Paul M. Postal

A Global Constraint on Pronominalization

I. Background*

In several recent works,[1] Lakoff has argued that grammars contain a class of rules which he has referred to as *Global Derivational Constraints*. Briefly, these are rules which filter out certain otherwise well-formed derivations on the basis of properties of two or more distinct structures[2] in the derivation. A crucial notion in the formulation of such rules is that of *corresponding constituent*, which, roughly, defines the various ancestors of a particular constituent in surface structure. The claim is that Global Derivational Constraints are defined on *n*-ads of constituents (hopefully restricted to a maximum of 2-ads) in particular trees of derivations and on the corresponding constituents of these in one, two, or perhaps more other trees. Thus such rules have a form which might be schematized very informally and imprecisely as follows:[3]

(1) Mark as ill-formed any derivation in which constituents A and B in Tree_i have properties $P_1 \ldots P_k$ and $Q_1 \ldots Q_l$ respectively, and the corresponding constituents of A and B (call them A' and B' respectively) in Tree_r have properties $R_1 \ldots R_m$ and $S_1 \ldots S_n$ respectively.

Thus the claim is that there are nontransformational, filtering rules which must be defined on at least pairs of trees, with the further claim that in many cases, if not all, these trees are not contiguous in the derivation.

As with every type of rule in grammatical theory, it is necessary in a serious account to propose the narrowest possible constraints on the class of rules to be allowed, if empirical content is to be retained. At least three distinct aspects of Global

* The present paper was written in April, 1970 and presented at the Conference on Formal Methods in Linguistics at Rocquencourt, France. A somewhat fuller version will appear in the proceedings of this conference, *Formal Analysis of Natural Languages*, to be edited by M. Halle, M.-P. Schützenberger, and M. Gross. I am grateful to the latter for kindly permitting earlier publication of the present version in this journal. I would also like to thank W. J. Plath and J. R. Ross for many helpful criticisms of an earlier version of this work.

[1] Lakoff (1969; 1970; to appear). Cf. also Casagrande (1970) and Postal (to appear).

[2] A problem arises with the generality of this statement because a significant class of constraints which one wishes the notion to cover seem to involve not so much particular structures, but rather the fact that these structures have been derived by some particular transformational rule. This is especially clear in the case of lexical exceptions to various transformations, but also appears to hold in many other cases.

[3] This schema is representative, of course, only of the subset of constraints definable on exactly two levels of structure.

Derivational Constraints can be considered in this regard, namely (a) the number of structures which can be referred to, (b) the number of constituents which can be referred to, and (c) the properties of the constituents which can be mentioned. With regard to the first consideration, we can tentatively hope that some very small maximum, say three or four structures, will suffice. Further, we might suspect that in all cases where there are more than two structures, one of the additional levels is the semantic representation. As for the number of constituents, so far all plausible cases seem to be such that a maximum of two is not in conflict with the empirical facts. It will be seen below that the constraint suggested here meets this condition. Finally, there are the properties which can be referred to, that is, the $P_1 \ldots P_n$, etc. in (1). Here a natural initial constraint would be that no property can be referred to in Global Derivational Constraints that is not available for the formulation of transformational rules. Thus such relations as *domination*, *command*, *to the left of*, *directly before a constituent of type A*, etc. would be available. Without going into these matters further here, it can be seen that the problem of restricting the class of Global Derivational Constraints is not different in kind from the one for transformational rules. In each case, the initial general statement of the rule definitions permits an excessively wide class of rules, the overwhelming mass of which are found in no language. Hence empirical work must move in the direction of restricting both types of rule.

Certain writers have, however, questioned the need for rules like those just described. Chomsky (to appear c), for example, is quite explicit in denying the existence of Global Derivational Constraints. This matter is of considerable importance because every fact that could be described by what have been called Interpretive Rules can be stated in terms of Global Derivational Constraints, but not conversely. Consequently if it can be shown that Global Derivational Constraints exist, it is thereby shown that Interpretive Rules in Chomsky's sense are not a real feature of human language.

A quotation will indicate in detail Chomsky's position on this matter (to appear c, 78):

> As to constraints internal to derivations, the revised theory [that is, Chomsky's position: PMP] maintains that there are none (beyond transformations and conditions on transformations), whereas generative semantics, in the formulations considered here, suggests that they are freely permitted. Since there seem to be no plausible cases, the narrower theory is to be preferred. As to constraints relating semantic representations and derivations, the revised standard theory holds that there is only one category of such "constraints"; namely, certain specific aspects of surface (or shallow) structure are relevant to semantic interpretation. Generative semantics again permits richer devices. Every rule of interpretation mapping surface (or shallow) structure into some aspect of meaning can be described as a "derivational constraint" (i.e., a "rule of grammar"),[4] but not conversely. Unless examples are presented to justify the extra

[4] This remark is simply false, since, as noted earlier, there are a variety of natural constraints on the class of Global Derivational Constraints which sharply separate them from the class of all possible rules. Moreover,

wealth of theoretical devices permitted in generative semantics, the narrower theory is again to be preferred, for reasons discussed in §2.2. There seems to me to be no such justification.

II. A Class of Restrictions on Pronominalizations

The facts I want to consider are a class of constraints on pronominalization, where by pronominalization I refer to relations between some antecedent nominal and a pronoun with which it is coreferential.[5] The constraints to be discussed are common to many speakers of English but by no means to all. I have discussed the same facts earlier in work on so-called "crossing" restrictions (cf. Postal 1968), and one of the many inadequacies of the earlier treatment was a failure to recognize that some of the "crossing" restrictions are not universal for English speakers. The restrictions to be discussed here are some of those involving *wh*-marked nominals, that is, involving the words *who, what, which, whose, when, where*, etc. I will indicate coreference in examples by identical subscripts. A star (*) preceding an example means that the sentence does not have the interpretation implied by the indices.

Consider now (2)–(4):

(2) a. *Who$_i$ did Mary say his$_i$ mother yelled at?
b. *What$_i$ did the fact that it$_i$ got rusty prove was no good?
c. *What official$_i$ did the policeman who arrested him$_i$ claim was drunk?
d. *Which tigers$_i$ were their$_i$ keepers eaten by?

(3) a. *The child$_i$ who$_i$ Mary said his$_i$ mother yelled at was crying.
b. *The car$_i$ that$_i$ the fact that it$_i$ got rusty proved was no good is outside.
c. *The official$_i$ who$_i$ the policeman who arrested him$_i$ claimed was drunk was drunk.
d. *The tigers$_i$ that$_i$ their$_i$ keepers were eaten by are not hungry just now.

Chomsky himself does not consistently maintain the view that the theory of Global Derivational Constraints is so unstructured as to be empty. Thus at several points in the same paper he argues that such constraints are unnecessary in an empirically adequate theory of grammar, arguments which are evidently not equivalent to the claim that rules per se are unnecessary. This shows that he distinguishes the class of rules under discussion from arbitrary grammatical principles. Finally, it seems inappropriate for a supporter of Interpretive Rules to criticize the precision and degree of restriction presently available in other theories of linguistic rules, for this interpretive theory is almost completely open ended. Chomsky's two papers on the subject (to appear b; to appear c) do not provide any general characterization of this class of rules, or any limitations on their possible descriptive power. There is general vagueness as to whether Interpretive Rules are "building" rules, "filtering" rules, or some combination of both; whether both types are to be allowed; what class of formal structures such rules take as input; what types they yield as output; and so forth.

[5] One must actually speak more precisely of a particular kind of coreference, which I shall call *stipulated*. This manifestation of coreference, which underlies pronominalization in general, must be distinguished from both *asserted* coreference and *inferred* coreference, which are found respectively in such examples as:

(i) *John* is *the one who rang the bell.*
(ii) *The boy* scratched *his real father's only son.*

Thus the italicized NPs in (i) are asserted by the speaker to designate the same individual, while in (ii) it is a logical inference from the meanings of the italicized NPs that they designate the same individual.

