On Derivational Constraints*

George Lakoff
Harvard University

1. Introduction

I would like to discuss some questions having to do with the theory of grammar. I assume that a grammar of a language is a system of rules that relates sounds in the language to their corresponding meanings, and that both phonetic and semantic representations are provided in some language-independent way. One possibility for phonetic representation is that discussed in Chomsky and Halle, 1968, though there are other possibilities that may turn out to provide better results. I assume that the notion 'possible surface structure' in a possible natural language is defined in terms of 'trees' or 'phrase-markers', with the root S, whose node-labels are taken from a finite set of node-labels: S, NP, V, ... The notion 'tree', or 'phrase-marker', is to be defined in one of the usual ways, in terms of predicates like precedes, dominates, and is labelled. Thus a grammar will define an infinite class of surface structures. In addition I assume that a grammar will contain a system of grammatical transformations mapping phrase-markers onto phrase-markers. Each transformation defines a class of well-formed pairs of successive phrase-markers, P_i and P_{i+1}. These transformations, or well-formedness constraints on successive phrase-markers, P_i and P_{i+1}, define an infinite class K of finite sequences of phrase-markers, each such sequence $P_1,\ldots,P_n$ meeting the conditions:

(1) (i) P_n is a surface structure

(ii) each pair P_i and P_{i+1} meet the well-formedness constraints defined by some transformations

(iii) there is no P_o such that $P_0,P_1,\ldots,P_n$ meets conditions (i) and (ii).

The members of K are called the syntactic structures generated by the grammar. I will assume that the grammar contains a lexicon, that is, a collection of lexical entries specifying phonological, semantic, and syntactic information. Thus, we assume

(2) a lexical transformation associated with a lexical item I maps a phrase-marker P containing a substructure Q which does not already contain a lexical item

* This work was supported in part by Grant GS-1934 from the National Science Foundation for Harvard University. The author wishes to thank P.M. Postal, J.R. Ross, D.M. Perlmutter, R.T. Lakoff, and J.D. McCawley for their extremely helpful discussion of the topics discussed here. This paper is part of a longer work called "On Generative Semantics" to appear in Jakobovits and Steinberg, University of Illinois, forthcoming.

into a phrase-marker P' formed by superimposing I over Q.

That is, a lexical transformation is a well-formedness constraint on classes of successive phrase-markers P_i and P_{i+1}, where P_i is identical to P_{i+1} except that where P_i contains a subtree Q, P_{i+1} contains the lexical item in question. Various versions of this framework will differ as to where in the grammar lexical transformations apply, whether they apply in a block, etc.

In this sense, transformations, or well-formedness conditions on successive phrase-markers, may be said to perform a 'filtering function', in that they 'filter out' derivations containing successive phrase-marker pairs (P_i, P_{i+1}) which do not meet some well-formedness condition on such pairs. A system of transformations is essentially a filtering device which defines a class of well-formed sequences of phrase-markers by throwing out all of those sequences which contain pairs (P_i, P_{i+1}) which do not meet some such well-formedness condition, that is, are not related by some transformation. Since transformations define possible derivations only by constraining pairs of successive phrase-markers, I will refer to transformations as 'local derivational constraints'. A local derivational constraint can be defined as follows. Let 'P_i/T_1' mean phrase-marker P_i meets tree condition T_1. (A structural description of a transformation is one example of a tree condition, further examples will be given below.) A transformation, or local derivational constraint, is a conjunction of the form P_i/T and P_{i+1}/T', where T and T' are tree conditions defining the class of input trees and class of output trees, respectively. Thus, the pair of tree conditions(T,T') defines a local derivational constraint.

In addition to transformations, or local derivational constraints, a grammar will contain certain 'global derivational constraints'. Rule orderings, for example, are given by global derivational constraints, since they specify where in a given derivation two local derivational constraints can hold relative to one another. Suppose (T_1, T_2) and (T_3, T_4) define local derivational constraints. To say that (T_1, T_2) is ordered before (T_3, T_4) is to state a global derivational constraint of the form:

$$(i)\ (j)\ (\ (P_i/T_1 \text{ and } P_{i+1}/T_2 \text{ and } P_j/T_3 \text{ and } P_{j+1}/T_4) \supset (i < j)\ ).$$

Another example of a global constraint is Ross' coordinate structure constraint which states that if some coordinate node A^1 dominates node A^2 at some point in the derivation P_i, then there can be no P_{i+1} such that A^2 commands A^1 and A^1 does not dominate A^2. This is a global derivational

constraint of the form:

Let $T_1 = A^i$ dominates and $A^{j_1} \ldots A^{j_m}$

$T_2 = A^i$ dominates $X^1 A^k X^2$

$T_3 = A^k$ commands A^i

CSC: (y) ($\sim$((P_y/T_1 and T_2) and (P_{y+1}/T_3 and $\neg T_2$)))

What the coordinate structure constraint does is to keep track of the derivational histories of pairs of nodes A^i and A^k. This is just what elementary transformations do: they define constraints on successive phrase-markers, keeping all but one or two nodes constant, and then tell what happens to those one or two other nodes in going from the first tree to the second. It seems reasonable to limit individual derivational constraints, both local and global, to tracing the histories of one or two nodes -- over a derivation in the case of global constraints, and over two successive trees in the case of local constraints. Other examples of global derivational constraints are Ross' other constraints on movement rules (Ross, 1967), the theory of exceptions (Lakoff, 1965 and R. Lakoff, 1968), Postal's crossover principle (Postal, 1968), output conditions for pronominalization (Lakoff, 1968) etc. It should be clear that all theories of transformational grammar include both local and global derivational constraints. The question now arises as to what kinds of local and global derivational constraints exist in natural languages. I will suggest in Section 2 below that there is a wider variety than had previously been envisioned.

Given a syntactic structure $(P_1,\ldots,P_n)$ we define the semantic representation SR of a sentence as SR = (P_1, PR, Top, F,...), where PR is a conjunction of presuppositions, Top is an indication of the 'topic' of the sentence, and F is the indication of the focus of the sentence. We leave open the question of whether there are other elements of semantic representation that need to be accounted for.

