OSU WPL 32.125-132 (1986)

Free Word Order in GPSG*

Arnold M. Zwicky
Ohio State University and Stanford University

1. **Free constituent order and free word order.** Recent versions of generalized phrase structure grammar (following Gazdar and Pullum 1981) provide an elegant scheme for describing free, or relatively free, order of constituents within a construct: immediate dominance, or ID, rules, which license the branching of a construct into certain constituents, are distinguished from linear precedence, or LP, rules, which stipulate that certain constituents must occur in a specified order whenever they are sisters.[1] Thus, given the ID rule VP ---> V, NP, PP and no LP rule imposing an order on any two of the three constituents, all six constituent orders are permitted.

The fewer LP rules a language has, the freer its constituent order. Consider, for example, a language with the ID rules in (1) but with no (relevant) LP rules whatsoever.

(1) S ---> NP, VP
NP ---> (A), N
VP ---> V, NP

In this language a subject-verb-object sentence with a structure as in (2)

(2) [[A N_1] [V [N_2]]]
S NP VP NP

has eight variant orders--with A and N1 in either order, V and N2 in either order, and the A-N1 and V-N2 combinations in either order. The eight orders are listed in (3).

(3) A N_1 V N_2 : A N_1 N_2 V : V N_2 A N_1 : V N_2 N_1 A
N_1 A V N_2 : N_1 A N_2 V : N_2 V A N_1 : N_2 V N_1 A

This is *free constituent order*: within any construct, the constituents can occur in all possible orders--but these constituents, taken together, always make a continuous unit. Free constituent order is not the same thing as *free word order*. In a language with free word order, (2) would have not 2^3 = 8 variants, but 4! = 24: within any clause, the words can occur in all possible orders.

2. **Liberation metarules.** Pullum (1982) proposed to use the metarule feature of the GPSG framework to describe free word order; a metarule predicts the existence of one set of ID rules (the *consequent* rules) from the existence of another set (the *antecedent* rules). A 'liberation metarule' (lm) à la Pullum can, for instance, scramble NP constituents within the VP:

(4) *IF* VP ---> NP[F], X
THEN VP ---> Det[F], N[F], X

According to (4), if it is permissible for VP to branch into certain constituents, namely an NP and any set X of other constituents, then it is also permissible for VP to branch into a somewhat different set of constituents, comprising Det, N, and the constituents in X. Thus, if /NP, V/, /NP, NP, V/, /NP, PP, V/, /NP, V, VP/, say, can all be constituent-sets of VP, then so can /Det, N, V/, /Det, N, NP, V/, /Det, Det, N, N, V/, /Det, N, PP, V/, and /Det, N, V, VP/.

As it happens, the literature on lms has concerned itself almost entirely with one particular lm, namely the one that liberates the constituents of VP into S as in (5), thereby permitting LP rules that express ordering conditions on the 'Satzglieder', that is, the major phrases of an S.

(5) *IF* VP ---> V, X
THEN S ---> NP, V, X

Versions of this lm are given for Makua by Stucky (1981, 1982, 1983), for Modern Greek by Horrocks (1983, 1984), and for German by Uszkoreit (1983).

3. **Problems with lms.** What's wrong with liberation metarules like (4) and (5)? Five things that I can see.

3.1. **Connection to mediators not expressed.** These lms are not just any old metarules. Rather, in each case the formula for the antecedent and the formula for the consequent stand in a special relationship. In (4), the left-hand sides of the two formulae are identical, and the right-hand sides are nearly so, differing only in that the right-hand side for the antecedent mentions a single category (NP), the *mediating construct*, where the right-hand side for the consequent mentions several (Det, N), the *mediating constituents*. In (5), the right-hand side of the antecedent is a subpart of the right-hand side of the consequent; here the mediating construct is the left-hand side of the consequent (S), while the mediating constituents are the right-hand side of the antecedent (VP) and the extra constituents in the left-hand side of the consequent (NP).

Indeed, the mediating construct (NP in (4), S in (5)) and the mediating constituents (/Det, N/ and /NP, VP/, respectively) are not randomly associated with one another. Rather, they are related via an existing *mediating rule* or *mediator*, NP ---> Det, N in (4), and S ---> NP, VP in (5).

