Move, Merge and Percolate are One!

On the Elimination of Variables in Grammar

Jan Koster, University of Groningen

1. 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 like 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 more recent formulation including that kind of condition, the Configurational Matrix (Koster 1987, 1999), maintains the traditional variables.

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 on variables" in the sense of Ross (1967) and subsequent work by others. Instead, I would like to claim that the proper reformulation of conditions on rules and/or representations is variable-free. In other words, I hypothesize that all core grammatical relations in all languages are characterized by the following formula (δ a category dependent on an antecedent α in a minimal domain β):

(3) **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:

(4) a. α precedes δ
    b. bi-uniqueness: one α for δ and one δ for α
    c. bilocality (covers 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).

Precedence (4a) entails that all movement is to the left and that all phrase structure is underlyingly head-initial (cf. Kayne 1994). For Dutch and German, this was convincingly demonstrated by Zwart (1993, 1994).

Recursion (4d) 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 (3).

2. 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 (5a), but Wh-movement usually shows a certain distance between moved element and trace (5b), as indicated by the dots:

(5) a. $[_{VP}$ V DP $]$
    b. $[_{CP}$ Wh_i .... t_i $]$

Both are instances of (3) and both are in accordance with the properties listed in (4). However, the variable (dots) has been supposed to be necessary for (5b) and not for (5a) 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 (5b)). 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 cannot be formulated in the minimalist framework.

This might, in fact, 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:

(6) $[_{PP}$ 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 (6). It bridges a fairly long distance in (6), 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 1984). 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 limit myself to the fact that Pied Piping ("Percolate") 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 (3) 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 (3). If this is correct, (3) accounts for the properties of both phrase structure (Merge) and chains (Move), but also for Gapping (further ignored here) and Pied Piping phenomena. This unification is possible by combining (3) with a set of filters, which are defined strictly in terms of the local notions of (3) 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):

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

Note that, apart from linear order, (7) defines objects that are already very close to being instances of (3), because the β of (3) can be interpreted as the label (γ) of an operation merging α and δ in (3).

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

(8) 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 (8a) and (8b), but from that it does not follow that (8c) is correct as Chomsky concludes, the reason being that (8) 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:

(9) 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 (8) is too narrow a range of options for upward percolation. A logical possibility not considered by Chomsky is that the label γ in (7) and (8) is a *subset* of the union of α and β (8b).

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:

(10) $[_{PP}\ [_{P}$ 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 (10) 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 (6) and (10). 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 (6) and (10)), while a new V and its functional projections block further percolation (in standard 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 (3)).

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*. 2001).

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.:

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

The presence of the gap is signalled on the successively 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 (3). In Dutch, for instance, PPs are islands (Van Riemsdijk 1978), which means that the following structure (an instantiation of (3)) is not well-formed and has to be filtered out (*met* means "with"):

(12) *[$_{PP/NP}$ 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 and reformulate it as (3). A situation like (13a) (= 11b), for instance, would never be considered, but instead we would only have configurations as in (13b):

(13) 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. 5a). In other words, Universal Grammar specifies only one mechanism, successive Merge, to bridge long distances rather than the traditional three (Move, Merge and Percolate). Formally, generalized Merge has a form defined by the Law of Grammar given in (3), supplemented with strictly local filters as to the subset of features actually transferred to the next level up.

References

Bouma, Gosse, Robert Malouf & Ivan Sag. 2001. Satisfying Constraints on Extraction and Adjunction. *Natural Language and Linguistic Theory* 19: 1-65.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1973. Conditions on Transformations. In *A Festschrift for Morris Halle*, ed. S. Anderson & P. Kiparsky. New York: Holt, Rinehart and Winston.

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*. New York: Academic Press.

Gazdar, Gerald. 1981. Unbounded Dependencies and Coordinate Structure. *Linguistic Inquiry* 12: 155-184.

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. 1987. *Domains and Dynasties*. Dordrecht: Foris.

Koster, Jan. 1999. The Word Orders of English and Dutch: Collective vs. Individual Checking. In *Groninger Arbeiten zur germanistischen Linguistik*, ed. Werner Abraham, University of Groningen, 1999: 1-42.

Koster, Jan. 2000a. Pied Piping and The Word Orders of English and Dutch. In *NELS 30: Proceedings of The North East Linguistic Society*, ed. M. Hirotani, A. Coetzee, N. Hall and J.-Y. Kim, 415-426. Amherst, Mass.: GSLA.

Koster, Jan. 2000b. Extraposition as Parallel Construal. Ms., University of Groningen.

Riemsdijk, Henk van. 1978. *A Case Study of Syntactic Markedness.* Dordrecht: Foris.

Ross, John Robert. 1967. *Constraints on Variables in Syntax*. PhD dissertation, MIT.

Zwart, Jan-Wouter. 1993. *Dutch Syntax: A Minimalist Approach*. PhD dissertation, University of Groningen.

Zwart, Jan-Wouter. 1994. Dutch is Head Initial. *The Linguistic Review* 11: 377-406.

Jan Koster
Department of Linguistics
University of Groningen
P.O. Box 716
9700 AS Groningen
The Netherlands
koster@let.rug.nl