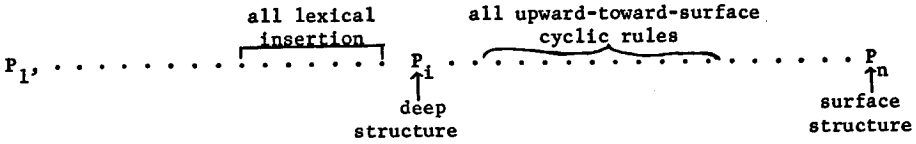
I will refer to any elaborations or refinements of this theory of grammar as a "basic theory", simply for convenience and with no intention of suggesting that there is anything ontologically, psychologically, or conceptually 'basic' about this theory. Most of the work in generative semantics since 1967 has assumed the framework of the basic theory. It should be noted that the basic theory permits a variety of options that were assumed to be unavailable in previous theories. For example, it is not assumed that lexical insertion transformations apply in a block, with no intervening nonlexical transformations. The option that lexical and nonlexical transformations may be interspersed is left open.

As should be obvious from the above discussion, the basic theory does not assume any notion of 'directionality of mapping' from phonetics to semantics or semantics to phonetics. The basic theory generates triples (P,s,S) (P a phonetic representation, s a surface structure, and S a semantic representation.) Some writers on transformational grammar have, however,

used locutions that might mislead readers into believing that they assume some notion of directionality. For example, Chomsky (1968) remarks that "... properties of surface structure play a distinctive role in semantic interpretation." However, as Chomsky points out a number of times in that work, the notion of directionality in a derivation is meaningless, so that Chomsky's locution must be taken as having the same significance as "Semantic representation plays a distinctive role in determining properties of surface structure" and nothing more. Both statements would have exactly as much significance as the more neutral locution "Semantic representation and surface structure are related by a system of local and global derivational constraints." The basic theory includes a notion of transformational cycle in the sense of Aspects, so that a sequence of cyclical rules applies "from the bottom up", first to the lowermost S's, then to the next highest, etc. We assume that the cyclical transformations start applying with P_k and finish applying (to the highest S) at P_ℓ where k is less than ℓ. We will say in this case that the cycle applies 'upward toward the surface structure' (though, of course, we could just as well say that it applies 'downward toward the semantic representation, since directionality has no significance).

It should be noted that a transformational cycle defines an 'orientation' on a derivation, and readers should be cautioned from confusing the notion 'cyclical orientation of a derivation' with the notion 'directionality of a derivation'. The former is a real and quite important notion; the latter is meaningless. To say that a cycle is oriented 'upward-toward-the-surface' is the same as to say that it is oriented 'downward-toward-the-semantics', and such terminology makes no claim about where a derivation 'begins'. Most theories of transformational grammar that have been seriously entertained have assumed a cyclical orientation that is upward-toward-the-surface. However, it is possible to envision a theory with an upward-toward-the-semantics cyclic orientation. Moreover, it is possible to imagine theories with more than one cyclic orientation. Consider a sequence of phrase-markers $P_1,\cdots,P_i,\cdots,P_n$. One could imagine a theory such that $P_1,\cdots,P_i$ had an upward-towards-the-semantics cyclic orientation and $P_i,\cdots,P_n$ had an upward-toward-the-surface orientation, or vice-versa.

The basic theory does not necessarily include a level of 'deep structure', and the question as to whether such a level exists is an empirical question in the basic theory. We assume that the notion of 'deep structure' is defined in the following way. (i) Assume that all lexical insertion rules apply in a block. (ii) Assume that all upwards-toward-the-surface cyclic rules follow all lexical insertion rules. We define the output of the last lexical insertion rule as 'deep structure'.



In the following section, we will look carefully at an example of a global derivational constraint and at the evidence it provides for the question of whether 'deep structure' in this sense exists.

2. A derivational constraint involving quantifiers

Let us consider in detail an example of a global derivational constraint.

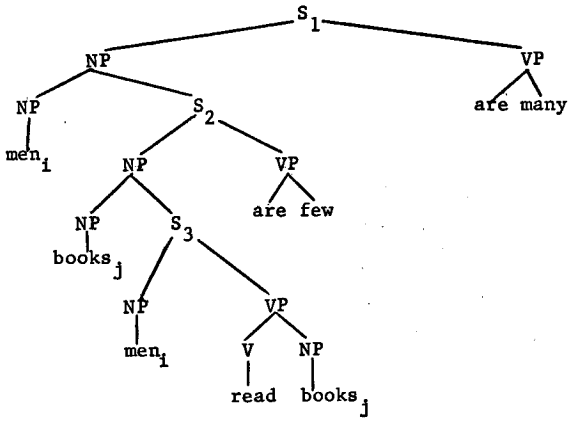
(1) Many men read few books.

(2) Few books are read by many men.

As is obvious, (1) and (2) are not synonymous.

Sentences like (1) and (2) were raised in a discussion by Partee (1969) with regard to certain inadequacies of a proposal made in Lakoff, 1965 for the derivation of quantifiers from predicates of higher sentences. That proposal was based on the observation that sentences like "Many men left" are synonymous with those like the archaic "The men who left were many". It was proposed that sentences like the former were derived from structures underlying sentences like the latter, with "many" as an adjective, which is then lowered. Under such a proposal, underlying structures like (3) and (4) would be generated.[1]

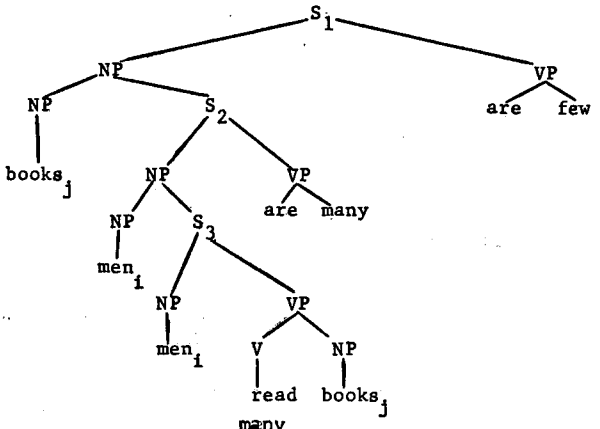
(3)



[1] I don't mean to suggest that P_1 would look like this in detail. The crucial point here is the relative height of many and few. In particular, I do not mean to suggest that the i in the men_i, that is the head of the relative clause would be a set index. Rather, an adequate semantic representation would have it as a variable as in {i:i has the property of being a man}.