What is important here is that in (4) and (5) the mediating construct and mediating constituents are not connected in any way to the mediating rule. Nothing says that (4) and (5) are much more probable metarules than, say, (4') and (5').

(4') *IF* VP ---> NP[F], X
THEN VP ---> P[F], A[F], X

(5') *IF* VP ---> V, X
THEN> AP ---> PP, N, X

3.2. **Connection among translations not expressed.** Second, as (4) and (5) are stated, the semantic interpretation principle for the consequent has no intrinsic connection to the translations for the antecedent or the mediator, even though it is in fact a composite of these.

That is, in (4), the way in which the translation for VP is built up out of the translations for Det, N, and the X constituents is that the translations for Det and N are combined in the same way that they are in the mediating rule NP ---> Det, N, and this translation is combined with the translations for the X constituents in the same way that the translation for NP is combined with the translations for the X constituents in an antecedent rule VP ---> NP, X. The way in which translations combine must be stipulated in a full statement of (4), at least in versions of GPSG advanced until recently.[2] Similar observations hold for (5).

3.3. **Inheritance of features stipulated rather than predicted.** Third, the inheritance of the features from the mediating construct to the mediating constituents has to be stipulated in (4)--as if this inheritance had no connection to other principles of feature inheritance, in particular the Head Feature Convention and the Control Agreement Principle. But in fact, N in the consequent of (4) should bear the features [F] as a result of the fact that NP in the antecedent bears those features, via the HFC; and Det in the consequent should bear those features as a result of the fact that N bears them, via the CAP.

3.4. **Spurious structural ambiguities predicted.** Fourth, clauses in which coconstituents occur contiguous to one another are assigned a number of constituent structures, ranging from the fully hierarchical (or 'configurational') to the utterly flat (or 'nonconfigurational'). But there is no reason to think that such clauses have any structure other than the perfectly flat. Certainly there is no reason to think that they are structurally ambiguous.

To see the problem, suppose we're looking at a language with both the lms (4) and (5), and consider a SVO sentence with the shape NP V Det N. This sentence is predicted to have four distinct structures (all with the same translation): one completely hierarchical structure, assuming neither (4) nor (5) has applied ([NP [V [Det N]]]); one flat structure, assuming both have applied ([NP V Det N], with neither Det N nor V Det N making a constituent); and two intermediate structures, assuming that only one lm has applied ([NP [V Det N]] and [NP V [Det N]]).

But students of free word order languages (like Hale 1983 discussing Warlpiri) observe that there is no evidence for internal constituent structure in these languages; it is this fact that has led some to suggest that every clause in these nonconfigurational languages has the form W*, each S branching directly into its component words.

3.5. **Restriction to lexical metarules violated.** Fifth, some lms will have antecedents (like VP ---> VP, AdvP or NP ---> NP, PP) that do not introduce lexical categories--hence, according to Flickinger (1983), should not be possible metarules at all. If we adhere to Flickinger's restriction, then it will be impossible to liberate the constituents of a

VP into a larger VP also containing an AdvP, or to liberate the constituents of an NP into a larger NP also containing a PP.

4. **The universal liberation metarule approach.** One response to these criticisms would be to posit a single lm as in (6), made available (though not made obligatory) by universal grammar. Here I follow a suggestion by Horrocks (1984: 119): 'Let us suppose that UG makes available a basic, fully hierarchical, X̄-theory, and ... devices for "flattening out" hierarchical structure, namely generalised versions of' his lms.

(6) *IF* A ---> B, X *AND* B ---> Y *THEN* A ---> X, Y,
where A is any category, B any category other than S,
and X and Y any sets of categories.

The restriction that B not be S is designed to prevent the liberation of material from a clause into a superordinate S; the generalization here is the familiar one that languages do not permit the interpolation of material from one clause within another.

