language* 6, 197-219.
Bouma, Gosse, Robert Malouf & Ivan Sag .1999. Satisfying Constraints on Extraction and Adjunction. Ms., University of Groningen, Stanford University. To appear in: *Natural Language and Linguistic Theory*.
Bresnan, Joan. 1970. On Complementizers: Towards a Syntactic Theory of Complement Types. *Foundations of Language* 6, 297-321.
Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.
Chomsky, Noam. 1973. Conditions on Transformations. In: S. Anderson & P. Kiparsky, eds., *A Festschrift for Morris Halle*. New York: Holt, Rinehart and Winston.
Chomsky, Noam. 1986a. *Knowledge of Language*. New York: Praeger.
Chomsky, Noam. 1986b. *Barriers*. Cambridge, Mass.: MIT Press.
Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, Mass.: MIT Press.
Emonds, Joseph. 1976. *A Transformational Approach to English Syntax: Root, Structure-Preserving and Local Transformations*, Academic Press, New York.
Gazdar, Gerald. 1981. Unbounded Dependencies and Coordinate Structure. *Linguistic Inquiry* 12, 155-184.
Greenberg, Joseph. 1966. Some Universals of Grammar with Particular Reference to the Order of Meaningful Elements. In: Jospeh Greenberg, ed., *Universals of Grammar*. Cambridge, Mass.: MIT Press.
Kayne, Richard. 1984. *Connectedness and Binary Branching*. Dordrecht: Foris.
Kayne, Richard. 1994. *The Antisymmetry of Syntax*. Cambridge, Mass.: MIT Press.
Kitahara, Hisatsugu. 1997. *Elementary Operations and Optimal Derivations*. Cambridge, Mass.: MIT Press.
Koster, Jan. 1975. Dutch as an SOV language. *Linguistic Analysis* 1, 111-136.
Koster, Jan. 1987. *Domains and Dynasties*. Dordrecht: Foris.
Koster, Jan. 1999. The Word Orders of English and Dutch: Collective vs. Individual Checking. In: Werner Abraham, eds, *Groninger Arbeiten zur germanistischen Linguistik*, University of Groningen, Groningen, 1999, 1-42.
Koster, Jan. 2000a. Pied Piping and The Word Orders of English and Dutch. To appear in: NELS 30.
Koster, Jan. 2000b. Extraposition as Parralel Construal. Ms., University of Groningen, Groningen.
Neeleman, Ad and Hans van de Koot. 1999. The Configurational Matrix. Ms., University College London, London.
Riemsdijk, H. van. 1978. *A Case study of Syntactic Markedness.* Dordrecht: Foris.
Ross, John Robert. 1967. *Constraints on Variables in Syntax*. PhD Diss., MIT, Cambridge, Mass.

Vanden Wyngaerd, Guido. 1989. Object Shift as an A-movement Rule. *MIT Working Papers in Linguistics* 11, 256-271.
Zwart, Jan-Wouter. 1993. *Dutch Syntax: A Minimalist Approach*. PhD Diss., University of Groningen, Groningen.
Zwart, Jan-Wouter. 1994. Dutch is Head Initial. *The Linguistic Review* 11, 377-406.