(3') Many are the men who read few books. There are many men who read few books.

(4)



(4') Few are the books that $\wedge$ men read. There are few books that many men read.

In (3), a cyclical rule of quantifier-lowering will apply on the S_2-cycle, yielding men$_i$ read few books. The same cyclic rule will apply on the S_1-cycle, lowering many onto men$_i$ and yielding (1), many men read few books. In (4) let us suppose that the passive, a cyclical rule, applies on S_3, before any quantifier-lowering takes place: this will give us books$_j$ are read by men$_i$. On the S_2-cycle, many will be lowered onto men$_i$, yielding books$_j$ are read by many men. On the S_1-cycle, few will be lowered onto books$_j$ giving (2), few books are read by many men.

These derivations work as they should and account for the synonymy of (1) with (3') and (2) with (4'). However, if nothing more is said, this proposal will yield incorrect results. For example, if the passive applies to S_3 in (3), we will get books$_j$ are read by men$_i$, and then quantifier-lowering on the S_2- and S_1-cycles will yield (2), few books are read by many men. But if (2) were derived from (3) in this fashion, it should be synonymous with (3'), many are the men who read few books. This is false, and such a derivation must be blocked. Similarly, if the passive were not to apply to S_3 in (4), the application of quantifier-lowering on the S_2- and

S_1-cycles would yield (1), many men read few books, again a mistake, since (1) does not have the meaning of (4'), few are the books that many men read.

Such a proposal would work in the first two cases, but would also predict the occurrence of two derivations that do not occur, at least for the majority of English speakers. If one inspects (1) - (4), one notices that the correct derivations have a very interesting property, namely, that the 'higher' quantifiers in (3) and (4) are the leftmost quantifiers in (1) and (2) respectively. Thus, we might propose a derivational constraint that would say something like: if one quantifier commands another in underlying structure (or rather, P_1), then that quantifier must be leftmost in surface structure. Such a constraint as it stands would be too strong. Consider cases like (5).

(5) The books that many men read are few (in number).

(5) would have an underlying structure like (4), where few is the higher quantifier; however, few is to the right of many in surface structure. Thus cases like (5) would have to be ruled out of any such derivational constraint. If one inspects (2) and (5), one sees that they differ in a very interesting way. In (5), few commands many, but many, being in a relative clause, does not command few; that is, few is higher in the tree than many in (5), just as it is in the underlying structure of (4). In other words, (5) preserves the asymmetric command-relationship between the quantifiers that occurs in (4). In (2), however, this is not the case. In (2), neither few nor many is in a subordinate clause, and so each commands the other and the command-relationship is symmetric. Thus the asymmetry of the command-relationship in the underlying structure, where few commands many but many does not command few, is lost in (2). It is exactly these cases where the quantifier that was higher in underlying structure must be leftmost in surface structure. Where the asymmetric command-relationship is lost it must be supplanted by a precede-relationship, which is necessarily asymmetric.

Such a derivational constraint may be stated as follows:

(6) Let $T_1 = Q^1$ commands Q^2

$T_2 = Q^2$ commands Q^1

$T_3 = Q^1$ precedes Q^2 '/' means 'meets condition'

Constraint 1: $P_1/T_1 \supset (P_n/T_2 \supset P_n/T_3)$

Constraint 1 states that if two quantifiers Q^1 and Q^2 occur in an underlying structure P_1, such that P_1 meets condition T_1, then if the corresponding surface structure P_n meets condition T_2, that surface structure P_n must also meet condition T_3. In short, if an underlying asymmetric command-relationship breaks down in surface structure, a precede-relationship takes over. Constraint 1 is a well-formedness constraint on derivations. Any derivation not meeting it will be blocked. Thus, the derivations (3) → (1) and (4) → (2) will be well-formed, but (3) → (2) and (4) → (1) will be blocked.

It is important to note that the fact that one of the two quantifiers is in subject position in the sentences we have discussed so far is simply an accident of the data we happened to have discussed. The difference in the interpretation of quantifiers has nothing whatever to do with the fact that in these examples one quantifier is inside the VP while the other is outside the VP. Only the left-to-right order within the clause matters.

(7) John talked to few girls about many problems.

(8) John talked about many problems to few girls.

These sentences differ in interpretation just as do (1) and (2), that is the leftmost quantifier is understood as the highest in each sentence.

Although (1) and (2) are cases where the asymmetry of the underlying command-relationship disappears in surface structure, it happens to be the case in (1) and (2) that condition T_1, which holds in underlying structure continues to hold in surface structure: that is, Q^1 continues to command Q^2. We might ask if there exist any cases where this does not happen, that is, where Q^1 commands Q^2 in underlying structure, but Q^1 does not command Q^2 in surface structure. A natural place to look for such cases, and perhaps the only one, is in sentences containing complement constructions. Let us begin by considering sentences like (9).

(9) Sam claimed that John had dated <u>few</u> girls.

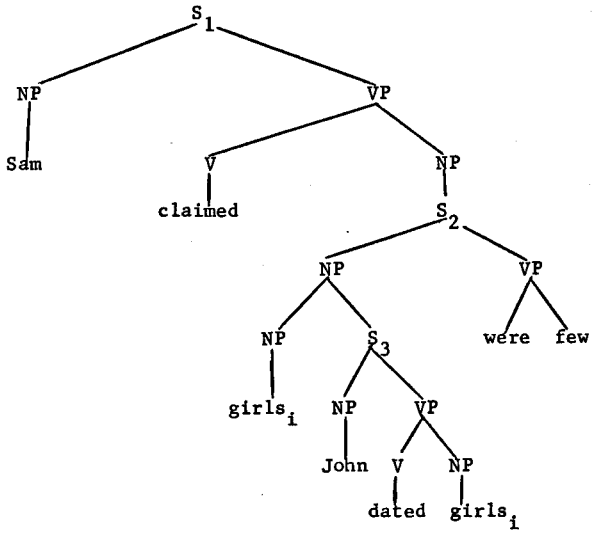
(9) is open to both of the readings (10) and (11), though (10) is preferable.