An analysis using the universal lm (ulm) in (6) would provide the basis for a response to the first three criticisms. As to the first: The ulm explicitly mentions two antecedent rules, one of which is the mediator, in addition to the consequent. The mediator in (4), NP ---> Det, N, is the second antecedent in the scheme of (6); the mediator in (5), S ---> NP, VP, is (in the form S ---> VP, NP) the first antecedent in the scheme of (6).

As to the second: Though (6) does not specify the translation for the consequent, this can in fact be given by the ulm - note, given universally, not stipulated for each lm - as composed from the translations for the two antecedents in the scheme of (6).

As to the third: Since the mediator is explicit as one of the antecedents in the scheme of (6), features can be assumed to be distributed in all the relevant rules by means of the HFC and CAP.

The fourth and fifth criticisms remain, however. In addition there is now a sixth criticism: This analysis predicts that any language is either perfectly nonconfigurational (if it has (6)) or else basically configurational (if it lacks (6) - though it might have some parochial lms, if any of these are allowed). Such a typological divide, labeled the 'bifurcationist view' by Pullum (1982: 215), seems too sharp, though it has been favored by some, for instance Hale (1982). Hale himself has abandoned strict bifurcationism; in Hale (1983: 44-6), he speculates that there might be three subtypes of nonconfigurational languages.

In any event, surprisingly little seems to be known about degrees of word order freedom in the languages of the world; the now immense literature on word order (much of it summarized in Hawkins (1983, esp. chs. 1 and 2)) is focused on linearization, in particular on potential universal principles of linearization, rather than on freedom. The Australian aboriginal languages Hale has studied appear to represent one typological extreme, while English happens to be at or close to the other extreme, but we have little systematic knowledge of what is possible in

between, and it would be inappropriate to enforce bifurcationism on the basis of a very small and skewed sample of languages.

5. **Parametrization of the ulm.** Intermediate degrees of configurationality are possible if (6) is constrained so as to apply only for certain specified category pairs (A, B) in a particular language, i.e. if L is parametrized as in (6') to a set P of mother-daughter pairs of phrasal categories. Again I follow a suggestion made by Horrocks (1984: 119): 'We might also suppose that the grammars of languages make varying degrees of use of these devices [lms], some, perhaps the so-called W* languages (see Hale 1982 and Pullum 1982), making free and extensive use of both, others, say Modern Greek and English, making very limited use of them.'

(6') Given a pair-set P for a language, then for any pair (A, B) in P,
   a. A and B are phrasal categories,
   b. B ≠ S, and
   c. *IF* A ---> B, X *AND* B ---> Y *THEN* A ---> X, Y,
      where X and Y are any sets of categories.

For a perfectly configurational language P would be empty, or as I shall say, has the value *NULL*, while for a perfectly nonconfigurational language P would include every relevant pair, or as I shall say, has the value *ALL*. Other order types are specified with reference to *NULL* or *ALL*. Nearly configurational languages would require a specification of P, and nearly nonconfigurational languages would require a specification of the pairs not in P. Assuming that a language is more complex as more conditions are required to describe its pair-set, this proposal would favor the pure order types, while still permitting intermediate ones to occur.

A very common sort of nearly nonconfigurational language - exemplified by Finnish, Tamil, and Korean, according to my informants - is one in which word order is free, except that the constituents of NP must make a continuous unit (and typically are subject to rigid ordering constraints of their own, describable in LP rules). For such a language, the pair-set P includes all except those in which B = NP.

I do not suppose that there is nothing more to be said about the pair-sets that occur in the grammars of the world's languages, that languages are, so to speak, free to pick any pair-set whatsoever. On the contrary, I expect that there are constraints on, and implicational generalizations within, pair-sets, and that these should be stated as part of universal grammar, though I am not now in a position to make specific proposals about the matter. As I pointed out above, much remains to be discovered about the extent of intermediate degrees of configurationality in language.

6. **Phantom rules.** The fourth and fifth criticisms have still not been addressed. To avoid predicting spurious structural ambiguities, we can arrange things so that only one structure is assigned, by generalizing the idea of 'phantom category' already proposed in GPSG (Gazdar and Sag 1981).