That these contrasts are not only real but relevant to the subject under discussion will be noted below.

(4) a. *The child$_i$, who$_i$ Mary said his$_i$ mother yelled at, was crying.
b. *The car$_i$, which$_i$ the fact that it$_i$ got rusty proved was no good, is outside.
c. *The official$_i$, who$_i$ the policeman who arrested him$_i$ claimed was drunk, was drunk.
d. *The tigers$_i$, which$_i$ their$_i$ keepers were eaten by, are not hungry just now.

The only relevant difference among (2), (3), and (4) is that in the first the *wh* forms are in question clauses, in the second in restrictive relative clauses, and in the third in appositive relative clauses. There are, however, no differences whatever among the pronominalization constraints in these cases, and as far as I know, this fact is completely general: the restrictions are the same for question clauses, restrictive relative clauses, and appositive relative clauses. Hence, in subsequent examples I shall restrict myself to one type of example without loss of generality.

On the basis of such facts, one might simply wish to claim that *wh* forms cannot be the antecedents of coreferential pronouns. But this is, of course, quite false, as (5) shows:

(5) a. Who$_i$ do you think said Mary yelled at his$_i$ mother?
b. The car$_i$ which$_i$ Bill said got rusty later developed a hole in its$_i$ exhaust pipe?
c. What official$_i$ claimed he$_i$ had been roughed up by the police?
d. Which tigers$_i$ did you claim indicated that they$_i$ wanted a better grade of keeper?

It follows that no general constraint banning a stipulated coreference relation between any *wh*-form and a pronoun is possible. The grammar of the dialect or dialects which do not accept the coreference linkages in sentences like (2)–(4) must contain some constraint that distinguishes cases of this type from constructions like (5), where the linkages are permitted.

What I shall now try to show is that this constraint (for this dialect group) is (i) naturally formulated as a Global Derivational Constraint, but (ii) not formulatable in any other way that I can see. In particular, I shall show: (iii) that the constraint is not definable on deep structure in Chomsky's sense; (iv) that the constraint is not definable on surface structure; (v) that the constraint is not definable on shallow structure; (vi) that the constraint is not definable on the structures which are input to the rules which move *wh*-marked nominals to clause initial positions; but (vii) that the constraint is definable on the pair of structures which are input and output of the rules which move *wh*-marked nominals.

Let us begin by giving an initial, theory-independent[6] characterization of the difference between examples like (2)–(4), which contain violations, and those like (5),

[6] That is, independent of theory at the level of theoretical dispute toward which the article is directed. Naturally, no characterization can be independent of all theory, since any descriptive vocabulary involves at least an implicit theory by its choice of units, possible relations, and so forth.

which manifest proper coreference linkages. It will be seen later that this description is oversimplified, but it suffices for this stage of the discussion. The characterization is simple enough. In each example of (2)–(4), where the linkage is ill-formed, the *wh* form is positioned (at the point where the *Wh* Movement rules would apply) *to the right of* the pronoun with which a coreference linkage is ill-formed. Hence at this pre-*Wh* Movement stage the sentences of (2) would have the following underlying forms:

(6) a. [Mary said his$_i$ mother yelled at who$_i$]
b. [the fact that it$_i$ got rusty proved what$_i$ was no good]
c. [the policeman who arrested him$_i$ claimed what official$_i$ was drunk]
d. [their$_i$ keepers were eaten by which tigers$_i$]

On the other hand, at the directly pre-*Wh* Movement stage for the sentences of (5), where the coreference linkages are proper, the *wh* form is in each case *to the left of* the pronoun. Thus the sentences of (5) would have at this stage the structures:

(7) a. [you think who$_i$ said Mary yelled at his$_i$ mother]
b. [the car$_i$ Bill said which$_i$ got rusty later developed a hole in its$_i$ exhaust pipe]
c. [what official$_i$ claimed he$_i$ had been roughed up by the police]
d. [you claimed which tigers$_i$ indicated that they$_i$ wanted a better grade of keeper]

In fact, the difference between having the pronoun to the left or right of the *wh* form at the stage immediately before the operation of the *Wh* Movement rules properly separates not only cases like (2)–(4) from those like (5) but properly separates all similar sets. This is, in fact, the defining characteristic of the contrast under analysis. In the dialect group being described, sentences like (2)–(4) always have ill-formed coreference linkages if, at the stage immediately before application of the *Wh* Movement rules, the pronoun is to the left of the *wh* form.

III. How the Constraint Can *Not* Be Stated

A. Remarks

We now must examine carefully the theoretical character of this constraint, asking how it can be stated (relatively)[7] precisely within an integrated grammatical framework, thus determining what, if anything, it can teach us about the general form of grammar.

The first thing to stress is that the constraint which precludes coreference linkages in sentences like (2)–(4), a constraint so far describable as a ban between such linkages in those cases where a *wh* form is to the right of the pronoun at the directly pre-*Wh* Movement stage, does not follow from any more general constraints on pro-

[7] Clearly, no one is in a position to state any rule with full precision without accepting a vast amount of arbitrariness.

nominalization, either for English in general, or for the certain dialect group in particular. Thus backwards, or right-to-left pronominalization (that is, cases where the pronoun is to the left of its antecedent) is in general permitted in a wide class of contexts like those in (6), where the antecedent is not a *wh* form. To see this, one need only note the well-formedness of sentences like those in (8), which manifest the analogous coreference linkages:

(8) a. Mary said his$_i$ mother yelled at Max$_i$.
b. The fact that it$_i$ got rusty proved the car$_i$ was no good.
c. The policeman who arrested him$_i$ claimed a certain official$_i$ was drunk.
d. *Their$_i$ keepers were eaten by the tigers$_i$.[8]
e. Their$_i$ keepers' wives were eaten by the tigers$_i$.

Hence we are discussing a special constraint[9] that seems to block backwards pronominalization at the pre-*Wh* Movement stage for *wh* forms.[10] Let me refer to it for convenience as the *Wh Constraint.*

[8] One must give (8e) here because of some constraint about relative depth of embedding which rules out examples like (8d) in such cases. This constraint has no bearing on the present discussion.

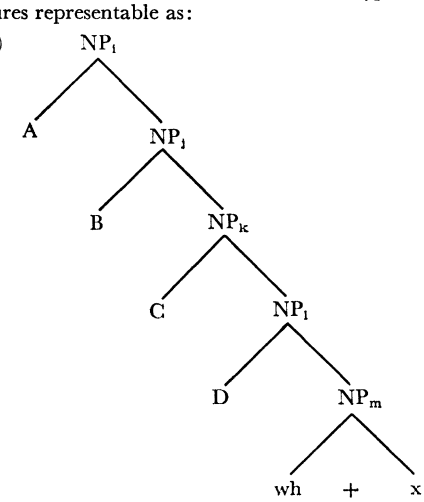
[9] A proposal made verbally by R. Jackendoff several years ago at one of the La Jolla conferences on English syntax would seek to reduce the constraint in question to a more general restriction blocking backwards pronominalization in certain cases. Cf. Appendix C for discussion.

[10] The statement in the text is inaccurate in at least one important respect. The restrictions in question govern not only *wh*-marked nominals themselves but also certain "larger" nominals in which these are embedded. Hence alongside of (i) one finds also (ii)–(vi):

(i) *Who$_i$ did his$_i$ mother criticize?
(ii) *Whose$_i$ mother did his$_i$ sister criticize?
(iii) *Whose$_i$ mother did the picture of him$_i$ annoy?
(iv) *Whose mother$_i$ did the picture of her$_i$ annoy?
(v) *Whose mother's$_i$ lover did the picture of her$_i$ annoy?
(vi) *Whose friend's$_i$ father did the claim that he$_i$ was a communist fail to disturb?