(10) Sam claimed that the girls who John had dated were few (in number).

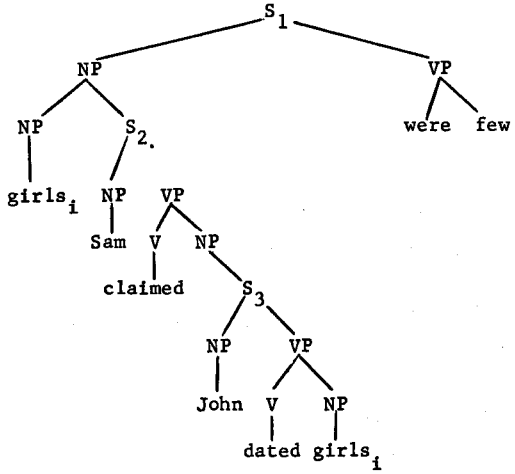
(11) The girls who Sam claimed that John had dated were few (in number).

(10) and (11) would have underlying structures like (44') and (45') respectively.

(10')

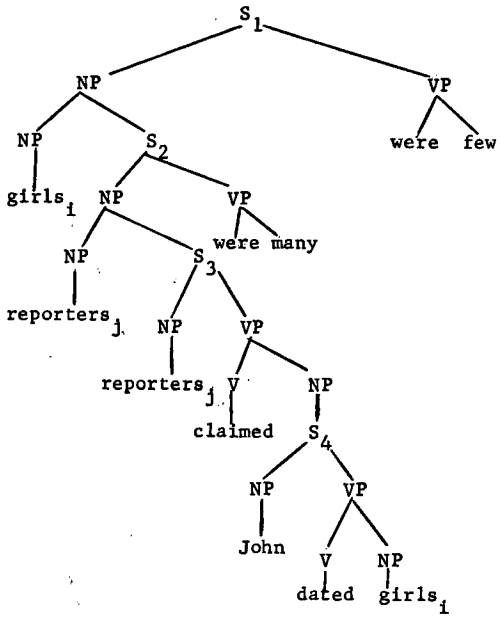

S1
NP
VP
Sam
V
NP
claimed
S2
NP
VP
NP
S3
were
few
girls i
NP
VP
John
V
NP
dated
girls i

(11')


S1
NP
VP
NP
S2.
were
few
girls i
NP
VP
Sam
V
NP
claimed
S3
NP
VP
John
V
NP
dated
girls i

In each case quantifier-lowering will move few down to $girls_i$. In (10), $girls_i$ occurs in the S immediately below few; in (11), $girls_i$ occurs two sentences down from few. We are now in a position to test the conjecture that one quantifier may cease to command another in the course of a derivation. Consider (12), where few commands many.

(12)



(12) would have the meaning of (12')

(12') Few were the girls who many reporters claimed John dated.

If we allow quantifier-lowering to apply freely to (12), many will be lowered to $reporters_j$ on the S_2 cycle, yielding many reporters claimed that John dated $girls_i$. The derived structure will now look just like (11), except that it will have the noun phrase many reporters instead of Sam. As in (11), quantifier-lowering will lower few onto $girls_i$ yielding (13).

(13) Many reporters claimed that John dated few girls.

In (13), few is in a subordinate clause and does not command many. Thus we have a case where few commands many in underlying structure, but not in surface structure. Note, however, that (13) does not have the meaning of

(46'). It has the reading of (14),

(14) Many were the reporters who claimed that the girls who John dated were few (in number).

where few is inside the object complement of claim (as in (10)) and where many would command few in underlying structure. Thus, we have a case where a derivation must block if one quantifier commands another in underlying structure, but not in surface structure. To my knowledge this is a typical case, and I know of no counterevidence. Thus, it appears that a derivational constraint of the following sort is needed.

(15) Let $T_1 = Q^1$ commands Q^2 '/' means 'meets condition'

Constraint 2: $(P_1/T_1) \supset (P_n/T_1)$

Constraint 2 says that if Q^1 commands Q^2 in underlying structure P_1, then Q^1 must also command Q^2 in surface structure P_n.

Constraints 1 and 2 are prime candidates for cases where grammatical constraints seem to reflect perceptual strategies. If one considers a perceptual model where surface strings are given as input and semantic representations are produced as output, constraints 1 and 2 guarantee that the relative heights of the quantifiers in the semantic representation of a sentence can be determined by the surface parsing of the sentence. If Q_1 commands Q_2 in surface structure but Q_2 doesn't command Q_1 then Q_1 commands Q_2 in semantic representation. If, on the other hand, Q_1 and Q_2 command each other in surface structure, then the leftmost quantifier commands the rightmost one in semantic representation. If Constraints 1 and 2 are reflections in grammar of perceptual strategies, then they would of course be prime candidates for syntactic universals. Unfortunately for such a proposal, there is a lot of idiosyncratic variation with such constraints.

Constraint 2 does not simply hold for quantifiers, but for negatives as well. Consider, for example, (16),

(16) Sam didn't claim that Harry had dated many girls.

where many does not command not. If quantifier-lowering worked freely one would expect that (50) could be derived from all of the following underlying structures.

(17) $[_S$ not $[_S$ Sam claimed $[_S$girls$_i$ [Harry dated girls$_i$] were many]]]

(18) $[_S$ not $[_S$ girls$_i$ $[_S$Sam claimed $[_S$Harry dated girls$_i$]] were many]]

(19) $[_S$girls$_i$ $[_S$not $[_S$Sam claimed $[_S$Harry dated girls$_i$]]] were many]

These have the senses of:

(17') Sam didn't claim that the girls who Harry dated were many.

(18') There weren't many girls who Sam claimed Harry dated.

