

ANALOGICAL GENERALIZATION IN
NATURAL LANGUAGE SYNTAX

by

Leland Maurice George

S.B., Massachusetts Institute of Technology
(1974)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 1980

Copyright (C) 1980 Leland Maurice George

The author hereby grants to M.I.T. permission to reproduce
and to distribute copies of this thesis document in whole
or in part

Signature of Author
Department of Linguistics and Philosophy
February 12, 1980

Certified by
Noam Chomsky
Thesis Supervisor

Accepted by ..
Samuel Jay Keyser
Chairman, Departmental Committee
on Graduate Students

ARCHIVES
MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

MAR 3 1980

LIBRARIES

To Thomas A. George
and
Evelyn E. Wegman-George
and their younger children

To Noam Chomsky

To James Ronald Havern

ANALOGICAL GENERALIZATION IN
NATURAL LANGUAGE SYNTAX

by

Leland Maurice George

Submitted to the Department of Linguistics and Philosophy
on February 12, 1980 in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Linguistics

ABSTRACT

It is observed that "analogy", understood as bearing on the projection problem for natural languages, has no sense outside of the context of the theory of universal grammar, whereas within this context it takes on several distinct meanings. The first of these "generative analogies" is the phrase structure symmetry expressed by the bar convention for the base, which we propose to limit to the relational schema "Specifier--Head--Adjunct", questioning the Uniform Projection Hypothesis. The second generative analogy is defined by the theory of derived constituent structure, which we do not further in this work. The third is the notion of parallelism controlling ellipsis rules such as those characteristic of coordinate conjunction; we attempt to establish a fairly broad scope for these rules by restricting base-generated coordination, and offer the first precise analysis of the relevant notion of parallelism in our Law of Congruity. The final generative analogy concerns the distinction between grammaticality and acceptability, as developed in the theory of derivative generation and filtering by analogy.

Thesis Supervisor: Noam Chomsky

Title: Institute Professor

Acknowledgements

Many thanks

to Noam Chomsky, Ken Hale, and Paul Kiparsky
for being my thesis committee

to Luigi Rizzi, Jean-Roger Vergnaud, and Edwin Williams
for being a shadow committee

to William J. Hawkins, Craig Hoffman, Robert Ingria,
Jaklin Kornfilt, Philip Le Sourd, Jindrich Toman,
and Amy Weinberg
for deepest friendship and philosophical camaraderie
(especially to Hawkins, for getting me into LSLT;
to Ingria, for his gargantuan labors in preparing
the final draft, and also to Hawkins and Kornfilt,
for considerable help on this score)

to the Linguistics Department of the University of
Connecticut (especially Howard Lasnik, Craig Hoffman,
David Michaels, and Arthur Abramson)
for a year's teaching experience that helped me write a
substantially better thesis than I would have
otherwise

to Sister Maria Joseph Corbett, Philip Le Sourd,
James Harris, and David Perlmutter
for coaxing me (in that order) in Chomsky's direction

to all my friends, within and outside of the West Bank
kibutz, too many to name here
for supporting me in this work

and

to Haj

Table of Contents

Acknowledgements	4
1. Taxonomic and generative analogies	7
2. How Phrase Structure rules are written	13
2.1 On the corner	20
2.2 Fragment of the English base	32
3. How Transformations are Written	37
4. How the rules work	46
4.1 Phrase structure	46
4.2 Transformations	57
4.3 Appendix: Filters and Mistakes	80
5. Coordination and Ellipsis	88
5.1 Restricting Phrase Structure Conjunction (Back to the Corner	89
5.2 Grammatical Ellipsis	126
5.3 Wh-Movement "Across the Board"	153
5.4 Comparative Subdeletion	160
6. Derivative Generation and Filtering by Analogy	164
Bibliography	173

1. Taxonomic and generative analogies

American "Structural Linguistics", as is well known, put forth the ideal of strictly ordered procedures of segmentation and classification, and admitted as the sole legitimate activity of our science the collecting of corpora to which to apply these procedures. For all their supposed rigor, however, the choice of procedures was arbitrary, and the temptation to understand the resulting analyses as showing anything about languages or their speakers was heroically resisted. Now obviously The American Structuralists didn't always live up to their taxonomic tenets, as is obvious from the generative value of many of their analyses. Nevertheless, their bare allegiance to these ideals plainly impeded their investigations. This situation was due to the prevailing positivism of the day, and, ironically, prevented the procedures from being put to the use to which they were best suited, namely, in detailing an empiricist-behaviorist model of language acquisition. Commentary on the projection problem was therefore confined to vague digressions on analogy and behavioral patterns.

Of course analogy must enter into the solution in the

trivial sense that any rational inquiry whatsoever must face questions of identity and difference. But what is a human language, and what is "the same" and what is "different" in this realm?

The Muse gives each poet to answer this question in his own way, but to the natural philosopher she sends only the intuitive categories of Grammar and Rhetoric. Thus the appeal to analogy in solving the projection problem takes on a determinate sense only in the context of a theory of Universal Grammar. In fact it takes on several different senses in this context, which it is my burden in this work to distinguish.

From a philosophical point of view--that of justifying grammars and delimiting the valid use of introspection--the most vital sense of "analogy" comes from distinguishing, among utterances, what is, in the strictest sense grammatical from what is intelligible or useful. This refines the competence-performance distinction in light of acceptable but ungrammatical sentences, and, possibly, grammatical but unacceptable ones.

Our view of these problems is quite speculative, and

we justify it mainly on the grounds of its technical value for linguists seeking structure in the phenomenal blur of native speakers' intuitions. We therefore reserve this matter for §5, where we will rely on its formal utility having been displayed.

This expository policy reverses the historical as well as the logical priorities, since the idea of "degrees of grammaticalness" and "semi-grammatical utterances" pervaded the earliest work in generative syntax. In fact, they developed originally in the theory of lexical selection (see *LSLT* *passim*, especially Chapter V). Only recently, beginning with "Remarks", has anyone seriously tried to generalize the idea of ungrammatical acceptability to non-lexical rules, and many generative grammarians still ignore the suggestion, or positively deplore it. It turns out, however, that the use of analogy, in this sense, overcomes the widest variety of obstacles to the construction of revealing theories of grammar, and we rely on it heavily as we proceed.

The remaining senses of analogy with which we will be concerned belong more properly to the technology of the theory as such. One shows the piecewise parallelism among

the major categories provided by the X-bar theory. Another concerns the concept of "corresponding constituent" or "root" (in the sense of LSLT, not of Emonds), the core notion of derived constituent structure. The last is the kind of internal parallelism of constituents required in the theory of grammatical ellipses; an explicit definition of this will be the main technical contribution of this thesis.

With this, we are almost ready to take up questions of grammatical analysis. Before we leap into the *mêlée*, however, we pause for a brief meta-theoretical discussion. This concerns the

Format{ Best Theory Rule}

If two theories of grammar provide comparably accurate accounts for the same range of linguistic structures, and if one theory provides a narrower range of accessible grammars than the other, then the burden of proof is on the proponent of the latter theory.

Here "accessible" means "below a certain threshold of complexity, relative to a given evaluation metric".

In spite of widespread skepticism and misunderstanding, this maxim has been presupposed in all serious discussion of natural language grammars since LSLT. True, one could argue that it was only implicit in

the classical idea of simplicity, that grammarians began applying it self-consciously only in the *Aspects* era. Nevertheless, LSLT surely followed it in effect, inasmuch as it treated the issue of the evaluation metric seriously.

We, at any rate, will adhere quite explicitly to the Best Theory Rule, with one qualification, hinted at in "Conditions on Rules":

We want UG to make as strong a statement as possible about the nature of language and thus to be subject to critical tests and to provide explanations for the phenomena attested in descriptive study. We can explain the fact that linguistic competence has the property *P* insofar as we can show that property *P* conforms to UG and is, furthermore, the special case of UG determined by experience. In the most interesting cases, the role of experience is limited or even nonexistent so that the property *P* simply reflects some property of UG and thus gives us direct insight into the nature of UG. We argue that a given language has the property *P* because UG requires that this be the case. Where it seems that speakers have been exposed to little if any relevant experience, but yet have acquired a language with the property *P* rather than some alternative, it is reasonable to attribute *P* to UG itself. The case of structure dependence of rules is a familiar simple example...

[Chomsky (1976, p. 303ff), italics mine]

These comments suggest a somewhat more refined interpretation of the Best Theory Rule than Chomsky himself has followed in recent work, including "Conditions

on Rules"--namely, one in which universals "less interesting" from the viewpoint of the italicized statements are inadmissible.

This methodology makes the Universal Grammar hypothesis in a sense more falsifiable, and certainly facilitates the evaluation of competing metatheories. Of course, I don't want to suggest that no universal is visible in a reasonable corpus; I only want to point out that arguments for Universal Grammar are more convincing if we resist the temptation to posit trivial universals.

As we proceed with our study, we will occasionally run into cases where the Best Theory Rule conflicts with the methodology just advocated. In such cases we will freely adopt the "less restrictive" view, especially when this contributes to the overall simplicity ("elegance" or "symmetry") of the metatheory.