I attempted to handle such facts in earlier work (Postal 1968) in terms of Ross's (1967) notion of Pied Piping, but I now see this was a complete mistake. The proper treatment must, I believe, simply refer in a statement like that in the text to *wh* forms and to certain types of nominal in which they are embedded. Hence roughly, in structures representable as:

(vii)



In the revised standard theory advocated by Chomsky, this constraint will have to be formulated in terms of Interpretive Rules. Chomsky seems to suggest in the paper quoted from above that coreference in general is to be described in terms of surface structure or shallow structure (Chomsky, to appear c, 80):

> Many aspects of meaning (scope and order of logical elements, coreference, topic, focus and certain types of presupposition) seem to involve surface (and perhaps shallow) structure in an essential way.

However, other interpretive treatments, in particular that of Jackendoff (1969), have suggested that the relevant Interpretive Rules for coreference might have to apply cyclically. Hence they would apply not only to points in a derivation thus far distinguished by linguists (deep structure, shallow structure, surface structure) but also to intermediate structures as well. I think that this cyclical idea can be disconfirmed.[11] Here, however, I will restrict myself to showing that the *Wh* Constraint cannot be defined over any single distinguished point in derivations, that it cannot be defined over any set of such distinguished points, and that, in fact, proper statement of it

---

where this schema covers nominals like those in (ii)–(vi), i.e. those of the form [whose(+NP+genitive(+NP+genitive+(...))+Noun], the restrictions on the *wh*-marked nominal NP_m apply as well to NP_l, NP_k, NP_j, NP_i, etc. What is crucial is to distinguish cases like (ii)–(vi) from those such as:

(viii) Which one of his_i friends did his_i mother say I criticized?
(ix) Which one of $them_i$ did $their_i$ parent's dog bite?

There are many possible structural and semantic properties which could be appealed to for the proper differentiation, and at this point I will make no proposal as to the correct basis. It can be noticed though that in cases like (ii)–(vi), which obey the constraints on simple *wh*-marked nominals, the *wh*-marking is not found on the head noun of the whole "maximal" nominal, while in cases like (viii) and (ix), which do not obey the constraint, the *wh*-marking is on the head noun of the "maximal" nominal.

[11] I am referring here, not to arguments like those in Postal (1968), which depend on assumptions no longer valid once Interpretive Rules can be appealed to, but rather to facts like the following. Any approach to pronominalization must be such that the italicized nominals in both (i) and (ii) can be marked as coreferents:

(i) People who have met *him* will tell you *Max* is insane.
(ii) People who have met *Max* will tell you *he* is insane.

Similarly, any approach to pronominalization must allow the italicized nominals in both (iii) and (iv) to be marked as coreferents:

(iii) *He* claims that people who have met *him* will tell you Mary is insane.
(iv) *He* claims that people who have met Mary will tell you *he* is insane.

Consider now, under the cyclical assumption, the situation where the structures in (i) and (ii) are embedded immediately below *he claims*, yielding:

(v) he claims that (people who have met him will tell you Max is insane)
(vi) he claims that (people who have met Max will tell you he is insane)

Under the cyclical assumption, the marking rules must apply to the complement first, with the results noted for (i) and (ii), that is, *him* and *Max* can be marked as coreferents in (v), *Max* and *he* as coreferents in (vi). Now we move to the next cycle, where, as the sentences in (iii) and (iv) show, *he* and *him* in (v) and *he* and *he* in (vi) must be markable as coreferents. But now the grammar has claimed that in sentences like (v) the triple of nominals *he*, *him*, *Max* can all be coreferents, and that in (vi) the triple of nominals *he*, *Max*, *he* can all be coreferents. But this is false in both cases. Such examples, which are infinitely extendable to cases where the complement is not immediately below the pronoun with which linkages cannot be established but as far below it in embedding as desired, show that a cyclical assignment of coreference linkages is by itself inadequate.

requires reference to two distinct trees in the derivation in addition to the semantic representation.

B. The Wh *Constraint Is Not Definable on Surface Structures*

That the *Wh* Constraint cannot be described properly in terms of surface structure is shown most easily and trivially by the fact that the relevant restrictions manifest themselves even in restrictive relative clauses from which the offending *wh* form has been deleted. Hence the parallelism between the (a) and (b) forms in (9) and (10) is perfect:

(9) a. *The astronaut$_i$ who$_i$ his$_i$ mother claimed was a psychotic was arrested.
b. *The astronaut$_i$ his$_i$ mother claimed was a psychotic was arrested.
(10) a. *An engineer$_i$ who$_i$ Mary claimed his$_i$ mother refuses to let out after dark arrived.
b. *An engineer$_i$ Mary claimed his$_i$ mother refuses to let out after dark arrived.

This shows that the *Wh* Constraint cannot be generally stated at the level of surface structure, since one of the elements necessary for specifying the constraint is not even present in some surface structures which manifest the restrictions. In addition, the arguments given in the next section to show that the constraint is undefinable in terms of shallow structure apply ipso facto to surface structure as well.

C. The Wh *Constraint Is Not Definable on Shallow Structures*

The term *shallow structure* is due to the present writer. It is intended to refer to a hypothetical level of grammatical structure more abstract than surface structure, but relevant for many grammatical statements. In particular, this level would be defined before such intuitively "late" rules as Topicalization or Υ Movement,[12] which apply in the derivation of structures like (11b):

(11) a. Max claimed that Sally won't go near Bill.
b. Bill, Max claimed that Sally won't go near.

Chomsky (to appear c) uses the term to refer to structures that appear prior to the application of what Emonds (1969) has called *root transformations*. This means at least that the structures so designated are those which exist after all but the last cycle of transformational rules has applied. It is clear, then, that both Chomsky and the present writer take this level to be less abstract than the one at which *Wh* Movement rules generally apply. This follows since the *Wh* Movement rules, at least in the case of relative clauses, apply at many different levels of structure and cannot be regarded as last cyclical.

[12] The former term was used in Ross (1967), the latter in Postal (1968).

It is certainly possible and in fact necessary to define shallow structure so that the rule which deletes relative pronouns, forming structures (9b) and (10b), applies at a point subsequent to the specification of shallow structure. In that way, the argument of Section B would not be applicable to shallow structures.

The question is, therefore, whether the *Wh* Constraint can be stated on a single level of shallow structure, defined subsequent to the derivation of structures by the *Wh* Movement rules. It is not hard to see that no general characterization at such a level is possible, for at this point, the relevant *wh* forms will be to the left of all the pronouns involved, both those with which coreference linkages are forbidden, as in (2)–(4), and those with which such linkages are proper, as in (5). Hence it would be necessary to propose some principle defined on such structures which distinguishes those pronouns to the right which are appropriately related from those which are not. But the fact that no such principle is available follows from the fact that the true generalization in this case must refer to just those pronouns which were to the left of the *wh* forms *before* these were transported by the *Wh* Movement rules. Thus there can be no other general principle which distinguishes between cases like (12a, b), (13a, b), and so forth:

(12) a. Who$_i$ did the police say persuaded her to meet his$_i$ sister?
b. *Who$_i$ did the police say his$_i$ sister persuaded her to meet?
(13) a. Who$_i$ did you testify Joan convinced to say that a girl he$_i$ knew was present at the accident?
b. *Who$_i$ did you testify Joan convinced a girl he$_i$ knew to say was present at the accident?

Any attempted formulation of a principle defined on post-*Wh* Movement structures to distinguish between cases like (12a, b) must simply reconstruct the difference between pronoun position to the left of the *wh* form and pronoun position to the right of the *wh* form in the pre-*Wh* Movement structures. Thus, a putative principle to accomplish this would have to infer from the structure of sequences like (12a) that the word *who$_i$* was to the left of the pronoun *his$_i$* in the input to the *Wh* Movement rules, while in the case of (12b) it would be necessary to infer that the word *who$_i$* was to the right of the pronoun *his$_i$* at that stage. But the only way such inferences could be constructed would be by an analysis of the clause structures intervening between the *wh* form and the relevant pronoun in the post-*Wh* Movement structures. Such an analysis might, for example, attempt to specify those positions in which a nominal is obligatory in underlying structures, thus determining a position from which the *wh* form has come. Thus, in the case of (12b), it might be determined that with a singular subject the verb *meet* requires an object nominal. Since *meet* has no object nominal in the post-*Wh* Movement structures in (12b), it could be inferred that the word *who* is just this object.