(19') There were many girls who Sam didn't claim Harry dated.

(17) is the normal reading for (16); (18) is possible, but less preferable (like (11)); but (19) is impossible. The regularity is just like that of Constraint 2. In (17) and (18), not commands many in underlying structure, just as in surface structure (16). In (19), many commands not in underlying structure, but many does not command not in surface structure (16). Thus we can generalize Constraint 2 in the following way. Let L stand for a 'logical predicate', either Q or NEG.

(20) Let T_1 = L^1 commands L^2 (L = Q or NEG)

Constraint 2': $(P_1/T_1) \supset (P_n/T_1)$

Conditions of this sort suggest that quantifiers and negatives may form a natural semantic class of predicates. This seems to be confirmed by the fact that Constraint 1 can be generalized in the same fashion, at least for certain dialects of English. Consider for example, the following sentences discussed by Jackendoff (1968).

(21) Not many arrows hit the target.

(22) Many arrows didn't hit the target.

(23) The target wasn't hit by many arrows.

Jackendoff reports that in his speech (23) is synonymous with (21), but not (22). I and many other speakers find that (23) has both readings, but that the (22) reading is 'weaker'; that is, (23) is less acceptable with the (22) reading. However, there are a number of speakers whose dialect displays the facts reported on by Jackendoff, and in the remainder of this discussion we will be concerned with the facts of that dialect.

Assuming the framework discussed above, (21) and (22) would have underlying structures basically like (24) and (25).

(24) $[_S$ not $[_S$ arrows$_i$ $[_S$ arrows$_i$ hit the target] were many]]

(24') The arrows that hit the target were not many.

(25) $[_S$arrows$_i$ $[_S$ not $[_S$arrows$_i$ hit the target]] were many]

(25') The arrows that didn't hit the target were many.

If Constraint 1 is generalized to include 'logical predicates', both negatives and quantifiers, then the facts of (21) - (23) will automatically be handled the new Constraint 1', given the underlying structures of (24) and (25) and the rule of quantifier-lowering. Constraint 1' would be stated as (26).

(26) Let $T_1 = L^1$ commands L^2

$T_2 = L^2$ commands L^1

$T_3 = L^1$ precedes L^2 (L = Q or NEG)

Constraint 1': $P_1/T_1 \supset (P_n/T_2 \supset P_n/T_3)$

Any derivation not meeting this condition will be ill-formed.[2]

(21) and (22) work as expected. Take (21): not(L^1) commands many (L^2) in underlying structure (24), many commands not in surface structure (21), and not precedes many in surface structure. So (24) → (21) meets Constraint 1'. Take (22): many (L^1) commands not (L^2) in underlying structure (25), not commands many in surface structure, and many precedes not in surface structure. So (25) → (22) meets Constraint 1'. Now consider (23), which is the interesting case in this dialect. If one allows the passive transformation to apply to the innermost S of (24) and (25), and then allows quantifier-lowering to apply, both (24) and (25) will yield (23). First consider the derivation (24) → (23). Not (L^1) commands many (L^2) in underlying structure (24), many commands not in surface structure (21),

---

2 I have ignored the role of stress in this discussion, though it is of course important for many speakers. Many people find that (i)

(i) Many men read few books.

where few has extra heavy stress can mean The books that many men read are few. Thus, the general principle here seems to be that where the asymmetric command relation is lost in derived structure, then either one or another of what Langacker (1966) calls 'primacy relations' must take over. One is the relation 'has much heavier stress than'; the other is the relation 'precedes'. These relations seem to form a hierarchy with respect to this phenomenon in such dialects:

1. Commands (but is not commanded by)
2. Has much heavier stress than
3. Precedes

If one quantifier commands but is not commanded by another in surface structure, then it commands in underlying structure. If neither commands but is not commanded by the other in surface structure, then the one with heavier stress commands in deep structure. And if neither has much heavier stress, then the one that precedes in surface structure commands in underlying structure. Letting $T_4 = Q^1$ has much heavier stress than Q^2, the constraint for such a dialect could be stated as follows, though the notation is not an optimum one for stating such a hierarchy:

$$P_1/T_1 \supset (P_n/T_2 \supset ((P_n/T_4 \vee P_n/T_3) \cdot (\sim P_n/T_4 \supset P_n/T_3)))$$

Since the dialect with this condition is far in the minority so far as I have been able to tell from a very small amount of study, I will confine myself to the normal dialect in the remainder of the discussion.

and not precedes many in surface structure. Thus, the derivation (24) → (23) meets Constraint 1'. Now consider the derivation (25) → (23). Many (L^1) commands not (L^2) in underlying structure, not commands many in surface structure (21), but many (L^1) does not precede not (L^2) in surface structure (55). Therefore, the derivation (25) → (23) does not meet Constraint 1', and so the derivation is blocked. This accounts for the fact that (23) is not synonymous to (21) and that there is no passive corresponding to (25) in this dialect.

It should also be noted that that part of Constraint 1' which says that L^2 must command L^1 in surface structure (P_n/T_2) if the precede-relationship (P_n/T_3) is to come into play, is necessary for the cases discussed. Consider, for example, sentence (25'), The arrows that didn't hit the target were many. In (25'), many (L^1) commands not (L^2) in P_1, and many also commands not in P_n. Thus, the if part of the conditional statement of Constraint 1' is not met, and the fact that not (L^2) precedes many (L^1) in surface structure (that is, (P_n/T_3) does not hold) does not matter; since the if-condition is not met, the constraint holds and the sentence is grammatical.

We have assumed so far that Constraints 1' and 2' mention the surface structure. But this is just an illusion which results from considering only simple sentences. Suppose, for example, that we consider complex sentences where deletion has taken place. Consider (27) and (28).

(27) Jane isn't liked by many men and Sally isn't liked by many men either.

(28) Jane isn't liked by many men and Sally isn't either.

Note that the sentence fragment Sally isn't either does not contain many in surface structure, but it receives the same interpretation as the full Sally isn't liked by many men either, and does not have the reading of There are many men who Sally isn't liked by. Constraint 1', as it is presently stated will not do the job, in that it mentions surface structure P_n rather than some earlier stage of the derivation prior to the deletion of liked by many men.