2. How Phrase Structure rules are written

Chomsky identified in LSLT itself one of the most serious deficiencies of the classical TG framework, namely, that it constrained the PS basis less narrowly than one might wish:

For example, in the abstract development of the level of syntactic categories, we might attempt actually to define "Noun," "Verb," etc., as primes of this level, fixed elements that may occur in the description of many languages. Or we may simply define "syntactic category" in such a way that the Nouns of English, for instance, turn out to constitute a single syntactic category, though there is no way of associating this category with some category in another language that we might also like to consider to be nouns. The former result would of course be a much more powerful one. In our discussion of syntactic categories and phrase structure we will not be able to approach such a construction. We will merely suggest certain formal conditions that the primes of these levels must meet, and we will make no attempt to construct primes with, in some sense, a fixed "content" for all languages. It is not clear what sort of basis of primitives would be required for this much more ambitious undertaking. In the discussion of transformational structure, however, we will be able to give a much more concrete interpretation of certain of the primes of this level, and it will make sense to ask, in some cases, whether distinct languages have the same transformations.

[op. cit., S19]

...That is, there will be no attempt here to define Noun Phrase within general linguistic theory, but only to define "constituents" in such a way that Noun Phrase, Verb Phrases, etc., will turn out to be constituents, and in fact, different constituents...

[op. cit., S50.1]

In the past dozen years, a good deal of work in generative grammar has sought to remedy this weakness. While it remains true that the primes of P are not assigned a fixed content, the variety of undefined primes on this level has been significantly reduced, so that many of the symbols that LSLT treated as elementary are now in effect defined in terms of others. For example, while the definability of Noun remains controversial, virtually everyone agrees on the definition of Noun Phrase in terms of Noun.

This kind of reduction is effected by the use of complex symbols as the non-terminal symbols of the base (see Aspects and "Remarks" for the application of this term to lexical items). LSLT anticipated this in the systematic use of subscripts for selectional subcategorization. The big move, however, was the introduction of superscripts, or bars, into the non-terminals; this notation, though implicit in choice of labels, had no systematic meaning in the earlier theory.

The empirical grounds for the choice of schemas came to light in the form of certain redundancies in the formalism for selection. Let NP_i, for example, be the

minimal Noun Phrase containing a certain Noun, N_i . Now the external distribution of NP_i will typically be constrained in terms of the properties of N_i . But the *Aspects* system, for example, missed this generalization because it required us to list all of the irrelevant material that could intervene between N_i and implicated conditioning context (*Aspects*, Chapter 4, S1.2, p.155). The X-bar theory, in contrast, with the attendant conventions of "feature percolation," makes what is admittedly the expected case expressible by the simpler rule.

Now let NP_i and N_i be as before, and let Det_i be a demonstrative or an article immediately contained in NP_i . Then Det_i and N_i will typically agree in various inflectional features. This can, of course, be stated transformationally; but the X-theory directly explains it.

These facts show that the name "Noun Phrase" should not be entirely "devoid of content," as the classical theory had claimed. That is, the occurrence of "Noun" in "Noun Phrase" is more than an arbitrary whim of the linguist's notation. This was noted by Lyons (1968), cited by Bresnan (1976a, p .19). Thus we put $N_i = NP_k$ for

some small positive integer k (say 3), and $N = N_{\emptyset}$. We say that each N_j , $0 \leq j \leq k$, is projected from N_{\emptyset} , calling N_{\emptyset} the Head of the Noun Phrase; similarly, Det_i is its Specifier.

Parallel arguments and terminology apply to the other major categories. In addition, we have certain "cross-categorical" generalizations which the new notation helps us express more naturally than before. For example, Chomsky shows in "Remarks" that NP's headed by derived nominals are not, contra Lees (1960) and Aspects (Chapter 4, S2.3, p.186), related to the corresponding sentences (Caesar's destruction of the city, for example, to Caesar destroyed the city) by a "Nominalization" transformation. But this means that we need another way to capture the parallels between the Possessive NP's in such constructions and the Subjects of analogous sentences as well as the parallels between the *of*-phrases of the NP's and the Objects of the sentences; similarly for the *by* phrases in the Passive alternates (the city's destruction by Caesar, for example, alongside of the city was destroyed by Caesar). The solution is to take NP and S as parallel domains for the relevant selectional rules and NP movements, generalizing the definitions of the grammatical relations across categories.

And so we find occasional but significant correspondences in grammatical relations and domains elsewhere in the grammatical system. Of course, this sort of parallelism is far from perfect; NP, for example, has no Complementizer or Auxiliary, and its Subject, unlike that of S, is optional. It is crucial not to be misled here by the use of the notation "[Spec,X]" for both Determiners and Aux; Specifier, like Head and Adjunct, is a relational notion, and as such has no formal value as a label in a phrase marker--see Aspects, Chapter 2, S2.2 p.68ff. In any case, the parallelisms among Specifiers are much weaker than the other symmetries mentioned above, since there is no categorial identity between Determiners and Auxiliaries as there is between, say, the Subjects of sentences and Noun Phrases.

Thus Noun Phrases, intuitively speaking, have less internal structure than Sentences (taken as projected from Verbs); the other major categories, PP's and especially AP's, have even less. I take it that the manifest asymmetries among the categories are as significant as the parallels. In fact, the very failure of a general isomorphism among the categories is part of what justifies our distinguishing them in the first place. If

language-particular conditions indicate a generalization of some grammatical relation across categories, the base schema may provide rules for that purpose; to the extent that the primes of P have a "fixed content for all languages," some parallels may even be favored between categories related in certain ways by their feature composition. But, in general, uniform projection is not forced.

From this point of view, attempts to impose structural parallelism mechanically across the board in the bar system are fundamentally misguided, so that it is no surprise if they have led to obscure notation and confusing analyses. The first to recognize this, as far as I know, is Williams (1971), who projects V through 4 ranks and N through only 3. While his treatment of the -ing constructions is probably wrong in some details, the general argument is certainly sound enough to support this insight.

Jackendoff (1977) also recognizes the principle, when he says

Note that the X Convention says nothing about what to do with non-parallel structures. Hornstein (1975) objects to the generalization of the Subject relation on the grounds that many other

aspects of S's and NP's are not parallel--for example, S's have Auxiliaries and Complementizers, and NP's have Determiners. But these differences are irrelevant: the X Convention says only that when parallels exist, they must be expressed. (Chapter 3, n. 5)

although his commitment to what he calls "the Uniform Level Hypothesis" (we say "Uniform Projection" instead, so as to reserve "level" for its classical meaning) compromises his understanding of it. Why else, in light of Williams' insight, would he insist upon imposing full "three bar" structure on even minor categories? The resulting prolixity of nodes is not illuminating in the least. Yet Jackendoff (1977) and most other linguists continue to assume that the bar theory is "devoid of content" unless uniform projection is universal across categories or even languages.

Thus the Uniform Projection Hypothesis sometimes leads us to attribute more complexity than is natural, as in the case just mentioned and in the case of most treatments of AP and PP. In other cases it leads to the converse formal artifice--that of depriving some categories of their inherent structure. Jackendoff himself commits this fallacy in a weak form, in taking

$$V \xrightarrow{3} (\text{Comp}) V \xrightarrow{3}, \quad \text{instead of, say, } V \xrightarrow{4} \text{Comp } V \xrightarrow{3}, \quad \text{thus}$$

allowing for arbitrary strings of Complementizers and disallowing structures in which Complementizers are obligatory (cf. LSLT Chapter 7, S54.2, Axiom 7, p. 188). He attacks, however, a stronger version of the fallacy, which is committed by Hornstein (1975): assuming a Uniform "two bar" Projection Hypothesis and observing that S, taken as headed by V, apparently has more complexity than the Hypothesis allows for, conclude that S does not project from V after all. It was in refuting Hornstein, in fact, that Jackendoff wrote the footnote quoted above.

Since in any case Sentences do have more complexity than Noun Phrases--Emonds (1979) suggests that two ranks really might suffice for nominals, by reanalyzing nonrestrictive relatives--Jackendoff's N³ adjuncts--as parentheticals--the defender of the Uniform Projection Hypothesis can only reply (page Jackendoff's ambivalence): "But why should we want to project S from V in the first place?"

2.1 On the corner

The only general answer to this is that the classical arguments for the X-bar theory become meaningless if we do not; I will explain this shortly. But first I must

introduce a symbol that will play a crucial role later on. I define X_1 (read "X-corner") as the category of highest rank projected from X , where X is a major category. Thus $N_1=NP$, $P_1=PP$, $A_1=AP$, and $V_1=S$ (not VP). Under the Uniform Projection Hypothesis this notation would be redundant, since we would have $X_1=X^k$ for some constant k , and we could thus use X^k in the statement of all generalizations in which I propose to use X_1 . It turns out, in fact, that virtually all of the good generalizations subsumed under the Uniform Projection Hypothesis concern X_1 , so that once a separate symbolism is available for this, Uniform Projection falls away. In intuitive terms, then, our proposal turns upside down the usual relation of rank to category.