I am not claiming that it would be impossible to formulate principles of this sort;

however, the objections are obvious and fundamental. First, the complexity of the proposed approach is extraordinary. It requires a complete analysis of the structure of every possible type of clause which can intervene between *wh* form and pronoun in contrasting pairs like (12). Since every complement clause type can so intervene, there is no bound on the complexity of intervening structures. Hence the principle of coreference assignment, if stated at this post-*Wh* Movement level, would have to include structural analyses of the full class of complement clause structures.[13] The complexity is the least problem, however: notice that in trying to formulate this principle we have in fact not succeeded in stating it on post-*Wh* Movement structures, but have had to appeal to the underlying structures of these, for example, the question whether or not a verb like *meet* requires an object under certain circumstances, and so forth. Hence the principle given is not stated over post-*Wh* Movement structures. The only way to make the formulation precise would seem to involve Global Derivational Constraints.[14] It is thus clear on these grounds that the proper generalization distinguishing pairs like (12a, b), and so forth, cannot be stated on post-*Wh* Movement structures. But since shallow structures are post-*Wh* Movement structures, it follows that the *Wh* Constraint cannot be stated on shallow structures alone.

D. The Wh *Constraint Is Not Definable on Deep Structures*

By *deep structures* I refer to a level like that posited by Chomsky, distinct from semantic representation. As one who accepts the framework of generative semantics as the currently best overall grammatical scheme, I do not believe in the existence of such deep structures. The demonstration that the *Wh* Constraint cannot be defined over deep structures is essentially trivial, and follows from the fact that the constraint differentiates among structures with essentially identical deep structures, according to whether or not certain rules have applied to them prior to the application of the *Wh* Movement rules. Thus, in a framework like Chomsky's, which accepts a transformation Passive,[15] sentences such as (14a, b) have deep structures that are identical in all respects relevant for the purposes of this discussion, namely they have the same nominal order.

[13] Ambiguities can arise to complicate the picture even further. Thus notice that in:

(i) Who did Carswell want to eat?

the *wh* form could have originated as either the subject of *eat* (with deleted unspecified object) or as the object of *eat* (with subject removed by deletion under coreference to higher verb subject). But such ambiguities are, naturally, associated with different coreference linkage possibilities in case a pronoun intervenes between the original possible positions of the *wh* form and its final location. Hence such sentences as (ii) have two sets of readings.

(ii) Who did Mary expect to promise his mother to eat?

According to one set, *who* is interpreted as the subject of *promise*, according to the other, *who* is interpreted as the object of *eat*. However, only the former set of readings contains members in which *who* and *his* are coreferents.

[14] But in this case, erroneous ones of incredible complexity—erroneous because there is an obvious alternative referring to relative positions of pronouns and *wh* forms.

[15] The argument is unaffected if Passive is broken up into two (or more) other rules which accomplish the same mappings, as in Chomsky (to appear a), Postal (1968).

(14) a. Who$_i$ did Joan claim visited his$_i$ mistress?
b. *Who$_i$ did Joan claim his$_i$ mistress had been visited by?

The deep structures in this case might be schematized as (15):

(15) [Joan claimed who visited his mistress]

But since the *Wh* Constraint differentiates between (14a, b) even though they have the same deep structures, it follows that the *Wh* Constraint is not definable over the class of deep structures.

The same point emerges for pairs like (16a, b):

(16) a. Who$_i$ did Joan talk to about his$_i$ illness?
b. *Who$_i$ did Joan talk about his$_i$ illness to?

These would again have a single deep structure of the form:

(17) [Joan talked to who about his illness]

with (16b) being derived by an optional rule which shifts the relative order of prepositional phrases, a rule which, like Passive, precedes the *Wh* Movement rules.

E. *The* Wh *Constraint Cannot Be Defined Over Any Set of the Levels (Deep Structure, Shallow Structure, Surface Structure)*

In previous sections I have shown that the *Wh* Constraint cannot be defined over any distinguished level of structure in Chomsky's theory, or, in fact, over any single level in any other theory as well. In this section, I will show that taking the distinguished levels in pairs or triples also does not overcome the limitations of each distinguished level separately.

Consider first the pair (shallow structure, surface structure). Clearly, the information afforded by either one of these is of no aid to the other in providing a statement of the constraint. Since both levels are post-*Wh* Movement, and the discussion of shallow structures in Section C shows that post-*Wh* Movement statement of the constraint in question is impossible, this pairing must likewise fail to provide a proper treatment.

Next take the pair (shallow structure, deep structure). This conjunction is also of no avail. For, as seen in Section D, deep structure does not indicate whether a *wh* form precedes or follows a pronoun in the input to the *Wh* Movement rules. But, as we saw in the discussion of shallow structures, just this information is required. Consequently, making available the information represented in deep structures does not overcome the limitations of shallow structure for stating the *Wh* Constraint. From the fact that the constraint cannot be stated over the pair (shallow structure, deep structure), it immediately follows that it cannot be stated over the triple (surface structure, shallow structure, deep structure). Obviously, the pair (surface structure, deep structure) does not contain the required information either.

It follows that no set of the distinguished levels in Chomsky's revised standard theory, or in any other theory of transformational grammar, provides the structure for defining the *Wh* Constraint.

IV. The Statement of the *WH* Constraint

A. The Constraint As Specified on the Input to the Wh *Movement Rules*

In Section III, I showed that the *Wh* Constraint is unstatable over any combination of levels distinguished in any type of transformational grammar (in particular, in the revised standard theory advocated by Chomsky). In Section II, I provided a tentative formulation of the constraint, defined over the level of structure that provides the input to the rules of *wh* movement. If such a formulation sufficed, this would be a small argument in favor of generative semantics and against the revised standard theory, since the former allows derivational constraints statable on particular single levels of structure defined by particular rules, or to pairs consisting of such single levels and the semantic representation, and the latter does not.[16] That is, I provided a formulation of the constraint which can be expressed as (18:)

(18) Mark as ill-formed all derivations in which there is a pronoun *A* to the left of a *wh* form *B* in those structures which are possible inputs to the *Wh* Movement rules, and in which *A* and *B* are coreferents.

This constraint refers to a level defined by a certain rule or set of rules and to the semantic representation (for indication of coreference).

Assuming that something like (18) were correct, a supporter of the revised standard theory would be forced to modify his theory in the direction of allowing a class of interpretive rules that could take levels of structure defined by transformation names as input and give aspects of semantic representation as output. This would be a significant change, though not an overwhelming one, so the demonstration of the existence of a constraint with the properties of (18) would not in itself provide that sharp a differentiation between the general principles which underly generative semantics and those which underly the kind of linguistic thinking which has led to the revised standard theory.

B. The Wh *Constraint Cannot Be Defined on the Input to the* Wh *Movement Rules Alone*

However, I will show now that the *Wh* Constraint cannot in fact be given in the form (18), where there is reference to the possible input structures of a class of rules and to the semantic representation. Such a constraint perhaps offers the possibility of re-

[16] For discussion of such derivational constraints, cf. Geis (to appear) and Postal (to appear). Jackendoff's (1969) theory of cyclical interpretive rules allows restrictions to be stated on particular single levels of structure which are not any of the distinguished levels, but only if these restrictions apply successively to each level of embedding. However, the cyclical idea has no application to the phenomena covered by the *Wh* Constraint.

formulation in terms of Interpretive Rules; however, the proper formulation of the *Wh* Constraint requires reference to two levels of structure besides the semantic representation.