This raises a general problem about constraints like 1' and 2'. Since they only mention underlying structures P_1 and surface structures P_n, they leave open the possibility that such constraints might be violated at some intermediate stage of the derivation. My guess is that this will never be the case, and if so, then it should be possible to place much stronger constraints on derivations than 1' and 2' by requiring that all intermediate stages of a derivation P_i meet the constraint, not just the surface structure P_n. Using quantifiers, we can state such a stronger constraint as follows

(29) Let $T_1 = L^1$ commands L^2

$T_2 = L^2$ commands L^1

$T_3 = L_1$ precedes L_2

Constraint 1": $P_1/T_1 \supset ((i)(P_i/T_2 \supset P_i/T_3))$

(29) will now automatically handle cases like (28), since it will hold at all points of the derivation up to the point where the deletion rule applies; after that point, it will hold vacuously. The reason is that many will not appear in any phrase-marker after the deletion takes place and so T_2 will not hold in such phrase-markers; and where T_2 does not hold, then T_3 need not hold.

However, this is still insufficient, since Constraint 1''' still requires that if T_2 holds in surface structure, then T_3 must hold in surface structure, as well as at earlier stages of a derivation. But there are late rules which make gross changes in derived structure and produce surface structures in which the constraints do not hold. Compare (30) and (31)

(30) Sarah Weinstein isn't fond of many boys.

(31) Fond of many boys, Sarah Weinstein isn't.

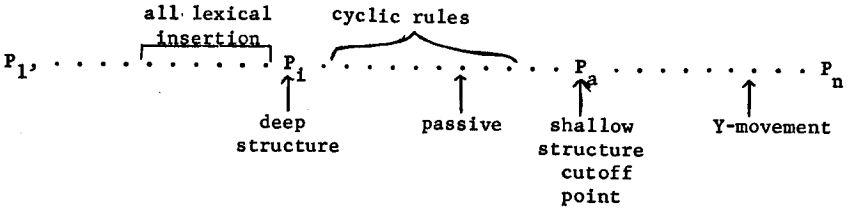
The rule of Y-movement as discussed in Postal, 1968, will produce (31) from the structure underlying (30). (Needless to say, (31) is not grammatical in all American dialects. We are considering only those in which (31) is well-formed.) Note that (30) works exactly according to Constraint 1''; the reading in which many commands not in P_1 is blocked since many does not precede not in surface structure. But (31), where the surface order of not and many is reversed, shows the same range of blocked and permitted readings as (30). The outputs of Y-movement do not meet Constraint 1'', though earlier stages of the derivation do. Thus, it appears that Constraint 1'' has some cutoff point prior to the application of Y-movement. That is, there is in each derivation some 'shallow structure' P_a defined in some fixed way such that

(32) Constraint 1''': $P_1/T_1 \supset ((i)(P_i/T_2 \supset P_i/T_3))$, where $i \leq a \leq n$

This raises the interesting question of exactly how the 'shallow structure' P_a is to be defined. One possibility is that P_a is the output of the cyclical rules. However, there aren't enough facts known at present to settle the issue for certain. Still, we can draw certain conclusions. Passive, a cyclic rule, must be capable of applying before P_a, since Constraint 1''' must apply to the output of passive. Y-movement must apply after P_a.

Let us now consider this phenomenon with respect to the process of lexical insertion. Let us constrain the basic theory so that some notion of 'deep structure' can be defined along the lines discussed above. Let it be required that all lexical insertion rules apply in a block and that all upward-toward-the-surface cyclic rules apply after lexical insertion. Since passive is a cyclic rule and since passive must be able to apply before P_a is reached, it follows that if there is such a 'deep structure' all lexical insertion must occur before P_a is reached. Thus, it is an empirical question as to whether such a notion of 'deep structure' is correct. If there exist lexical items that must be inserted after P_a, then

such a notion of 'deep structure' would be untenable since there would exist upward-toward-the-surface cyclic rules (e.g., passive) which could apply before some case of lexical insertion.



Empirical question: Does there exist a lexical item that must be inserted between P_a and P_n? If so, then 'deep structure' does not exist.

Let us pursue the question somewhat further. Consider the sentences

(32) a. I persuaded Bill to date many girls.

b. I persuaded many girls to date Bill.

(33) a. I persuaded Bill not to date many girls.

b. I persuaded many girls not to date Bill.

(34) a. I didn't persuade Bill to date many girls.

b. I didn't persuade many girls to date Bill.

If we consider the meanings of these sentences, it should be clear that these cases work according to the two derivational constraints that we have stated thus far, except for that minority of speakers for whom these constraints do not hold, or who have somewhat different constraints.

The difference in the occurrence of not is crucial in these examples. In (33), not in semantic representation would occur inside the complement of persuade, while in the semantic representation of (34) not will occur in sentence above persuade. That is, in (33) persuade commands not in SR, while in (34), not commands persuade. This difference in the occurrence of not accounts for the fact that (33a) is unambiguous, while (32a) and (34a) are ambiguous. (32a) can mean either (35) or (36)

(35) There were many girls that I persuaded Bill to date.

(36) I persuaded Bill that the number of girls he dates should be large.

(34a) can mean either (37) or (38)

(37) There weren't many girls that I persuaded Bill to date.

(38) It is not the case that I persuaded Bill that the number of girls he dates should be large.

But (33a) can mean only (39)

(39) I persuaded Bill that the number of girls he dates should not be large.

The reason (39) is unambiguous is that since not precedes many in derived structure, not must command many in underlying structure (by Constraint 1"'). Since not originates inside the complement of persuade, and not must command many, many must also originate inside the complement of persuade. In (32a) and (34a), many may originate either inside the complement of persuade or from a sentence above persuade, which accounts for the ambiguity.

Now compare (33b) and (34b). In (33b) many both precedes and commands not in derived structure; therefore, many must command not in underlying structure. (33b) only has the reading of (40)

(40) There were many girls that I persuaded not to date Bill.