Perhaps the deepest claim to be framed in this notation is the Fundamental Law of Features: features percolate unconditionally up to the minimal containing X_1 node. Occasionally, under certain rather narrow restrictions, they percolate beyond this point, as we will see in SS4-5, where we will also defend another X_1 -law: only X_1 's are directly conjoined in Phrase Structure (except lexical compounds like surf and turf, firm and flatter, if these have internal phrase structure--Ross

(personal communication) on "Freezes"). For now, however, we are content to state the perfectly general rule that percolation doesn't stop until it reaches the corner. Another use of the corner is in the definition of "proper analysis", where X_1 represents the "bounding nodes" for Subjacency-Opacity. And finally, we claim that in the unmarked case of complementation, the adjunct to the head must be X_1 : in more familiar terms, only S, not S or VP, may be a sister to V, for example--cf. Grimshaw (1977, S7.3.1) and Chomsky (1977).

The need for an expression of this kind is probably the only (theoretical) reason that the Uniform Projection Hypothesis has until now been questioned only by Williams(1971, 1974); as noted above, if all categories have the same number of ranks k , then our X_1 will reduce to X^k for all X . But if we do not expect the symmetry among the categories to be perfect--that is, to be expressed as an isomorphism--then there is no reason for our cross-categorical generalizations to be stated over categories of identical rank. We require at most homomorphisms preserving the order of ranks among the categories.

Again, it is important to keep in mind that dropping the Uniform Projection Hypothesis doesn't necessarily make our theory less restrictive. On the contrary, if we were right in hinting above that the relative internal complexity of the various major categories is a function of universal conditions on the interpretation of features, then it is the standard practice that represents the less restrictive theory, by requiring the unconstrained use of vacuous rewriting rules wherever one category has less structure than another.

With this, we can return to the question, "why project S from V, that is, why put $S=V1$?" As far as I know, the Uniform Projection Hypothesis is the only reason for hesitating to do so. If we reject it, then a number of good reasons come to light for treating Sentences as verbal, or equivalently, VP's as "small clauses" (Williams again!).

In particular, re-consider the original argument for introducing the bar theory: that NP's have in the base Subjects, Objects and ~~by~~-phrases just like S's. If we now say that the Noun destruction is the head of the NP Caesar's destruction of the city don't we have to say

that ~~destroy~~ is the head of the S~~Caesar~~ ~~destroyed~~ the city so that the associated selectional rules can apply in parallel fashion in parallel domains?

Or consider Sag's dilemma over the specification of Gapping remnants (1976, S3.4, pp. 262-266): his Uniform Projection system has two ranks and five categories, N, V, A, P, and S; thus he does not project S from V. Now on his analysis, Gapping may leave as remnants N^2 , A^2 , P^2 , and S^2 , so that one would like to write X^2 at the relevant position in the structural description. But this would overgenerate by permitting V^2 remnants, and Sag as much as admits that he has no independently motivated means of filtering out the resulting bad outputs, first observed by Ross. But if we take VJ as S, and put XJ for the remnant term in the statement of Gapping, then we eliminate the overgeneration and restore the formal symmetry of the rule. (Actually, Sag's exclusion of VP remnants is probably not right in general. Compare Daphne has drunk the acid and I eaten the brownies, comparable to Sag's examples, with ?Daphne has drunk the acid and Ron eaten the brownies. Also, we probably can't specify the rank of a term in a rule (cf. S3, *infra*). If so, there is no argument here for $S = VJ$.)

The final argument we consider here is the co-occurrence restriction between Complementizers and Auxiliaries stated by Bresnan (1970): that : Tense :: for : to :: \emptyset : to... If we put $VJ=S$ and adopt the Fundamental Law of Features already argued for, then we can assimilate Bresnan's analysis to the Noun-Determiner agreement treated above. Now it might be objected that this reduction is available even without projecting S from V, if we assume that in English, Tense is in the base a daughter of S rather than of VP, and ends up in the VP only by virtue of the Affix Hopping transformation. But the argument goes through anyhow, since the opposite is surely true in other languages, such as Italian and French, while the Complementizer-Auxiliary relation is plainly universal. Moreover, the assumption that Tense is a verbal category could explain why English has Affix Hopping to begin with--namely, in order to restore on the surface the unmarked connection, which our base disrupts, between the head and the specifier of a single category. Finally, we want to preserve the LSLT (S62.1, p. 218) demand that heads must be lexical: "only non-heads can be dependant on the context for their morphological 'realization' (i.e., can belong to long components)".

We therefore assume that V projects through at least four ranks, N through at most three, P through at most two, and A through one only. The figure for V may be as high as seven or eight, if we follow Sag (1976) in insisting on full binarism in the Auxiliary, as:

He [may [have [been [running]]]]

Sag does not reach this conclusion, because he uses direct recursion, $V \xrightarrow{i} Aux \xrightarrow{i} V$, to get the binary structure. But of course this move leads us to expect arbitrarily long chains of Auxiliaries in arbitrary orders. To use successfully as much nesting as Sag does, we must assign V as many ranks beneath the Subject as we have to assign the other categories altogether. Instinctive recognition of problems of this kind is no doubt what lead proponents of the Uniform Projection Hypothesis to eliminate S from the V category.

Typically such phrase structure as we have been advocating will be imposed by canonical rewriting rules, instantiations of the following schemata:

$$\begin{array}{c}
 X^{j+1} \rightarrow \dots X^j Y^j \dots \\
 \dots Y^j X^j \dots \\
 \dots \text{Spec } X^j \dots \\
 \dots X^j \text{ Spec} \dots
 \end{array}$$

where X , Y stand for major categories and Spec for minor categories, with j an appropriate integer. We call Y^j an Adjunct in the resulting phrase marker, and if $j=0$, we call X^j the lexical Head and Y^j its complement; otherwise X^j is the extended Head.

Normally, Specifiers and Adjuncts will appear on opposite sides of the head; when this is not the case we say the Adjunct is also the Subject (or major Specifier); we then also designate, from among the Specifiers between the Subject and the Head, the one that is furthest from the Head, calling this the main Specifier. Tense, for example, is the main specifier in a sentence.

Normally also at least one of the "... " in the schemata is instantiated as null. This accounts for the typically "binary" character of basic phrase structure--

cf. Chomsky (1961, F+K p135).

Observe that these schemata disallow "nonmaximal" complementation, such as the use of bare VP's or reduced S's in V^1 . As far as I know, no one has ever proposed such a structure for anything but complement clauses--NP reduced, say as a direct Object--so symmetry demands that we exclude it in general. As noted above, this turns out to be justified by a closer examination of the facts of clausal complementation--Grimshaw (1977), Chomsky (1978).

There is one kind of construction, however, in which a head does not project its highest rank. Gerunds, for example, are probably introduced by a rule something like $NJ \rightarrow NJ \text{ -ing VP}$; this would improve Williams' analysis by giving the correct placement of *-ing* a free ride on the standard Affix Hopping transformation--cf. Jackendoff (1977, Chapter 9). We will treat this as an instance of the specialized *non-canonical* rule schema:

$$X^j \rightarrow \dots \text{Spec } Y^i \dots$$

$$\dots Y^i \text{ Spec} \dots$$

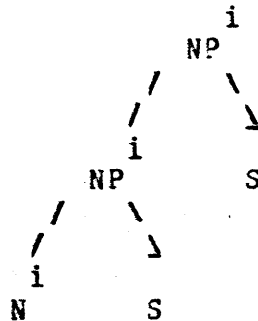
where X, Y, and Spec are as before.

Observe also that the canonical schemata do not provide for direct recursion (cf. LSLT S54.2, Axiom 7, and our earlier discussions of Jackendoff and Sag). In fact, we assume that the only case of this is coordination of XJ, given by the non-canonical schema:

XJ --> XJ Conjunction XJ

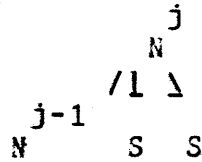
This forces us to use grammatical ellipsis to derive many constructions commonly treated as resulting from direct phrasal conjunction. We defend this result, as promised above, in S5.

The restriction also excludes the current treatment of stacked relatives. Vergnaud (1974), for example, advocates a tree of the form



for these constructions (note that this analysis violates phrase structure Axiom 7, if we follow Aspects in introducing Relative clauses in the base. Cf. LSLT,

S108.1), while Jackendoff (1977) has



Neither of these is generated by our schemata. But this is actually an advantage, since the excluded structure is in fact superfluous.

This is so because of the possibility, apparently envisioned by Kuroda (1968, n. 19), that relative clause stacking involves extraposition of one relative within the other--this is Vergnaud's (1974, Chapter 1, n. 13) exegesis, which seems sound. The idea is to use a constituent structure like $[N [\begin{smallmatrix} i \\ V \end{smallmatrix} \dots V]]$. It doesn't matter if "extraposed" relatives are base generated in place or really moved there, as long as the construal rule replacing Kuroda's Extraposition can take a restrictive relative pronoun as its antecedent. Vergnaud's only objection to Kuroda's idea (Jackendoff seems to ignore it entirely) is the failure of relative clause extraposition within nonrestrictive relatives. But on his own analysis, these have a radically different syntax from that of restrictives, so Vergnaud's argument is inconclusive.