A phantom category is one that appears as the construct in at least one ID rule but not as a constituent in any ID rule. As a result, it will not appear as a node label in constituent structures, but rules with the phantom category as their construct will nevertheless be available as antecedents for metarule application.

The extension of this idea that I have in mind is the following: Not all ID rules will actually license branchings; instead, some will act as 'phantom rules', serving only to provide antecedents for applications of (6'). In particular, I propose that the effect of (6') be reined in by (7), which is a revision of one of the conditions defining well-formed constituent structures:

(7) A rule A ---> B, X
licenses the branching of A into B and X
only if (A, B) is not in the pair-set P.

To see how (7) works, consider a language with the ID rules in (8a-e) and with a pair-set P containing all relevant pairs in which B ≠ NP.

(8) a. S ---> NP, VP
b. NP ---> Det, N
c. VP ---> V, NP, (AdvP)
d. VP ---> V, VP
e. AdvP ---> Deg, Adv

In this language the following pairs (among others) are in P: (S, VP), (VP, V), (VP, AdvP), (VP, VP). According to (7), then, the ID rules (8a), (8c), and (8d) fail to license branchings and so are phantom rules. From (8a) and (8d), by (6') it follows that

(8) f. S ---> NP, V, VP

is an ID rule, but according to (7) it is a phantom rule, since (S, VP) is in P. From (8f) and (8c), by (6') it follows that

(8) g. S ---> NP, V, V, NP

is an ID rule, *and* an actual one, since neither (S, NP) nor (S, V) is in P (the first is out because its second member is NP, the second because its second member is a lexical category, V). The set of actual rules also includes (8h-j):

(8) h. S ---> NP, V, NP
i. S ---> NP, V, NP, Deg, Adv
j. S ---> NP, V, V, NP, Deg, Adv

On this analysis the structure assigned to a construct is the flattest one available. In my example, those are the structures provided by (8g-j).

Adopting (7) changes the character of (6') entirely. Its effect is now to specify the syntax of a language, its set of actual rules, as a subset of a larger set of rules, rather than to express implicational generalizations about the set of actual rules (and so in effect extending

the set of actual rules). That is, (6') is not a metarule at all, but a (universal) principle of a quite different type. The fifth criticism of (6) is thus averted, and our proposal no longer depends on the metarule feature of GPSG at all - a welcome consequence in light of recent moves (on independent grounds) to abandon this feature in favor of other theoretical constructs, as in Pollard (1984).

A final note. In a framework combining (6') and (7), many of the properties that have been claimed to characterize nonconfigurational languages do not follow from the fact that for these languages P = *ALL*. Consider the list in Hale (1982: 86-7) - (a) free word order, (b) discontinuous constituency, (c) pronoun drop, (d) lack of NP-movement transformations, (e) lack of dummy NPs, (f) rich case system, (g) complex verb words - to which I would add (h) the occurrence of fixed-position (usually second-position) clitics. Properties (a), (b), and (d) all follow from my ulm treatment, but the others do not. Properties (c) and (e) are two sides of the same coin, but the fact remains that nothing I have said would connect free word order with pronoun drop, rich case systems, complex verb words, or fixed-position clitics.

This does not seem to me to be an unwelcome state of affairs. Though there might well be some universal associations within this set of properties, I doubt that any of them entails or is entailed by free word order, so that it would be no virtue for a theoretical framework to necessitate one or more of these entailments. There are the familiar functional reasons motivating a tendency towards a trade-off between rigid order and such properties as case marking, complex verb words, and fixed-position clitics, and in my view these reasons need not be framed as principles of universal grammar. Consider Lisu, which has extraordinarily free constituent order; according to Hope (1973: 85-6), the sentence 'This morning beside your house I gave Asa a slap on his ear' has 720 grammatical, and synonymous, translations into Lisu - one for each of the orders of the six NPs involved. Thus, Lisu is a nearly nonconfigurational language of the same type as Finnish, Tamil, and Korean. But unlike the latter three languages, with their rich case systems and extensive agreement morphology on verbs, Lisu lacks nominal inflection (though topic NPs are marked by an enclitic) and has no verbal morphology beyond the marking of mood. As Hope points out, widespread ambiguity results with respect to the grammatical relations NPs bear to the verb of their clause; speakers manage disambiguation nonsyntactically - by reference to (linguistic or nonlinguistic) context and real-world knowledge.