In (34b), on the other hand, not precedes many in derived structure and so must command many in underlying structure. So (34b) can mean (41) but not (40).

(41) There weren't many girls that I persuaded to date Bill.

Let us now consider the lexical item dissuade.

(42) a. I dissuaded Bill from dating many girls.

b. I dissuaded many girls from dating Bill.

In (42) the word not does not appear. The only overt negative element is the prefix dis-. Thus, the postlexical structure of (42) would have the negative element not inside the complement of -suade, but in the same sentence, as in (43).

(43) a. I NEG-suaded Bill from dating many girls.

b. I NEG-suaded many girls from dating Bill.

Moreover, the negative element would precede rather than follow the object of dissuade. In terms of precede and command relations, the postlexical structure of (42) would look like (34) rather than (33). Suppose P_a were postlexical, that is, suppose that the command relationship between not and

many in semantic representation were predictable from the precede relationship at some point in the derivation after the insertion of all lexical items. We would then predict that since NEG precedes many in (42), NEG must command many in the underlying structure of (42). That is, we would predict that the sentences of (42) would have the meanings of (34). But this is false. (42a) and (42b) have the meanings of (33a) and (33b).

Summary of Majority Dialect

Persuade --	32a	means	35, 36
Persuade not --	33a	means	39
	33b	means	40
Not persuade --	34a	means	37, 38
	34b	means	41
Dissuade --	42a	means	39
	42b	means	40

Dissuade means persuade - NP - not rather than not-persuade-NP, and the constraints on the occurrence of quantifiers in derived structure reflect this meaning, and must be stated prelexically. The lexical item dissuade must be inserted at a point in the derivations of (33) and (34) after Constraint 1"' has ceased to operate. Now recall that Constraint 1"' must operate on the output of the passive transformation. Consider (44).

(44) a. Many men weren't dissuaded from dating many girls.

b. Not many men were dissuaded from dating many girls.

c. I didn't dissuade many men from dating many girls.

(44) shows both the characteristics of (34) and (21) - (23). In (44), Constraint 1"' must operate both after the passive transformation and before the insertion of dissuade. Thus we have cases where an upward-toward-the-surface cyclic rule must apply before the insertion of some lexical item. This shows that any conception of 'deep structure' in which all lexical insertion takes place before any upward-toward-the-surface cyclic rules apply is empirically incorrect. It also shows that the passive transformation may apply to a verb before the overt lexical representation of the verb is inserted, which means that prelexical structures must look pretty much like postlexical structures. In the case of dissuade, one might be tempted to try to avoid such a conclusion with the suggestion that dissuade is derived by a relatively late transformation from a structure containing the actual lexical item persuade. Under such a proposal, dissuade would not be introduced by a rule of lexical insertion, but rather by a rule which changes one actual lexical item to another. Such a solution is untenable since lexical items like prohibit, prevent, keep, forbid, etc., which do not form pairs like persuade-dissuade, work just like dissuade with respect to the properties we have discussed.

A particularly tempting escape route for those wishing to maintain a level of 'deep structure' might be the claim that the lexical item dissuade is inserted precyclically, that dissuade requires a not in its complement sentence, and that this not is deleted after shallow structure. Thus, the not would be present at the time that the constraints shut off, and all of the above facts would be accounted for. This proposal has some initial plausibility since similar verbs in other languages often have a negative element that appears overtly in its complement sentence. For example, in Latin we have "DissuasI Marcō nē iret" (I dissuaded Marcus from going), where nē, the morphological alternant of nōn in this environment, occurs in the complement sentence.

Let us suppose for the moment that such a solution were possible. This would mean that the complement sentence of dissuade would contain a not at the level of shallow structure, but not at the level of surface structure. Now consider the following sentences:

(45) I dissuaded Mary from marrying no one.

(46) *I persuaded Mary not to marry no one.

(47) *Mary didn't marry no one.

(48) I didn't persuade Mary to marry no one.

(46) and (47) are ungrammatical in standard English, and in all dialects if the two negatives are both considered as underlying logical negatives (e.g., if (47) has the reading It is not the case that Mary married no one). As is well-known, this prohibition applies only for negatives in the same clause (cf. example (48), where the negatives are in different clauses). The question arises as to where in the grammar the No-double-negative (NDN) constraint is stated. (a) If 'deep structure' exists, it could be stated there; (b) it could be stated at shallow structure; (c)it could hold at all levels between deep structure and shallow structure; or (d) it could hold only at surface structure.

Now consider (45). Under the above proposal for post-shallow-structure deletion of not in dissuade-sentences, not would still be present at the level of shallow structure. (In fact, it would be present at all points between deep structure and shallow structure, that is, all points in the derivation where constraints 1 and 2 hold.) At the level of shallow structure, (45) would have the form:

(45) I dissuaded Mary from not marrying no one.

Thus under the above proposal, the No-double-negative (NDN) constraint could not hold at the level of shallow structure, or at any previous point in the derivation back to deep structure; if it did, (45) would be ruled out. Thus, under the proposal for post-shallow-structure not-deletion, the NDN-constraint could only be a constraint on surface

structure, not on shallow structure. Let us now take up the question of whether this is possible.

The following sentences are in accord with the NDN-constraint, whether it holds at shallow or surface structure, since these sentences have essentially the same representation at both levels.

(49) Max said that Sheila Weinstein was spurned by no one.

(50) Max didn't say that Sheila Weinstein was spurned by no one.

(51) *Max said that Sheila Weinstein wasn't spurned by no one.

Note that in the appropriate dialects, Y-movement can apply to sentences of this form moving the participial phrase of the embedded clause.

(52) Max said that Sheila Weinstein wasn't spurned by Harry.

(53) Spurned by Harry, Max said that Sheila Weinstein wasn't.

In (53), Harry has been moved from a position where it was in the same clause as n't to a position where it is in a higher clause. Now consider once more:

(51) *Max said that Sheila Weinstein wasn't spurned by no one.

If the NDN-constraint holds for surface, not shallow, structures, then the application of Y-movement to (51) would move no one out of the same clause as n't and should make the NDN-constraint nonapplicable at surface structure. Thus, (54) should be grammatical.