I think that the system developed here strengthens the classical theory of phrase structure in a rather more natural way than other versions of the X-bar Convention. The alert reader will have noticed, however, that we have omitted altogether discussing a certain application of the notation, namely the use of " F^i " as a term in the Structural Description of a transformation, where F is a specified value of some feature and i is a fixed rank. This symbolism has been argued for at length by Bresnan (1976a), who unfortunately frames her proposals within a rather too loose theory of Structural Descriptions and uses a Uniform Projection Base.

Both of these defects could probably be remedied, but we will not here pursue Bresnan's approach. The reason is the apparent soundness of van Riemsdijk's attack on the decomposition of the major categories into features. While I am not convinced that these arguments are correct, it would take us too far afield to decide the issue here. This is unfortunate inasmuch as our schemata fail to force the kind of asymmetries which, in attacking the Uniform Projection Hypothesis, we have supposed to be necessary in principle. One would like to overcome this by restricting the instantiation of the schemata as a function of the

feature composition of their terms. For a rather more constructive treatment, based on a new theory of subcategorization, see George and Hoffman (1979); for a rather different view, see Jackendoff (1977).

2.2 Fragment of the English base

Here are some of the rewriting rules we will take over from previous analyses of English phrase structure, reformulated under the corner notation. We put $V = V^k$, assuming $k \geq 4$, following Williams as before. Again we leave the exact value of k undetermined, so as to avoid the problem already mentioned of nesting in the Auxiliary; for this reason we also omit most of the Predicate Phrase rules.

$$V^k \rightarrow \text{Complementizer } V^{k-1}$$

$$V^{k-1} \rightarrow N^1 V^{k-2} \dots$$

...

$$V^1 \rightarrow V^0 (N^1) \dots$$

The "... " on the right represent adverbs of various kinds and other clausal adjuncts. Where the position of these is crucial, we will adopt Williams' results.

As for the structure of N1, we assume essentially the framework of "Remarks on Nominalizations". We depart from this, however, in three respects: (i) we adopt Kuroda's analysis of stacked relatives, as in S2.1 infra; (ii) we assume that non-restrictive relatives in English are in fact parentheticals, as Emonds (1973, n.6; 1976, pp. 7-8, n.6; 1979) in an article applying Banfield's (1973b) penetrating analysis of parentheticals (recall that Emonds does not "lower" the parenthetical but rather upgrades a preceding constituent by shifting it to the right); and (iii) we reanalyze certain "Prepositions" as Case markers introduced by the rule expanding N1.

The third point is surely the most difficult. It is based on the conjecture that no significant generalization is stated over the whole class of elements that have been treated as Prepositions in the bar theory--Directionals; Heads of Time and Place Adverbs; various Particles; of, about, Dative to, Passive by, etc.--that does not hold outside that class. Of course many criteria isolate singly the proper subclasses named, and some may unify the whole class with other categories; but perhaps nothing would be lost if we held that Preposition, as usually understood, does not represent a natural class. This

appears to have been anticipated in LSLT, S39; compare Siegal (1974). Now this conjecture, even if true, admittedly does not prove in itself that the class in question does not form a major category, much less that a proper subclass of it should be shunted off into the Case system for N1, as we claim. Still, the absence of "defining characteristics" for P is suggestive, given the apparent availability of plausible tests for the other major categories, especially since the usual assumption that P is a major category causes difficulties elsewhere in the general theory.

One such difficulty arises from the fact that various Adverbial categories supposedly realized as P1 are distinguished by their "nearness" (in the sense of embeddedness) to the Verb--cf. S2.1 supra on Williams and Ross (1973). This would evidently lead us to frame phrase structure rules that, in introducing the hypothetical major category P1, specify which of its strict subcategories must develop from it, even though such statements are beyond the power of Context Free rules and apparently not needed in the distribution of the other major categories. But if we treat each of the supposed subcategories of P as a separate major category, this

problem does not arise.

Similarly, certain Prepositions (*of, to, about...*) allegedly heading V Complements are individually governed by the associated V₀. This is already problematic, and it becomes positively alarming when we observe that some of these have their "Objects" further selected by the governing V₀; for example, *talk* requires a Concrete N after *to*, where *complain* requires a Human N. Such behavior is completely atypical of selectional rules if we take the Prepositions in question to be lexical Heads. Hence we feel safe in reanalyzing them as Case markers, i.e., as Specifiers of N]. Note also that this theory affords us a possibility of deriving the core effects of Chomsky's (1978) "Case Filter": "*lexical N, unless it is assigned Case." This would hold in Deep Structure because the only rule expanding N] introduces Case. It would hold further in Surface Structure iff there is no rule deleting the "dummy carrier" Case. Actually, I believe that English does have a Case deletion system that for the most part neutralizes the Nominative-Accusative distinction (see (Sapir (1921, Chapter VII)), so that the full effect of the Case Filter may not even be wanted at the surface. (Consider, for example, sentences like *Me and my brother*

are leaving, which are quite acceptable for most speakers--up to the effects of schoolteachers. The analogues of such forms are out of the question in languages systematically distinguishing Nominative and Accusative; cf. the German sentence *Mich und meinen Bruder gehen. For independent evidence that morphological neutralization of Case can indicate loss of the opposition in syntactic structure, see van Riemsdijk and Groos (1979), who do not, however, draw our conclusion about Prepositions.)

Observe that the suggested reanalysis avoids the fallacies of earlier attempts to eliminate the category P (Generative Semantics, "The Case for Case"), in that we are not reducing the whole class to any one of the other categories--see George and Hoffman (1979), Hoffman (1980).

3. How Transformations are Written

As in the classical theory, each transformation is characterized by a pair (Q, t) , where Q is a restricting class ("structural description" in today's parlance) and t is an elementary transformation ("structural change"). Elementary transformations will be defined as in LSLT, although we use some (self-explanatory) abbreviations below.

We define a restricting class as a finite set of sequences all of a fixed length over the vocabulary $\{U, Cat, Str\} \cup P$, where

U is the identity element of the concatenation semigroup, sometimes written e ,

Cat is an element unique to the transformational level, the so-called "constituent variable" more properly, "constituent schematic letter"),

$Str \neq Cat$ is another element unique to the transformational level, the so-called "string variable" ("schematic letter"), and

P is the set of primes of P , the syntactic category symbols, excluding designation of rank.

Our approach differs from that of LSLT on the following points:

- o We expressly use schematic letters to effect what LSLT achieved only indirectly, through the construction of "families of transformations"; the latter are eliminated here, along with the

associated uniqueness requirements. ("Conditions on Rules," n. 9)

- o We require each term of a sequence in a restricting class to be simple--a single symbol, with no designation of rank--whereas LSLT permitted such terms to be arbitrary strings of primes, freely distinguishing rank (though only implicitly, since the bar theory was only implicit--cf S2, above).

Normally a restricting class Q contains just one sequence, say $R = (R_1, \dots, R_k)$, none of whose terms is \bar{U} . In this case, the complexity of Q is $@g(S) \prod_{i=1}^k @g(m)(R_i)$, where

(1) $@g(m)(P) = 2$, where P is a major category

$@g(m)(Cat) = 3$

$@g(m)(Str) = 4$

$@g(m)(Str) = 5$, where P is a minor category

Naturally, we always choose the least complex formulation that guarantees, under the interpretation of grammatical statements to be sketched in S4, that the transformation being defined will generate the positive evidence motivating its inclusion in the theory. "The logic of this approach is essentially that of the theory of markedness." (Chomsky, 1973).

Occasionally it is useful to combine rules whose

restricting class and elementary transformations are sufficiently similar. It is for this purpose that we use restricting classes with two sequences and sequences including U terms. If, for example, we were to diagram Wh-Movement as

(2) Comp ... (Case) Wh ...
 ↑ |

we could write more formally

(3) $Q = (\text{Comp}, \text{Str}, \text{Case}, \text{Wh}, \text{Str}),$
 $(\text{Comp}, \text{Str}, \text{U}, \text{Wh}, \text{Str})$

as the associated restricting class. In determining the complexity of such a restricting class, we precede as before, but count "repeated" elements only once. In the style of LSLT, this fact is indicated by the following notation for Q:

(4) $\text{Comp}, \text{Str}, \text{Case}, \text{Wh}, \text{Str}$
 U

We can easily extend the evaluation procedure given above to incorporate this idea. Formalizing the extension is straightforward, and we omit to do so here; henceforth, we take the extension as given. Under the weightings we have proposed, the "logic of markedness" now guarantees that non-singular restricting classes will be used only to

effect parentheses, as in the Wh-Movement example. If we further put

$$(5) @g(m)(U) = 1$$

we guarantee that this is also the only use of U in restricting classes.