Notes

*The main ideas in this paper were developed in a GPSG course at Ohio State in autumn 1984 and expanded in a seminar in summer 1985; essentially the current version was presented at the Linguistic Institute summer meetings at Georgetown University in July 1985. Thanks to Gerald Gazdar and Geoffrey K. Pullum for their comments on an earlier draft and to Belinda Brodie, Paul Chapin, and Joel Nevis for their comments on the July 1985 version. This is the version of 25 July 1985.

[1]As Pullum (1982) observes, adopting this version of the ID/LP format restricts the set of languages with GPSG syntaxes to a proper subset of the context-free languages; the restriction follows from the stipulation that an LP rule X < Y requires that X precede Y whenever they are sisters

--regardless of what mother category dominates them and regardless of what ID rule licenses this configuration. If LP rules are permitted to refer to the mother category, then no restriction of the set of CF languages results (though generalizations could still be stated across separate ID rules licensing similar configurations, for instance separate rules licensing the branching of VP into V and VP).

[2]In Klein and Sag (1985) and Gazdar et al. (1985), in contrast, all translation schemes are derivable from the forms of rules.

References

de Geest, W. and Y. Putseys, 1984. Sentential complementation. Dordrecht: Foris.

Flickinger, Daniel P. 1983. Lexical heads and phrasal gaps. WCCFL 2.89-101.

Gazdar, Gerald, Ewan Klein, and Geoffrey K. Pullum (eds.). 1983. Order, concord and constituency. Dordrecht: Foris.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum, and Ivan A. Sag. 1985. Generalized phrase structure grammar. London: Basil Blackwell.

Gazdar, Gerald and Geoffrey K. Pullum. 1981. Subcategorization, constituent order, and the notion "head". In Moortgat et al. 1981: 107-23.

Gazdar, Gerald and Ivan A. Sag. 1981. Passive and reflexives in phrase structure grammar. In Groenendijk et al. 1981: 131-52.

Groenendijk, J.A.G., T.M.V. Janssen, and M.B.J. Stokhof (eds.). 1981. Formal methods in the study of language. Part I. Amsterdam: Mathematisch Centrum.

Hale, Kenneth. 1982. Preliminary remarks on configurationality. NELS 12.86-96.

Hale, Kenneth. 1983. Warlpiri and the grammar of non-configurational languages. NLLT 1.1.5-47.

Hawkins, John A. 1983. Word order universals. New York: Academic Press.

Hope, Edward R. 1973. Non-syntactic constraints on Lisu noun phrase order. FLang 10.1.79-109.

Horrocks, Geoffrey. 1983. The order of constituents in modern Greek. In Gazdar, Klein, and Pullum 1983: 95-111.

Horrocks, Geoffrey. 1984. The lexical head constraint, $\bar{X}$ theory and the "pro-drop parameter". In de Geest and Putseys 1984: 117-25.

Klein, Ewan and Ivan A. Sag. 1985. Type-driven translation. Linguistics and Philosophy 8.2.163-201.

Moortgat, Michael, Harry van der Hulst, and Teun Hoekstra (eds.). 1981. The scope of lexical rules. Dordrecht: Foris.

Pollard, Carl J. 1984. Generalized phrase structure grammars, head grammars, and natural language. Ph.D. dissertation, Stanford Univ.

Pullum, Geoffrey K. 1982. Free word order and phrase structure rules. NELS 12.209-220.

Stucky, Susan. 1981. Free word order languages, free constituent order languages, and the gray area in between. NELS 11.364-76.

Stucky, Susan. 1982. Linearization rules and typology. WCCFL 1.60-70.

Stucky, Susan. 1983. Verb phrase constituency and linear order in Makua. In Gazdar, Klein, and Pullum 1983: 75-94.

Uszkoreit, Hans. 1983. A framework for processing partially free word order. ACL Conference Proceedings 21.106-12.