That reference to two levels of structure is necessary follows from the fact that in the dialect group being described, (18), while true for a certain class of cases, is nonetheless false. Notice that (18) refers to pronouns to the left of *wh* forms at the stage of input to the *Wh* Movement rules. However, all the actual sentences considered so far which illustrate (18), and seem to support it, are such that the relevant pronoun ends up to the right of the *wh* form, because of actual application of one or another *Wh* Movement rule. One could, therefore, provide a crucial test of (18) by finding a class of sentences which contain *wh* forms, but in which these forms do not undergo any of the *Wh* Movement rules, and remain in their original positions. Fortunately, the dialects in question offer several classes of such sentences, all interrogative in nature. First of all, there is the well-known group of *Incredulity Question Clauses*, characterized by a special sharply rising intonation:

(19) Nixon appointed who to the Supreme Court?

Second, there is a little-known class of what I will call *Legalistic Question Clauses*, which seem natural only in the mouths of courtroom attorneys, police investigators, and quiz program announcers. These have falling intonation.

(20) The person you saw was walking in what direction?

Finally, there is the subclass of ordinary *wh* interrogatives in which more than one nominal is questioned and *wh*-marked:

(21) Which president sent an ostrich to which tapdancer?

The relevance of these three classes is that in each case backwards pronominalization between a *wh* form and a pronoun on its left is possible. First, with Incredulity Question Clauses:

(22) a. The newsman who criticized him$_i$ later belted $\left\{\begin{matrix}\text{what}\\\text{which}\end{matrix}\right\}$ official$_i$?

b. Finding out he$_i$ won surprised $\left\{\begin{matrix}\text{what}\\\text{which}\end{matrix}\right\}$ candidate$_i$?

This possibility also exists with Legalistic Question Clauses:

(23) a. Mr. Jones, for $100,000, the man who appointed him$_i$ later said $\left\{\begin{matrix}\text{what}\\\text{which}\end{matrix}\right\}$ secretary of state$_i$ was an imbecile?

b. Remembering you are under oath, the witness who claimed he had never seen it$_i$ was walking toward $\left\{\begin{matrix}\text{what}\\\text{which}\end{matrix}\right\}$ building$_i$?

Finally, the linkages are equally possible in question clauses with multiple *wh*-marked nominals:

(24) a. Which columnist reported her$_i$ victory to which actress's$_i$ mother?
b. What company had his$_i$ wife spy on what well-known industrialist$_i$?

Examples like (22)–(24) show then that constraint (18) is incorrect, since (18) claims that the coreference linkages should be impossible in all such cases.[17]

One must therefore ask what the difference is between the class of cases for which (18) appears to have no exceptions, and the class represented by (22)–(24) for which it fails. The answer is obvious: (18) works just in case the *wh* forms are moved by the *Wh* Movement rules, that is, just in case the pronoun which was originally to the left ends up after these movements on the right. It is this movement which is absent from (22)–(24) and all similar sentences. Hence compare the well-formed linkages in (24) with the ill-formed ones in:

(25) a. *To which actress's$_i$ mother did the columnist report her$_i$ victory?
b. *What well-known industrialist$_i$ did the company have his$_i$ wife spy on?

C. A Multilevel Formulation

I propose therefore that it is exactly the property of altered relative position that must be referred to in the correct reformulation of (18). The *Wh* Constraint appears to block coreference between a *wh* form[18] and a pronoun which is to the left of it in the input to the *Wh* Movement rules just in case this constituent order is reversed in the output of these rules. Hence one can provide a tentative statement of the *Wh* Constraint along the lines of (26):

(26) Mark as ill-formed any derivation in which:

a. there are two nominal constituents, *A* and *B*, in the input structure of a *Wh* Movement rule, where:
(i) *A* is a pronoun
(ii) *B* is a *wh* form
(iii) *A* is to the left of *B*; and
b. the corresponding constituents of *A* and *B* in the output structure of the *Wh* Movement rule, call them *A′* and *B′* respectively, are aligned such that *B′* is to the left of *A′*; and

[17] For some reason which I do not understand at all, backwards pronominalization in sentences like (22)–(24) does not seem well-formed with single word *wh* nominals, i.e. *who* and *what*. Thus:

(i) ?The witness who claimed he had never see it$_i$ was walking toward what$_i$?
(ii) *Which columnist reported her$_i$ victory to who$_i$?

This has no bearing on the argument of the text, however, since nothing precludes the existence of separate constraints on backwards pronominalization for certain *wh* forms in certain contests.

[18] As pointed out in footnote 10, this statement is too narrow.

c. in the semantic representation, *A* and *B* (or, more precisely, their corresponding elements) are marked as stipulated coreferents.[19]

The crucial properties of this filter rule are, then:

(27) a. that it refers to two levels of structure in addition to semantic representation;
b. that the two levels required are not the distinguished levels of any theory, in particular not surface structure, shallow structure, or deep structure;
c. that the two levels can be taken to be the input and output structures of certain rules, namely the *Wh* Movement rules; and
d. that the constraint makes crucial use of the notion *corresponding constituent.*

These properties show that the constraint is of the type posited by Lakoff under the term *Global Derivational Constraint.* One minor matter, however, is that in the case of (26) the two levels of structure referred to can be contiguous, whereas in general Lakoff allows Global Derivational Constraints defined on noncontiguous trees. And indeed (26) also manifests the noncontiguous property, if one takes into account the semantic representation.

I conclude then that the *Wh* Constraint is a Global Derivational Constraint of the type suggested by Lakoff, the existence of which is claimed in generative semantics and denied by Chomsky. If, as seems clear, (26) is a realistic approximation to a formulation of this constraint, there is no possibility of reformulating it as an Interpretive Rule in the sense of the recent work of Chomsky and his students.

However, it is clear that any work done by Interpretive Rules can be accomplished by Global Derivational Constraints. Thus the situation is as follows:

(28) a. Some facts are describable by Interpretive Rules.
b. These facts are also describable by Global Derivational Constraints.
c. There are many theoretically possible types of linguistic situation which would be describable by Global Derivational Constraints but not by Interpretive Rules. That is, Global Derivational Constraints are a "stronger" notion in one sense.
d. But, as the earlier discussion shows, this theoretical possibility is realized

[19] That the difference between stipulated coreference and asserted coreference is actually relevant for the proper statement of the *Wh* Constraint is shown by such examples as:

(i) *Who* did Mary say the one who visited *his* mother was?

In (i) the italicized NPs can be understood as coreferents in spite of the *Wh* Constraint because they are asserted coreferents. That is, (i) can correspond to a more remote structure:

(ii) Mary say the one$_i$ who$_i$ visited his$_i$ mother was wh + some one$_j$

In this structure, *one*$_i$, *who*$_i$, and *his*$_i$ are stipulated coreferents, but *one*$_i$ and *one*$_j$ are asserted coreferents, a relation represented in surface structure by the verb *was*. Hence the repositioning of *wh* + *some one*$_j$ (*who*) to the left of the pronoun *his*$_i$ is unaffected by constraint (26), which refers to stipulated coreference.

in English. There exist bodies of linguistic data, in particular those defining the basis for the *Wh* Constraint, which are describable in terms of Global Derivational Constraints but not by Interpretive Rules. It follows that linguistic theory needs Global Derivational Constraints, but does not need Interpretive Rules. On the basis of present evidence, one must conclude that the proper way to overcome the limitations of transformational rules per se is by the application of Global Derivational Constraints.