We have seen that this theory of restricting classes is somewhat more restrictive than that of LSLT. *A fortiori*, it is narrower than recent theories proposing to enrich the classical descriptive apparatus with Boolean or quantificational conditions on restricting classes. Compare, for example, Peters and Ritchie (1973) or Bresnan (1976a). Bresnan uses the full power of Boolean description to account for the spectrum of "Pied-piping" possibilities in various Wh-Movement constructions, and, to the best of my knowledge, this is the only halfway careful attempt to show such apparatus to be helpful in explaining grammatical phenomena. Unfortunately, however, this effort fails at a crucial juncture: nothing in Bresnan's system rules out in principle the possibility that the variation in Pied-pipability should be arbitrarily different from what is observed in the literature. Though I do not take at face value the

judgements Bresnan accepts, there is, in fact, a good deal of agreement among speakers. Therefore, we must assume that grammar-specific statements are not the right way to describe the phenomena. To solve the problem we must find the right interpretation of the A/A Condition, as Bresnan observes; for this reason, we must postpone further discussion until S4, where it will appear that a more restrictive theory may explain the situation better.

Our theory of restricting classes thus allows less descriptive latitude than most. In fact, it seems that one version of Chomsky's (1976) Minimal Factorization Condition is a theorem of our system, given the evaluation procedure proposed above:

(6) Weak Minimal Factorization:

a sequence in a restricting class cannot contain two successive categorial terms (Cat or elements of P) unless one or the other is satisfied by a factor changed by the rule.
(Adapted from *op. cit.*, p. 312)

As Chomsky points out, "This condition in effect extends to the context of the structural change the requirement of analyzability as a single constituent that we have tentatively imposed on the factors changed by the rule."

In the same work Chomsky suggests an even more "radical restriction in the expressive power of transformations":

(7) Strong Minimal Factorization:

a sequence in a restricting class cannot contain a categorial term not satisfied by a factor changed by the rule.

(Adapted from *loc. cit.*)

Chomsky says that he thinks that this "is probably too strong a condition," *op. cit.*, p. 314. This is because eliminating contextual restrictions from transformations causes grammars to "overgenerate massively". Despite his reservations, however, Chomsky considers the strong version at length, proposing to eliminate at least some of the overgeneration via Emonds' Structure Preserving Hypothesis and via considerations of "surface interpretation". This is the sort of reasoning that has since led to the "Move @g(a)" format for transformations, which was already anticipated in "Conditions on Rules".

Evidently, the theory of restricting classes we have outlined permits more variation in the statement of transformations than Strong Minimal Factorization does. Nevertheless, we do not adopt the more restrictive theory

here as the null hypothesis; this is the first of the cases mentioned in S1 where we suspend the Best Theory Rule in order to eliminate a universal of the "less interesting" variety. Thus we do not need a special universal to tell us to include a "center variable" in the major movement rules--any corpus adequate to show the bare existence of Wh-Movement or NJ-Preposing, for example, will already force us, given the evaluation procedure already proposed and the interpretation of transformations sketched in S4, to include the schematic letter Str explicitly between the categorial terms of the restricting class--so from our view-point, Strong Minimal Factorization is nonetheless ad hoc for being universal.

In any case, I share Chomsky's early misgivings about the "empirical" tenability of Strong Minimal Factorization. I doubt, in particular, that his strategies for eliminating the "massive overgeneration" entailed by the elimination of context predicates are sufficiently general, despite his recent infatuation with the "Move @g(a)" scheme. As Chomsky himself has pointed out (in his Fall, 1979, lectures at MIT), the Structure Preserving Hypothesis is really plausible only in the case of the substitution elementary (as for NJ-preposing)--

adjunctions, even though not structure preserving in the narrow sense, should probably be allowed to apply freely in subordinate domains. And it is precisely in its treatment of Wh-Movement and various rightward movements as engaging the Substitution elementary that Emonds' theory appears badly strained. But if we relax the Structure Preserving Hypothesis as indicated--in effect taking simple adjunctions as automatically structure preserving--then its potential for equilibrating the "Move @g(a)" system is severely reduced.

As for the role of interpretive principles in eliminating overgeneration, I fear that "only their unclarity disguises their irrelevance" (LSLT S13.1). Of course we cannot conclusively rule out advances in this direction, but I, for one, cannot make any sense out of the proposals now in the "Move @g(a)" literature, such as Chomsky's (1976, p. 317ff) attempt to explain, in terms of "predication", differences in the operation of N1-Preposing in N1 and V1 domains.

In conclusion, I would like to mention two points worth keeping in mind even if Strong Minimal Factorization does hold for the major movement rules. First, the

condition does not inevitably lead to the "Move @g(a)" format for these transformations; rather, it leaves room for rule-particular specification of "landing sites", because they, like the moving phrase, are "changed" by the operation of the rule--that is, mentioned in the (reduced) statement of the elementary transformation. Secondly, Strong Minimal Factorization is almost certainly hopeless for transformations other than the major movement rules--Chomsky admits as much when he speaks of it as characteristic of "core grammar"--such as the (so-called "construal") rules responsible for the distribution of grammatical formatives, rules like those dubbed "strictly local" by Emonds, and certain deletion operations (the "grammatical ellipses" of S5, *infra*, for example).

4. How the rules work

No argument about how grammatical rules are written can really make sense except in the context of an understanding of how the rule notation is applied in constructing structural descriptions. For this reason, our technical discussion has so far verged on the uncomfortably abstract and perhaps on the dangerously vague. We may of course console ourselves with the thought that we had to start somewhere, and that comparable woes beset the exposition of any deep theory. But we can no longer put off deciding, at least in part, how the rules work.

4.1 Phrase structure

For present purposes we assume that (initial) Phrase-markers are produced as in LSLT, Chapter VII, with the conventions modified in the obvious way to permit the use of complex symbols (category-rank pairs, in the sense of S2), in addition to primes, in the phrase structure description.

Readers of George (1975) may realize that I do not in fact advocate the system that we will be using, but rather favor a view like that of Clark (1975), too feebly

endorsed in my (op. cit.), which takes phrase structure rules to be cyclic in the same sense as transformations. This is, of course, contra the later Chomsky (1965, Chapter 3, especially p. 134ff), who indeed treats both rule types as cyclic, but in opposite senses. Since, however, a defense of my real beliefs is not now feasible, we will rely on the so-called "standard theory" for the remainder of this dissertation.

Readers of LSLT will recall that the possibility of recursive application of phrase structure rules, though admitted in Chapter VII, was later excluded under the finite kernel hypothesis (Condition C, S54.1; cf. S569.1, 91.4) of the "Transformational Analysis" Chapters IX-X. They should also see that the cyclic application of transformations, while simulated in the clumsy mechanism of S91.5, was by Chomsky's own admission never adequately grounded in the formalism:

This effect of Condition 5 indicates that very likely this condition is not the correct one. It seems that it should be possible to have the proper mappings apply to each transform X before it enters into a further transformation, and still to have these mappings reapply in the correct way to any further transform Y including X, without disturbing X internally. I have, however, found no good general formulation that meets these requirements.

[p. 395]

Against this concession, we can now only repeat that our present use of the Aspects sketch is not to imply that the "standard" system, with its reinstatement of recursion in the base and its informal use of the transformational cycle, does provide the "good general formulation" approximated by Condition 5.

Now it was repeatedly stressed in the classical studies that the fundamentals of constituent structure cannot be determined if attention is confined to the base rules--in case after case, the behavior of phrases under transformation provides the decisive test between conflicting grammatical theories. Its not so much that we have to decide "how transformations work" in order to decide "how phrase structure rules work"; rather, we have to see the transformations in action just to figure out what the phrase structure rules are.

Nevertheless, "description in terms of phrase structure" is logically prior to "transformational analysis", essentially because two vital functions of the latter, characterizing the concepts of factorization and proper analysis, are keyed by definition to the phrase structure relation E. This, the so-called "is a"

relation, gives the categorial labelling of substrings of a terminal string, relative to a stated occurrence and to a stated "phrase structure interpretation". If B is a constituent of a Phrase-marker K and occurs as a P in K , we write $E(B, P, K)$. Our interpretation of E will differ considerably from that of LSLT.

In the first place, we have reduced E from a five-place predicate to a three-place one, mainly for clarity of exposition. Specifically, the fourth term Z of the classical relation, which represented the terminal string under analysis, is here absorbed into the class K of "representing strings" that give the intended interpretation of Z (this will always include Z --cf. LSLT, end of S81.1). Also, a variety of difficulties concerning the definition of "occurrence" are hidden under our schematic letter A , which stands for the LSLT pair (X, W) .

A more important respect in which we depart from the classical understanding of E proceeds from our adoption of the bar theory (S2, supra). To see this, recall that the new theory of phrase structure uses complex symbols in the place of primes to label constituents of major

categories--reconstructing "N" and "NP", for example, as the pairs (N, \emptyset) and (N, k) , respectively, where k is the rank of the "maximal projection" N_1 . But the argument P of $E(B, P, K)$ is by definition a prime, so that the classical relation no longer has the intended effect under the bar theory. We could, of course, alter the definition quite simply so as to use complex symbols for the argument P , but it appears that a coarser function, not distinguishing the rank of a constituent, is sufficient for transformational analysis; we have already hinted as much by confining terms of restricting sequences to categorial primes (S3, *supra*). No rule, for instance, that takes a nominal constituent N for its i -th factor will have to specify directly whether the nominal must be an N or an N_1 . In the unmarked case, by the A/A Condition (*infra*), it will always take the latter; it will reach lower only in the presence of an independently motivated "context predicate"--in English, say, the category Det or Case for term $i-1$. Similarly, VP-Preposing is a "marked rule", presumably using an Aux context to access its non-maximal target.