(54) *Spurned by no one, Max said that Sheila Weinstein wasn't.

But (54) is just as bad as (51) -- the NDN-constraint applies to both. As we saw, the NDN-constraint cannot apply to the surface structure of (54), since no one is not in the same clause as n't, having been moved away by Y-movement. Thus, in order to rule out (54), the NDN-constraint must apply before Y-movement; that is, it must apply at the level of shallow-structure.

But this contradicts the post-shallow-structure not-deletion proposal, since under that proposal, (45), which is grammatical, would contain two negatives in the same clause at the level of shallow structure (cf. (45')). Thus, the post-shallow-structure not-deletion proposal is incorrect.

A similar proposal might say that prior to shallow structure, from replaced not with verbs like dissuade and prevent, while to replaced not with a verb like forbid, and that from and to 'acted like negatives' (whatever that might mean) with respect to constraint 1 and 2.at shallow structure and above. However, it is clear from sentences like (45) that from does not act like a negative with respect to the NDN-constraint at shallow structure. Such a proposal would thus require from both to act like a negative and not to act like a negative, a contradiction. Neither of these routes provides an escape from the conclusion that dissuade must be inserted after shallow structure.

Another route by which one might attempt to avoid this conclusion would be by claiming that derivational constraints did not operate on the internal structure of lexical items, and that therefore dissuade could be inserted before the passive, preserving the notion of 'deep structure'. It would then not interact with the constraints. Such a claim would be false. Dissuade does interact with the constraints. As the summary of meaning-correspondences given above shows, dissuade acts just like persuade not, not like persuade. In particular, (42a) is unambiguous, just like (33a). It only has a reading with many originating inside the complement, as the constraints predict. If dissuade were impervious to the constraints, then we would expect it to act like persuade; in particular, we would expect (42a) to be ambiguous, just like (32a), where many may be interpreted as originating either inside or outside of the complement. But as we have seen, (42a) does not have the outside reading. John dissuaded Bill from dating many girls cannot mean Many were the girls who John dissuaded Bill from dating.

It should be noted that this argument does not depend on the details of Constraint 1"' being exactly correct. It would be surprising if further modifications did not have to be made. Nor does this argument depend on any prior assumption that semantic representation must be taken to be phrase-markers, though the discussion was taken up in that context. It only depends on the facts that persuade-not and not-persuade obey the general constraints on quantifiers and negatives, and that dissuade acts like persuade-not. Thus, in any version of transformational grammar there will have to be stated a general principle relating semantic representations of of sentences containing quantifiers and negatives to the left-to-right order of those corresponding quantifiers and negatives in 'derived structure'. If the general principle is to be stated, the notion 'derived structure' will have to be defined as following the passive rule, but preceding the insertion of dissuade. Thus, in no non-ad-hoc transformational grammar which states this general principle, will all lexical insertion precede all cyclical rules.

Let us sum up the argument.

(i) Suppose "deep structure" is defined such that all lexical insertion rules apply in a block and precede the application of any upward-toward-the-surface cyclic rules.

(ii) Evidence was given for derivational constraints 1 and 2, which relate semantic command-relationships to precede- and command-relationships at some level of 'derived structure.'

(iii) The appropriate level of 'derived structure' must (a) precede surface structure, (b) follow the passive transformation, and (c) precede the insertion of lexical items such as dissuade, prohibit, prevent, keep, etc.

(iv) Since passive is an upward-toward-the surface cyclic rule, (iii) shows that the concept of 'deep structure' given in (i) cannot be maintained.

Note

At the conference at which this paper was presented, Paul Postal made the following observation: After more than a decade of research on English syntax, there are still very few solid arguments for rule ordering, and those that are solid usually involve cases where one rule is postcyclical and therefore must follow all cyclical rules, or cases where one rule must apply later than another, but where this will follow from the cyclical principle given the correct underlying structure. Of course, a multitude of arguments for rule-ordering have been presented, but the vast majority of these are shaky arguments since they depend on inadequate analyses. Very often it turns out that one can state a rule one of several ways, with no empirical consequences attached to the choice, except that the different choices will involve different rule orderings, any one of which is acceptable. Other rule ordering arguments disappear when derivational constraints are considered. Such situations, Postal observed, are surely an artifact of our present inadequate notion of linguistic structure. He suggested that the following may eventually turn out to be the case: Some rules are cyclical; others are postcyclical (shallow structure would be defined as the output of the last application of a cyclical rule). Beyond this no global derivational constraints on rule ordering would have to be given, though other kinds of global derivational constraints would replace some of them, for example, the output conditions for pronominalization mentioned in Lakoff, Pronouns and Reference. The application of rules in their correct order would be given entirely by the principle of the cycle given an adequate account of underlying structures, and by the postcyclical grouping of other rules. At present, this is simply speculation, but it is well worth looking into.

References

[1] Chomsky, Noam, "Deep Structure, Surface Structure, and Semantic Interpretation," to appear in [4].

[2] Chomsky, Noam and Morris Halle, The Sound Pattern of English, Harper and Row, 1968.

[3] Jackendoff, Ray, "An Interpretive Theory of Negation," M.I.T. mimeo, 1968.

[4] Jacobovits, L. and D. Steinberg, Semantics: An Interdisciplinary Reader in Philosophy, Psychology, Linguistics, and Anthropology, University of Illinois Press, forthcoming.

[5] Lakoff, George, On the Nature of Syntactic Irregularity, Indiana University dissertation, 1965.

[6] Lakoff, George, "Pronouns and Reference," Harvard mimeo, 1968.

[7] Lakoff, George, "Repartee", Foundations of Language, 1969.

[8] Lakoff, Robin, Abstract Syntax and Latin Complementation, M.I.T. Press, 1968.

[9] Partee, Barbara H. "Negation, Conjunction, and Quantifiers," Foundations of Language, 1969.

[10] Postal, Paul, The Cross-over Principle, IBM mimeo, 1968.

[11] Ross, J. R., Constraints on Variables in Syntax, M.I.T. dissertation, 1967.