V. Appendix A: A Single Level Statement of the Constraint in Terms of "Features" or Other Ad Hoc Markings

I have concluded in the main body of this paper, specifically in Part IVB, that the *Wh* Constraint cannot be stated on any single level of structure (ignoring in the count the semantic representation). Rather, I claimed, reference to both the input and output structures of the *Wh* Movement rules is required. In this appendix I would like to consider one inadequate attempt to avoid this conclusion which might appeal to some who are familiar with the notion of "syntactic features" in the sense of Chomsky (1965) or Chomsky (to appear a), to name only two works utilizing this concept.

Clearly, if (counter to my earlier claim) it is possible to state the constraint on any single level, the only possible candidate is the one defined by the input structure to *Wh* Movement rules. I argued in Part IVB that no such account was possible because the property of having the pronoun on the left or right of the *wh* form at this stage fails to differentiate fully between the cases. As shown by sentences like (22)–(24), it turned out that whether or not a pronoun present at this stage to the left of a *wh* form could be coreferential with it depended on whether or not it subsequently ended up on the right in the output of the *Wh* Movement rules. Hence the need to appeal to both the input and output structures in (26).

However, a proponent of deep structures in part incorporating syntactic features might well try to avoid this (for him unhappy) conclusion in some way like the following. Recall that what is needed is some way to distinguish between cases such as (29a, b):

(29) a. The newsman who criticized him$_i$ later belted which official$_i$? (Legalistic Question)
b. *Which official$_i$ did the newsman who criticized him$_i$ later belt?

In the formulation which I proposed, it is the position of the pronoun in (29b) to the right of the *wh* form in the output of the *Wh* Movement rules which is crucial. The advocate of feature marking might then claim that the structures of (29a) and (29b), and all similar pairs, are in fact distinguished at the level of input to the *Wh* Movement

rules. In particular, he might claim that there is some feature, say *G*, whose positive and negative values distinguish between such pairs. He might, for instance, claim that *G* is assigned to *wh* forms. Thus he might say that, at the stage of input to the *Wh* Movement rules, *wh* forms like those in sentences of the type (29a) are marked $[-G]$, while those in sentences of the type (29b) are marked $[+G]$. In support of this hitherto unmotivated feature, he might say something like the following: clearly, at the stage where the *Wh* Movement rules apply it is necessary to distinguish (29a) from (29b). Otherwise, there would be nothing to indicate that the rule applies in one case and not in the other. The feature *G* fulfills just this function.

But, given the availability of a device like *G*, the feature-marking theorist can quickly show that reference to the output structures of *Wh* Movement rules for the purpose of stating the *Wh* Constraint can be avoided. For the constraint can now be reformulated to say (in one way or another) that backwards pronominalization between a pronoun and a *wh* form on its right at the stage of input to the *Wh* Movement rules is all right just in case the *wh* form is marked $[-G]$. Hence it might be claimed that the *Wh* Constraint can be formulated in terms of one level of syntactic structure proper together with the semantic representation, and hence that it is reformulable as an Interpretive Rule in Chomsky's sense, with the minimum modification of the revised standard theory mentioned in Section IVA; that is, a modification allowing Interpretive Rules to apply to nondistinguished levels of structure picked out by the names of transformations. Global Derivational Constraints, in the sense of rules which refer to two or more levels of structure neither of which is the semantic representation, would have been shown to be (in this case at least) unnecessary.

By discussing such an approach only in an appendix, I hope to have suggested already that it is not a serious alternative to the description given earlier involving explicit Global Derivational Constraints. For the point is not that the feature marking approach fails to provide a description in accord with the facts in this case, but rather that if arbitrary features are available, any possible Global Derivational Constraint[20] can be reformulated in terms of a single level of structure. Thus, suppose we consider

[20] Actually, as Ross points out, it is not literally true that *any* Global Derivational Constraint can be recoded using ordinary syntactic features. In particular, the example of Ancient Greek case agreement described in Lakoff (1970) and Casagrande (1970) is indescribable in terms of such features. The reason is that the relevant constraint in Greek keeps track of the earlier stage properties of an unbounded number of constituents. The rule is, roughly, that a participle agrees with its subject in case, where the relevant subject is the one at the end of the cycle defined by the sentence node immediately above the participle. The crucial point is, however, that the case with which the participle agrees may be determined by very late operations defined after the "subject" in the relevant sense has ceased to be such. Since there is no bound on the number of relevant subject-participle pairs which have to agree in a Greek sentence, it follows that recoding this constraint in terms of features would require linguistic theory to provide an unbounded set of integers as features such that relevant nominal-participle pairs could be given as: nominal_{13}-participle_{13}, nominal_{5678}-participle_{5678}, and so forth. Hence the case rule would then be statable as: participle_n agrees in case with nominal_n, assuming that linguistic theory is also expanded to allow variables (like *n*) over integer features, and to allow agreement rules to be stated in terms of such variables. The degree to which an integer-feature approach to Greek case agreement is mistaken as an alternative to the Global approach thus hardly needs stress. One must conclude then that Global Constraints are recodable in terms of ordinary features only when they keep track of the multitree properties of a limited number of constituents.

some arbitrary Global Derivational Constraint which refers to two constituents, A and B, in tree (T_i), and their correspondents, A' and B', in tree (T_m). Suppose further that the relevant properties of A and B in (T_i) are that they are terms of some relation R, while in tree (T_m) the relevant relation between A' and B' is R'. One can quickly eliminate reference to one or the other tree by use of features. Thus, assign to all constituents like B that bear the relation R to some A the feature $[+M]$. One can then reformulate the constraint with only reference to a single tree by saying that the constraint marks as ill-formed all derivations in which there is a tree containing two constituents, A' and B', related by R', such that B' manifests the feature $[+M]$.

The point should be clear. Marking structures with arbitrary features is not a way of avoiding Global Derivational Constraints. It is simply a way of coding these constraints in a pointless and obscuring notational framework, that of "features". As far as I can see, there can be no possible empirical properties representable by Global Derivational Constraints which can not be recoded in this way. Thus, the reformulation makes no serious empirical claims and is not an alternative description. I should note that this conclusion is unaffected by whether or not the feature solution is claimed to be "universal", since the availability of arbitrary features in universal grammar provides the same lack of content as their availability in particular grammars.

In short, a theory which makes available arbitrary syntactic features is not weaker than one with Global Derivational Constraints; it is at least as powerful, and no doubt more so if we consider reasonable constraints on Global Derivational Constraints as mentioned in Section I. Thus, insofar as appeal to arbitrary features has been possible in syntactic theory over the last five years or so, at least the power of Global Derivational Constraints has been implicitly available. It would be reassuring to believe that no use has been made of this deceptive approach; but in fact this is not the case. The present writer himself is guilty of this very tactic. Thus I proposed (Postal 1970) an arbitrary feature, [*Doom*], whose sole function was to code properties of derivations at the stage of the transformational cycle to keep them available at postcyclical stages, since I argued that certain deletion rules which must apply at the latter level are restricted by generalizations statable only at the former.[21] In defense, it can be said that an explicit theory of Global Derivational Constraints had not been proposed at the time [*Doom*] was postulated.

A further point. The argument that the feature coding mechanism is empirically empty as a way of avoiding Global Derivational Constraints is obviously not affected if (instead of arbitrary features) one appeals to arbitrary dummy elements, that is, to segmental markers which are later deleted.[22] These will provide the same power, and they burden the user only with the need to delete them at some point after statement of the constraints they have coded.

[21] For discussion of this case, cf. Lakoff (to appear), Postal (to appear).

[22] I made use of such arbitrary elements in my doctoral thesis (Postal 1962). Chomsky (to appear a) apparently continues to permit their use.

Returning now to the *Wh* Constraint, it is important to note not only that the feature coding mechanism fails to avoid Global Derivational Constraints, but also that it encounters adverse arguments on factual grounds. Thus I considered the possibility of distinguishing between cases like (29a, b) at the level of input to the *Wh* Movement rules in terms of the markings $[-G]$ and $[+G]$ respectively. This solution has a grain of plausibility in this case because there is no other formal difference apparent in the structures of these sentences at that stage. But now consider contrasts such as those of (30a, b):

(30) a. What press agent reported the fact that she$_i$ had won an Oscar to what actress$_i$?
b. *To what actress$_i$ did Max report the fact that she$_i$ had won an Oscar?