(Depending on how we formalize our use of complex symbols in representing strings, we may have to modify the

definition of E even to characterize the coarser relation; since the needed changes would be trivial, we won't bother with them here.)

Another generalization of E is needed to state the Fundamental Law of Features, "don't stop until you reach the corner" (cf. S2 *supra*), and the allied conventions for the percolation of syntactic features. It appears that the required operations fall into four general cases, three obligatory and one optional:

(1) Fundamental Law of Features

If B is the Head or minor specifier of C, then all features of B are assigned obligatorily to C.

(2) Law of Addition

If P is a categorial prime added to B by a transformation, then the feature [+P] is assigned obligatorily to B.

(3) Major Specifier Law

If B is a major specifier of C, then all features of B are assigned obligatorily to C that are not basic features of any Q such that C "is a" Q in the intended phrase structure interpretation.

(4) Free Percolation

Optionally assign to B all the features of all its immediate constituents.

Now the first of these provisions, as we have noted, is implicit in any version of the bar theory, so it should

be familiar to everyone, and we will not dwell on it here.

The second Law is nicely illustrated by the action of the rule that prefixes the formative WH to the categories NJ, AJ, etc. (but apparently not to VJ--cf. "Conditions on Transformations" S13). By the familiar constructions of derived constituent structure, when this "WH Placement" rule applies to some category Q, the resulting sequence again belongs to the same category Q; that is, we get the so-called "Chomsky adjunction" configuration $[WH^Q Q]$. What the Law of Addition adds to this is that the derived term is also marked [+WH].

Somewhat trickier is percolation under the Major Specifier Law. To understand this, we need first to define "basic feature":

P is a basic feature of Q iff P is a lexical subcategory of Q or a minor Specifier of Q.

Number, for example, is a basic feature of N, but not of V, A, etc; similarly for the "inherent" selectional features of N (Concrete, Human, etc--Cf. Aspects). Conversely, Past, Progressive are basic features of V, but not of N, A, etc. And if we now assume that WH is introduced in the base as the Specifier of VJ (i.e. as the "Complementizer"), but assigned (as above) exclusively by

transformations to the other categories, then WH is a basic feature of V, but not of N, A, etc. In this case, WH will obligatorily percolate "past the corner" of a major Specifier of any category other than V. For example, the N] whose books is marked [+WH] because its Possessive Subject whose is under the Law of Addition marked likewise. In contrast, the relative clause V] in The man [who introduced you] left is not marked [+WH] by virtue of the Major Specifier Law (even though its Subject who is [+WH]), because WH is a basic feature of V. This line of reasoning enables us to bring the descriptive core of the Left Branch Condition (Ross, 1967) under A/A without transgressing Minimal Factorization, i.e. without the use of "mixed terms" in the restricting class of Wh-Movement--compare Bresnan (1976a) and Woisetschlaeger (1976).

The Free Percolation provision should present no difficulties of interpretation; note however that it is an all-or-nothing proposition.

Our statement of the percolation conditions, even if more precise than most treatments of the problem, are not quite adequate from the point of view of formalism. This

is best remedied for the three obligatory statements, we conjecture, by a suitable extension of the theory of derived constituent structure--specifically, by defining three new "is a" relations E in the recursive construction of $g(r)$ (see LSLT, Chapter IX, especially 7-9 S87.7, p. 368). For optional percolation, the best apparatus would seem to be a special kind of transformation. Such a rule cannot be formulated within the framework we have proposed for the statement of language-particular transformations. This is not too alarming, however, since the provision for Free Percolation will presumably be a universal in anyone's theory.

As already intimated, our particular characterization of feature percolation will figure below primarily in the theory of proper analyses. There are, however, independent measures for the correctness of our decisions here. Implicit in every version of the bar theory proposed so far, for example, is the following

(5) Feature Conflict Filter

If the values [+F] and [-F] percolate to the same node, then assign *.

This, in tandem with the Fundamental Law of Features,

accounts for a wide variety of typical bar theory evidence, such as the obligatory morphological agreement between Determiners and their Head Nouns.

Perhaps more obscure, but no less important, is the fact that this filter, under our theory of percolation, also explains a phenomenon which we may call Conjunct Agreement--that is, the requirement that conjoined constituents, in addition to belonging to the same major category, must share certain "internal" features of syntactic structure. The general observation appeared in LSLT (SS77.1-2, p. 301ff; S115.1, p. 557ff), but it was there left open just which features are relevant in a given case. Our theory answers this question precisely, and, as far as I can see, correctly. We do not have, for instance,

(6) *The boy who and the girl embraced is my neighbor

(7) *Who and the girl embraced?

(8) *What made who and the girl so gay?

because the feature values [+WH] and [-WH] percolate obligatorily from their respective conjuncts (by the Law of Addition and the Major Specifier Law), thus causing a filterable conflict at the immediately dominating N] node.

Now Ross (1967, 4.134) indeed attempts to explain the first case (6) under his Coordinate Structure Condition, but the third example (8) shows that such an analysis is not sufficiently general, since the star appears even when there is no requirement of movement (cf. S5.3, *infra*).

The feature of Number, conversely, being (unlike WH) a basic feature of N, need not percolate "past the corner", so two conjoined N1's may differ in Number without provoking a vicious conflict under (5).

A third case provides even more interesting evidence. As we saw above, the selectional feature Concrete is a basic feature of N; hence it need not percolate from two conjoined N1's under the Major Specifier Law, so that N1-conjuncts may in general disagree in this feature:

(9) The French admired Joan and her piety
On the other hand, to replace admired with a Verb taking a Concrete object, we must apply Free Percolation, so that a feature conflict rules out the result:

(10) *The Inquisition burned Joan and her piety
(Again we assimilate ordinary selection to strict subcategorization by making it strictly local--cf.

Aspects and S2, *supra*.)

These three specimens seem typical of the way that Conjunct Agreement sorts features; if they are, then the whole range of phenomena supports our theory of percolation and the Feature Conflict Filter. Nevertheless, the work on selection presented in George and Hoffman (1979) suggests that the need for stating (5) as a filter (in the narrow sense--see *infra* on "Filters and Mistakes") arises as an artifact of an inadequate analysis of features. Our provisions for percolation, however, would survive unscathed in the more restrictive theory, as would the logic of our explanation of Conjunct Agreement. For this reason, and because the results of George and Hoffman do not for the most part bear on the fundamental issues of this dissertation, we will stick with the formulation (5) for the time being.

4.2 Transformations

At last we turn to the operation of transformations, beginning with the concept of *factorization*:

- (11) Let $q = (W_1, \dots, W_r) \in Q$, a restricting class. Then (Y_1, \dots, Y_r) is a factorization of K by q iff both (i) and (ii) hold:
 (i) $Z = Y_1 \dots Y_r$ is the terminal string of K
 (ii) $\exists (Y_j, W_j, K_j)$ for each j , $1 \leq j \leq r$
 In this case we also say that (Y_1, \dots, Y_r) is a
 *factorization of K by Q .

This is modelled on the LSLT definition of proper analysis; following more recent practice, we reserve the latter term for those factorizations meeting certain additional "conditions on transformations": Strict Cycle, A/A, etc.

In fact, we will probably want to relativize the concept of proper analysis to level assignment, excluding different factorizations depending on which "component" a given transformation belongs to. Lexical insertion rules and minor deletions, for example, will require "strictly local" factorizations, and reordering transformations will meet some "bounding" condition like Subjacency--cf. Chomsky (1973) and the literature on "Freezing": Ross (1974), Culicover and Wechsler (1977), and Cattell (1976); see also Hawkins (1980). Major deletions ("ellipses")

will obey a parallelism constraint--cf. S5.2 *infra*, on the Law of Congruity. We thus end up with a series of concepts @g(a)-proper analysis, one for each grammatical level. We then define "generate" so that an @g(a)-level rule applies only on a factorization that is an @g(a)-proper analysis.

Unfortunately, this is not the place to offer a definitive theory for what we have just sketched. For now, we state just two conditions, apparently met by all proper analyses:

(12) Definition

A factorization (Y_1, \dots, Y_r) of K by $q = (W_1, \dots, W_r)$ is cyclic iff there is no proper constituent Brof K such that $U \neq W_i \neq \text{str}$ implies Y_i is included within B .

(13) Definition

A factorization (Y_1, \dots, Y_r) of K by $q = (W_1, \dots, W_r)$ is inclusive iff there is no distinct factorization (Y'_1, \dots, Y'_r) of K by q such that $U \neq W_i \neq \text{str}$ implies that Y_i is included within Y'_i .

(14) Strict Cycle Law (Chomsky-Williams)

If a factorization is an @g(a)-proper analysis, then it is cyclic.

(15) Law of Inclusion ("A/A", Chomsky-Kayne)

If a factorization is an @g(a)-proper analysis, then it is inclusive.

The Strict Cycle Law in effect excludes a factorization if all of its categorial terms are part of the same proper constituent; the Law of Inclusion disallows a factorization if one of its categorial terms is a proper constituent of the corresponding term of another factorization, unless the converse holds for another categorial term.