This distribution of acceptable and nonacceptable coreference linkages follows from (26). But notice that no feature $[-G]$ is necessary to mark, in the input to the *Wh* Movement rules, the fact that these rules are inapplicable to (30a); this is a general fact about clauses containing two *wh*-marked NPs, at least when one of them is the subject nominal. Thus the rule which assigned $[-G]$ to structures like (30a) would not only be arbitrary in several respects (where should $[G]$ be marked in these cases, for example?) but would also be redundant since the natural structural representations provide the information that *Wh* Movement rules do not apply. But if $[-G]$ is not marked in these cases, the reformulation of the *Wh* Constraint must be replaced by a less general statement in which there is reference not only to $[G]$ but also to multiple *wh*-marked NPs within a single clause, where one of them is a subject nominal. Hence the feature marking mechanism cannot provide a truly general statement of the *Wh* Constraint, no matter which choice is made.

A much deeper point emerges from this discussion, but I do not have space here to do it justice. A wide variety of proposals within transformational grammar has, in one way or another, involved the postulation of arbitrary, unnatural linguistic structures, in order to permit the statement of various linguistic restrictions within an overly narrow, nonglobal rule framework. This is, I believe, a basic mistake. The most fundamental virtue of Global Derivational Constraints is that they can eliminate at least a great variety of such devices, permitting statement of restrictions in terms of a sequence of natural structures,[23] either those for which there is direct empirical evidence, like semantic representation and surface structure, or those automatically derived by natural rules connecting these. I claim, then, that too weak a theory of rules has led to much invention of artificial structures. These include arbitrary features

[23] It can of course be claimed that the notion of "natural structure" is vague and unclear, that no general criterion for assigning structures to one or the other category exists, and so forth. While true, this would surely miss the point, namely that there is a notion to be characterized here and that there are many clear cases on both sides. In particular, a theory which allows arbitrary syntactic features and/or arbitrary segmental markers is incompatible with a constraint to natural structures. The problem is analogous to that in phonology, where a serious theory of phonological representation cannot be constructed unless there are very narrow constraints on the class of phonological features available.

such as those discussed above; arbitrary segmental markers similar to the one in the present writer's thesis or various recent proposals by Chomsky; but, more seriously, the whole apparatus of deep structure as distinct from the natural, empirically confirmable level of semantic representation. One of the fundamental contrasts between generative semantics and the recent work of Chomsky and his students, a contrast which has received little or no attention, is that the former insists on a restriction to natural structures, while the latter makes wide use of arbitrary markers of various sorts.

VI. Appendix B: Comparison with the Cross-Over Formulation

In previous work on pronominalization restrictions such as those discussed in this paper (Postal 1968), I attempted to describe the facts in terms of a principle blocking the application of movement transformations in cases where such movement would "cross" one nominal over another with which it was a coreferent under various quite complicated conditions. It should be obvious that this is in some respects quite close to the formulation given in (26) above. Because the *Wh* Movement rules transport constituents to the left, a formulation that says a *wh* form having a (coreferential) pronoun on its left in the input to the *Wh* Movement rules cannot have that pronoun on its right in the output is very close to one which prevents a *wh* form from being crossed over a coreferential nominal.

It might be asked what the basis is for choosing between these two descriptions, and whether the earlier formulation would not in such cases open the way for avoidance of the need for Global Derivational Constraints. I shall answer the second question first. Treating the *Wh* Constraint and similar restrictions in terms of a principle that blocks the application of transformational rules does seem to permit the restriction to be stated on a single tree, namely the input to the *Wh* Movement rules. However, this is only possible if two assumptions are made. First, the relation of coreference between nominals must be marked not only in the semantic representation but also in derived structures at least up to the point of input to *Wh* Movement rules. Second, transformations must be permitted to be defined at least partly in terms of such information. However, since the coreference is marked in semantic representations, there is no independent reason for assuming it is marked in later stages. In a way, the claim that it is marked in the intermediate levels would represent a kind of coding of information from more abstract structures into less abstract ones of the type discussed in the previous appendix. It is thus not even clear that such a marking would present a true alternative to Global Derivational Constraints. Further, even if such markings are permitted in syntactic structures like the input to the *Wh* Movement rules, a complication of the structural application statements for transformational rules is required to permit them to take account of such markings.

The latter problem might be avoided if it could be shown, as I attempted in

earlier work, that the "crossing" restrictions were universal. Hence the constraints in question would not be part of individual transformations but metaconditions on applicability of nominal movement rules stated in linguistic theory. But it is now clear that in general this is wrong, since the *Wh* Constraint is not even operative in the dialects of all English speakers. Hence this constraint must be present in the grammars of some speakers and not in those of others. Therefore, the crossing account commits us to language- and dialect-particular conditions on particular transformations. But this is an unhappy result in the extreme, since it would claim that the *Wh* Movement rules in one dialect are not the same as those in another, just because of "crossing" constraints. Clearly, the right solution is to leave these rules in identical form for all dialects and to state the constraints as separate filter conditions. This is just what the Global Derivational Constraint formulation permits, since one can simply say that some dialects contain (26) and others do not. I thus conclude that the "crossing" formulation was mistaken, and that it (i) cannot be taken to avoid Global Derivational Constraints, and (ii) can be seen to be definitely inferior to a description in terms of global conditions.

I might note that the "crossing" formulation is unacceptable to interpretivists in any event because it requires reference in the syntax to what they claim is a purely "semantic" relation. Hence appeal to the "crossing" formulation would not be a possible way around the arguments given earlier for statement of the *Wh* Constraint as a Global Derivational Constraint. Thus Jackendoff (1969, 73–74) states:

> In an interpretive theory of reference, the Crossover Principle cannot be stated, since coreferentiality, a purely semantic concept, cannot be referred to in transformations, and conversely, semantic rules cannot depend on what transformations have taken place, but only on the resulting structural configurations.

But we showed earlier that the *Wh* Constraint cannot be stated on the "resulting configurations" of any rule, or on the configurations prior to any rule, although it can be stated on the pair of configurations given in the input and output of the *Wh* Movement rules, that is, it can be stated in the form of a Global Derivational Constraint.

VII. Appendix C: A Proposal to Reduce the *Wh* Constraint to an Independent Restriction

I have claimed that the *Wh* Constraint is not reducible to any more general blockage on pronominalization. In footnote 9, however, I referred to an attempt to eliminate this constraint by taking it as a subcase of an independent restriction.

This account would seek to reduce the constraint to a more general limitation banning backwards pronominalization between indefinite nominals and pronouns. I rejected this proposal in Postal (1968, Chapter 21) for what I now think are largely

wrong reasons having mostly to do with dubious assumptions about rule ordering. This proposal would, for instance, try to reduce the constraint in (31a) to that in (31b):

(31) a. *Who$_i$ did his$_i$ mother yell at?
b. *His$_i$ mother yelled at someone$_i$.

on the basis of at least two assumptions:

(32) a. *Wh* forms are *wh* + *some* forms, hence indefinites; and
b. The constraint illustrated by (31b) can be stated at a level of structure prior to the application of the *Wh* Movement rules.

(32b) is required because of the well-formedness of examples such as:

(33) Someone$_i$ yelled at his$_i$ mother.