By the way, the investigators named in our statement of the Laws ((14) and (15)) don't necessarily subscribe to our formulation of them. Williams (1971, 1974), for example, presents his version of the Strict Cycle as an extrinsic ordering condition, while Kayne (1975) appears to favor the "absolute" interpretation of A/A. By citing these authors, then, we mean only to attribute to them the first (comparatively) clear development of the relevant concepts. Also, our implementation of the Chomsky-Kayne A/A should not be confused with Bresnan's (1976a) formalization of the same idea. Our Law of Inclusion differs from Bresnan's "relativized A/A" not only in its simplicity, but also in the tameness of the associated

theory of restricting classes, outlined in S2 supra. As promised there, we want to offer some evidence that our (methodologically preferable) theory is in fact adequate.

Before we proceed to this task, however, we pause to make some further observations about the formal status of our conditions.

To begin with, our Laws ("if-then") may be strengthened to a definition ("if and only if") once they are combined with needed level-specific conditions on proper analyses, as we remarked above. But this is not to imply that there are no additional components of the desired definition that are completely general--that is, conditions stated over proper analyses at every level. In particular, something in the neighborhood of the Tensed-S and Specified Subject Conditions of Chomsky (1973) is most likely such a completely general component of the definition of proper analysis. See George and Kornfilt (1978) for a preliminary formulation as the "Finite Phrase" and "Subject Accessibility" Conditions, which we may conveniently group together under Chomsky's term "Opacity". Of course, our use of this word is not meant to commit us to Chomsky's view that the Conditions it

covers should be translated into "filters on Logical Form"--see below on "hyperindexing". In fact, we continue to interpret them as Conditions on Rules (components of the definition of proper analysis), assuming in particular that a rule penetrates an opaque domain by mentioning in its restricting class a minor Specifier of that domain.

Secondly, our statement of the Strict Cycle Law is too strong under "standard" assumptions about the base; it would, taken literally, prevent the application of all but root transformations. We could, of course, complicate it so as to avoid this result--cf. Peters and Ritchie (1973), Lasnik and Kupin (1977)--but such a complication is unnecessary under the theory advocated in S4.1 supra, in which base rules are cyclic "in the same sense" as transformations. In fact, our primary reason for supporting this "bottom-up" construction of (initial) Phrase-markers is the attendant simplicity of (12).

Finally, consider that the definition of "occurrence" (LSLT S20, p. 109ff) constitutes a rudimentary notion of indexing, which is refined and elaborated somewhat under the conception of factorization/proper analysis. This more sophisticated indexing system is in turn further

advanced via the successive constructions of derived proper analyses, "roots", and derived phrase structure interpretations (LSLT SS86-7). We now conjecture that this classical development of indexing is sufficient for the correct statement of the theory of grammar--in particular, that the "inscription" à la trace theory of numerals on the "nodes" of collapsed diagrams is superfluous. On the basis of this conjecture, we henceforth call the indexing procedures presupposed in more recent work "hyperindexing".

From our present viewpoint, then, grammatical theorems stated over "Logical Form" in the sense of Chomsky (1976) and much later work would in general "fall out" as artifacts of a somewhat too reified analysis. Now I have already tried to support this reduction (which, to be sure, may be a "reduction" only in the loose sense) in public discussion (George, 1979; George and Kornfilt, 1979) by arguing that the Logical Form literature wrongly conflates two kinds of interpretation ("deep" and "surface") that were properly separated in EST (Chomsky, (1970)), thus destroying a good account of the fact that certain aspects of meaning are preserved under transformations while others are not.

More recently I have found another line of reasoning with the same purport; this newer attempt to refute the arguments for hyperindexing proceeds from the perspective of filter theory, as follows. We find the primary indications for hyperindexing in the possibility it affords us of translating various conditions on rules (in our terms, components of the definition of proper analysis) into filters on Logical Form, where this reformulation is supposed to have special plausibility on analogy to certain morphophonemic statements, like the "clitic templates" of Perlmutter (1963) and the ~~*for-to~~ rule of Chomsky and Lasnik (1977). But this, I submit, is a false analogy, because it relies on a mistaken view of these "surface filters". For these are generally seen as transformational rules that "assign *" to arbitrary surface configurations, and on this view it is only natural to extend the device to eliminate "overgeneration" at other levels. But I propose a different understanding of the surface filter data, one which does not easily generalize to Logical Form. This alternative is based on a hypothesis I call the "Spec-Spec Constraint", which holds that any sequence of adjacent minor Specifiers must be mapped into morphophonemic representation by a special

rule particularly designed for the sequence in question. If this claim can be sustained, then the surface filters would in general fall out as items in a list of the residue of language-particular spelling rules. Thus they would have, individually, no theoretical significance, and there would technically be no sense to the operation "assign *", at least at the surface. See S4.3 *infra* for further discussion.

Of course this argument does not prove that there can be not filters on Logical Form. On the other hand, it does undermine the independent plausibility of the idea, thus supporting, at least indirectly, our conjecture against hyperindexing.

Now it should be apparent to readers of Chomsky (1973) how our interpretation of various Conditions as components of the definition of proper analysis goes hand in hand with our choice in S3 *supra* of a moderately restrictive theory of restricting classes over the broader theories represented by LSLT and the narrower ones tending toward the "Move @g(a)" rubric. To illustrate the workings of our system, we consider now the Wh-Movement phenomena known as "Preposition stranding" and the

"Wh-Island Condition".

Given the reanalysis of Prepositions suggested in S2 ^{supra}, the term "Preposition stranding" is of course a misnomer. As is well-known, "Prepositions" heading Time and Place Adverbs in fact resist stranding; so do many Directionals (compare He jumped out (of) the window with Which window did he jump out *(of)?, where the usually optional Case marker of becomes obligatory just to protect the Directional from stranding). In fact, it is fair to say that free stranding is limited to Case as against the other elements that have been analyzed as Prepositions. This holds, I predict, for all languages, because of the Finite Phrase Condition (Dutch stranding of Directionals does not falsify my prediction because it is, by virtue of its limitation to "R-pronouns" (van Riemsdijk, 1977), not free in the intended sense).

Case stranding, in turn, is obviously marked from a cross-linguistic point of view, as has often been observed. This follows from the Law of Inclusion and the Fundamental Law of Features, given that the simplest formulation of the restricting class for Wh-Movement is in our framework as

(16) $Q = (\text{Comp}, \text{str}, \text{WH}, \text{str})$

For suppose we have assigned WH to $N]^{-1}$ in the configuration

(17) $\begin{matrix} & & -1 \\ & & [WH^N N]^{-1} \\ & N & \end{matrix} \text{Case}^{\wedge} [\quad]]]$

Then the feature [+WH] will obligatorily percolate to $N]^{-1}$ from its Head $N]^{-1}$ under the Fundamental Law, so that the smaller Wh-phrase will never be analyzable under the third term of Q, by A/A. Thus we need a more complex Q, such as we suggested in S3 supra, to permit Case stranding:

(18) $Q = (\text{Comp}, \text{str}, \text{Case}, \text{WH}, \text{str})$

But under the logic of markedness, we will proceed to this more costly Q only for cause, whence the comparative rarity of Case stranding among the languages of the world.

Similarly, consider the "Wh-Island Condition", that is, the fact that subordinate interrogatives in English are islands to Wh-Movement. Within the general framework of "Conditions on Transformations", there are several possible explanations of this fact:

- o A/A. Assuming as usual that interrogative V]'s are marked [+WH], a statement of Wh-Movement not complicated to distinguish the major category of the "target predicate" will never reach into an indirect question.

- o **Superiority.** Since A/A is a special case of Superiority, as noted independently by Hendrick (personal communication), Wh-Islands are explained *a fortiori* on the assumptions of the last item. For those interrogatives that are not the domain of internal Wh-Movement (i.e. subordinates without *if/whether*), moreover, the explanation holds even without the assumption that Comp-->WH distinguishes interrogatives, given the Strict Cycle Law.
- o **Opacity.** The familiar "escape hatch" theory of the Complementizer follows from a sufficiently complex statement of the Tensed-S and Specified Subject Conditions, again, given the Strict Cycle.
- o **Subjacency.** The familiar "escape hatch" theory of the Complementizer follows from a sufficiently complex statement of the Subjacency Condition, again given the Strict Cycle.

Thus there is massive redundancy in the "Conditions" treatment of Wh-Islands. Of course redundant characterizations of "overgeneration" are not as such objectionable--in fact, they can often be used to account for gradations in the acceptability of various ungrammatical forms. Nevertheless, there is evidently considerable overkill among the Conditions on this score, so much so that it is hard to see how to even begin using the redundancy to explain graded judgements. Much recent work has sought to remedy this situation.

Such efforts fall into two general cases: one based on a rather feeble reconsideration of the English

phenomena motivating the Superiority Condition, and one based on a much more interesting (brilliant, in fact) study of a language which appears to violate the WH Island Condition, namely, Italian.