There are, however, fundamental objections to any attempt to reduce facts like those in (31a) to the regularity underlying those like (31b). First, to make the argument go through, it is necessary to derive not only question clause *wh* words but also those in both restrictive and appositive relatives from indefinite forms, since, as noted earlier, the coreference constraints analogous to (31a) are identical for the three clause types. But, regardless of one's opinions about restrictive clauses, it is surely out of the question to derive the *wh* words of appositive relative clauses from indefinites. That is, a structure like (34b) for sequences like (34a) is impossible:

(34) a. *Johnson$_i$, who$_i$ he$_i$ believes I attacked
b. [Johnson$_i$, he$_i$ believes I attacked wh + some X]

Rather, the element underlying *who* in appositives like (34a) must be a definite pronoun like the object of *attack* in:

(35) Johnson$_i$, and he$_i$ believes I attacked him$_i$

Hence on this ground, the attempted reduction fails since it would predict the absence of the constraints in appositive relative clauses when in fact the constraints in such clauses are identical to those in question clauses.

Second, on grounds independent of the above argument, the proposal fails in a much more radical way. That is, it does not even work for question clause *wh* words. For as we have seen there are a wide variety of contexts in which backwards pronominalization with a *wh* form antecedent is well-formed provided the *wh* form is unmov ed by the *Wh* Movement rules. Thus contrast:

(36) a. *The tyrant tortured her$_i$ mother in front of some helpless princess$_i$.
b. What tyrant tortured her$_i$ mother in front of what helpless princess$_i$?

If it were truly the case that the *Wh* Constraint was a consequence of a general condition banning backwards pronominalization involving indefinites, how could one

explain the well-formed linkages in examples like (36b), or those mentioned earlier of the Incredulity Question Clause or Legalistic Question Clause type?

Third, and finally, the proposal is incompatible with the fact, noted in footnote 10, that the *Wh* Constraint holds not only for *wh* words per se, but also for a certain class of nominals in which these are embedded. Thus the constraint holds not only for the italicized nominal in (37) but also for the parenthesized one and the square bracketed one:

(37) *[(*whose*$_i$ friend's$_j$) father$_k$] did $\left\{\begin{matrix}\text{his}_i\\ \text{his}_j\\ \text{his}_k\end{matrix}\right\}$ sister criticize

But it is generally understood that, while the italicized nominal in such cases is indefinite, the parenthesized and bracketed nominals are in fact definite. That is, initial "maximal" nominals like those in (37) would have a more remote structure like:

(38) the father of the friend of whose

And the nonquestion analogues of these, as expected, do not involve any constraints on backwards pronominalization of the type that could explain the restrictions on pronominalization for the *j*- and *k*-subscripted nominals:

(39) a. His$_j$ sister criticized the father of the friend$_j$ of someone (I know).
b. His$_k$ sister criticized the father$_k$ of the friend of someone (I know).

Thus, while the suggestion under discussion here is perfectly reasonable and appropriate and deserves consideration, there seem to me to be three basic reasons why it fails. Hence I see no alternative to the claim that the *Wh* Constraint is an independent restriction in English grammar, even considering only the dialect group that manifests the restrictions I have designated as the *Wh* Constraint. But, equally seriously, I have already noted that there are speakers who do not manifest these restrictions, that is, speakers who do not have the *Wh* Constraint. These speakers, however, do have the constraint illustrated in examples like (31b). Thus there are even clearer grounds to conclude that that constraint, whatever it is, and the *Wh* Constraint are independent.[24]

[24] It should be stressed, however, that some of the cases of banned coreference involving *wh* forms originally discussed in my work on "crossing" constraints appear to be reducible to independent pronominalization constraints. For instance, in the present paper I have dealt with no cases like:

(i) *Who$_i$ did he$_i$ claim Mary wanted to kiss?

cases in which, in the input to the *Wh* Movement rules, the pronoun would both command and precede the *wh* form. In my original work I claimed that the ill-formedness of examples like (i) was due to the same principle as the ill-formedness of examples like:

(ii) a. *Who$_i$ did his$_i$ father claim Mary wanted to kiss?
b. *Who$_i$ did the claim that he$_i$ was a drunk fail to bother very much?
c. *Who$_i$ did the nurse that he$_i$ kissed report to the FBI?

which have been discussed in this paper. But I now deny this. The ill-formedness of examples like (ii) is, I claim, due to principle (26) above, the *Wh* Constraint. But I think the ill-formedness of examples like (i) can be reduced to the same principle which blocks examples like:

(iii) *He_i claimed Mary wanted to kiss $\left\{\begin{matrix}\text{Max}_i \\ \text{the captain}_i\end{matrix}\right\}$.

That is, I suggest that what is wrong with (i) is simply a violation of the principle that a pronoun cannot both command and precede its antecedent. To make this suggestion obtain, however, it is necessary for this latter principle to be defined on pre-*Wh* Movement Rule structures so that the violation in (i) would be defined on structure:

(iv) [he_i claimed Mary wanted to kiss who_i]

This causes certain problems, but I believe they can be surmounted.

One very strong reason for distinguishing cases like (i) from those like (ii) is that the former seem to be truly dialect universal. I have found no speaker who accepts the coreference linkages in examples like (i). The linkages like (ii) are, however, dialect particular. But it is, of course, also the case that no speakers accept the linkages in examples like (iii). Thus (i) and (iii) correlate with respect to their universality, while (i) and (ii) do not. This is accounted for by a universal constraint prohibiting pronouns which both command and precede their antecedents (at some appropriate stage) and by a nonuniversal constraint, (26), blocking coreference in cases like (ii) for *wh* forms only (ignoring the extensions mentioned in footnote 10).

References

Binnick, R. I. et al. (1969) *Papers from the Fifth Regional Meeting*, Chicago Linguistic Society, Department of Linguistics, Univ. of Chicago.

Casagrande, J. (1970) "A Case for Global Derivational Constraints," *Papers in Linguistics* 2.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*, MIT Press, Cambridge, Mass.

Chomsky, N. (to appear a) "Remarks on Nominalization," in Jacobs and Rosenbaum (1970).

Chomsky, N. (to appear b) "Deep Structure, Surface Structure, and Semantic Interpretation," in Jakobovits and Steinberg (to appear).

Chomsky, N. (to appear c) "Some Empirical Issues in the Theory of Transformational Grammar," in Peters (to appear).

Emonds, J. E. (1969) *Root and Structure-Preserving Transformations*, unpublished Doctoral dissertation, MIT.

Geis, M. (to appear) "A Derivational Constraint on the Simplification of Certain Comparative Constructions," to appear in *Linguistic Inquiry*.

Jackendoff, R. S. (1969) *Some Rules for English Semantic Interpretation*, unpublished Doctoral dissertation, MIT.

Jacobs, R. and P. S. Rosenbaum, eds. (1970) *Readings in English Transformational Grammar*, Ginn-Blaisdell, Waltham, Mass.

Jakobovits, L. A. and P. Steinberg (to appear) *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology and Psychology*, Cambridge University Press, Cambridge.

Lakoff, G. (1969) "On Derivational Constraints," in Binnick et al. (1969).

Lakoff, G. (1970) "Global Grammar," *Language* 46, No. 3, 627–39.

Lakoff, G. (to appear) "On Generative Semantics," in Jakobovits and Steinberg (to appear).

Peters, P. S. (to appear) *Goals of Linguistic Theory*, Prentice-Hall, Englewood Cliffs, New Jersey.

Postal, P. M. (1962) *Some Syntactic Rules in Mohawk*, unpublished Doctoral dissertation, Yale University.

Postal, P. M. (1968) *Cross-Over Phenomena*, in *Specification and Utilization of a Transformational Grammar* (Scientific Report No. 3), IBM Thomas J. Watson Research Center (published by Holt, Rinehart and Winston, New York, 1971).
Postal, P. M. (1970) "On Coreferential Complement Subject Deletion," *Linguistic Inquiry* 1, 439–500.
Postal, P. M. (to appear) "The Best Theory," in Peters (to appear).
Ross, J. R. (1967) *Constraints on Variables in Syntax*, unpublished Doctoral dissertation, MIT.

Thomas J. Watson Research Center
IBM
Box 218
Yorktown Heights, New York 10598