The former effort centers on the observation that certain Superiority violations in English are not as bad as those originally noted by Chomsky (1973). Any native speaker, for example, will find

(19) *I don't remember what books which people read more acceptable than

(20) **John knows what who saw
as Chomsky (1978, n. 43) has pointed out, citing Kayne (personal communication). Now this is supposed to show that "Superiority is not what is causing the problem" in (20). But of course the stated contrast shows no such thing, at most, it indicates that Superiority is not the only thing wrong with (20). To use this kind of data to show that Superiority is in fact irrelevant to (20), we would have to claim that (19) is perfectly grammatical, which Chomsky expressly declines to do--and wisely enough, since (19), if better than (20), is plainly worse than

(21) I don't remember which people read what books
Similarly

(22) *Who did you read what to?
is considerably less acceptable than

(23) What did you read to who?
So the whole pattern would actually ~~support~~ the Superiority analysis of Wh-Islands if we had another Condition to explain the difference between (19) and (20). And in fact we do find one, in the form of the Spec-Spec Constraint already suggested. Thus (20)** violates two conditions and (19)* just one, while (21) is perfectly grammatical.

The moral of the story is that one cannot refute a Condition on rules merely on the grounds that some violations of it cause worse unacceptability than others--cf. Chomsky (1977a, p. 21). As long as the lesser stars are there at all, we should rather use such evidence to validate redundancies in our analysis of "overgeneration", as mentioned before.

The principle here stands, of course, even if it turns out that we want to replace the Superiority

Condition with some analogue, on other grounds. Suppose, for example, that finding

(24) What did you read ^{when}
^{where}
 grammatical, we reject Superiority because it would prevent the needed Wh-Movement derivation, given likely assumptions about the phrase structure of Time/Place Adverbs, viz., that they are superior to clause-mate Direct Objects. We should then replace it with the corresponding "Priority Condition", putting "precedes" in place of "is superior to" in the statement of the definition (cf. Lasnik and Kupin), and the earlier remarks about graded judgements will apply *ceteris paribus*. (~~Where did you read what~~ is no counterexample, since it can be generated by the independant rule preposing place adverbs.)

I henceforth assume that the Priority Condition, as just sketched, does correctly replace the Superiority Condition. Note now that the A/A Condition holds once more independently (if at all), since it does not follow from Priority as it did from Superiority. On the other hand, the formulation of the Law of Priority is so similar to that of the Law of Inclusion that we already suspect

that they jointly follow from some stronger Condition. Though we cannot offer the indicated reduction here, its desirability is further suggested by another graded set of judgements:

(25) I wonder who read what

(26) *What do you wonder who read

(27) **Who do you wonder what read

Now (25), being perfectly grammatical, violates none of the conditions, the intermediate case (26) violates just A/A-Priority; while the worst offender (27), depending on which Wh-word moves first, involves either a simple violation of Priority, followed by an operation simultaneously violating A/A and Priority, or a simple violation of A/A, followed by a violation of the Strict Cycle.

I now propose that these Conditions--A/A-Priority, Spec-Spec, Strict Cycle--are in fact the only ones relevant to the core phenomena of the Wh-Island Condition. Opacity and Subjacency, in particular, while they have their own separate motivation (Raising, CNPC, etc.), have nothing essential to do with it.

This would allow us to sustain the simpler

formulation of Opacity and Subjacency, taking X] to be the "bounding categories", as suggested in S2, at the same time explaining the acceptability gradations just discussed; which brings us to Rizzi's (1977) "Violations of the Wh-Island Condition in Italian..." The violations reported involve mainly (but cf. *infra*) movement of a WH phrase to form a relative clause, as in Rizzi's (6a):

- (28) Il solo incarico che non sapevi a chi
avrebbero affidato è poi finito proprio a te
"The only charge that you didn't know to whom
they would entrust has been entrusted exactly
to you"

Now of the four explanations of Wh-Islands made available in "Conditions on Transformations", Rizzi considers seriously only the two that we have already rejected, Opacity and Subjacency; he dismisses Superiority on the grounds we have already refuted (see his n. 11), and he, like everyone else heretofore, disregards the A/A explanation altogether.

Of the two approaches he does contemplate, Rizzi seems to favor Subjacency on "empirical" grounds, though he can appeal to no considerations of symmetry (such as our "simplicity of bounding") to confirm his choice. Thus it becomes imperative to review his arguments in search of

an alternative interpretation of the data.

At bottom, Rizzi sees the problem as one of deciding between two strategies for explaining the contrast between Italian and English: one rooted in the differential applicability of Subjacency, and the other in the differential applicability of Opacity. The former takes ⁻¹VJ to be a bounding category in English but not Italian, exempting Wh-Movement from Opacity in both languages, while the latter distinguishes English from Italian by including "WH traces" among the "anaphors" for purposes of the Logical Form Filter interpretation of Opacity. Rizzi, as we have mentioned, seems to favor the Subjacency-based approach (see his n. 25), primarily because he can use it to explain a further difference between the two languages, namely, that Italian extracts constituents of NJ to the front of VJ more readily than English. For example,

- (29) Gianni, di cui so che una foto è apparsa
recentemente sul "Gazzettino",...

is apparently perfectly grammatical, while its
English translation

- (30) *Gianni, whom I know that a picture of appeared
recently in the "Gazzettino",...

is out of the question. Following Chomsky, (1977b)

"On Wh-Movement", Rizzi assumes that the English star is due to Subjacency, taking V⁻¹I to be a language-specific bounding category, where Italian apparently contents itself with the universal bounding categories X₁. Assuming then that neither A/A-Superiority(-Priority) nor Opacity interferes, he also gets out of the putative bounding category divergence the Wh-Island difference between the two languages.

Observe, however, that (30) violates Subjacency even if we don't take V⁻¹I as bounding, as long as we take the category [of whom] to be bounding, which we must if we identify it as N_I (as proposed above), and probably even if we admit P as a unitary major category (cf. van Riemsdijk). Thus the English example we really need to test against the Italian is the "pied-piping" counterpart of (30):

(31) (*)Gianni, of whom I know that a picture appeared recently in the "Gazzettino",...

This, though probably not fully grammatical (because of the Finite Phrase Condition) is surely far better than (30), and, given the stylistic markedness of pied-piping elsewhere in English, hardly assimilates to better-known violations of Subjacency (CNPC, etc.). Furthermore, Rizzi

himself admits that this kind of fronting out of N1 is extremely limited--apparently, di is the only really mobile Preposition in such cases (op. cit., n. 16), and even with di many forms we would expect on Rizzi's account do not work:

- (32) ??L'uomo di cui la sorella maggiore è innamorata
 di te è Gianni
 "The man whose older sister is in love with you
 is Gianni"

Thus Rizzi's attempt to correlate Wh-Island differences with differences in extraction from N1 does not seem really convincing. Furthermore, his interpretation of Subjacency deprives it of any naturalness it may derive by identifying "bounding category" with the independantly motivated concept of "cyclic domain", which is in enough danger under our interpretation of the Williams cycle, which takes all major categories to be cyclic domains.

But of course the general strategy of tying together apparently unrelated contrasts between the two languages is correct. Accordingly, we now offer an alternative hypothesis which is in the same spirit, but which restores the symmetry of Opacity-Subjacency, thus returning us to our A/A-Priority approach to WH Islands.

The intuition guiding our reanalysis of Rizzi's observations arises from a review of the work of Cinque (1977), who argues persuasively that Italian analyzes its equivalent of the so-called "Left Dislocation" construction of English into two subsystems: one, limited to root V1, is base-generated as in English, while the other involves a movement rule which we may (pace Cinque) schematize as

(33) @g(f)-Movement
 Comp str @g(f) str
 ↑_____↓

where @g(f) is a suitable feature with reflexes in intonation and interpretation--the moved phrase must represent "old information" (op. cit., S2.1, p. 406).

What I propose is very simply that the apparent violations of the Wh-Island Condition in Italian are in fact effected by this independently motivated rule, not by Wh-Movement, which presumably functions just as in English. This would mean that there is no need for any special statement in the grammar of either language to account for their difference on this point, which would follow ~~ohne weiteres~~ from the fact that Italian, but not English, utilizes a movement analysis of "Left

Dislocation". There are several obvious strategies for putting this hypothesis to the test, but we can only sketch these here.

First, the theory requires the Relative Pronouns in Rizzi's examples to have the feature @g(f), so that a more precise analysis of this feature as it works in Cinque's construction is urgently needed. Given such an analysis, we might hope to explain why Interrogative Pronouns, for the most part, do not escape from Wh-Islands--thus we do not have

(34) *Chi ti domandi chi ha incontrato?[Rizzi's (7a)]
 "Who do you wonder who met?"

(35) ??A chi non ti ricordi quanti soldi hai dato?
 "To whom don't you remember how much money you gave?"

would follow if we found that question words normally do not have @g(f). Note also that this kind of example is "considerably improved if the [escaping] WH phrase...is made 'heavier'":

(36) A quale dei tuoi figli no ti ricordi quanti soldi hai dato?
 "To which one of your sons don't you remember how much money..."

(see Rizzi's n. 5). We conjecture that such

"heavier" interrogative phrases differ from their less marked counterparts precisely in having @g(f). We might also hope to find the fronted Partitives mentioned above to be distinguished by this feature.

A second problem for our hypothesis concerns the fact that all of Cinque's examples have Clitic Pronoun "doubles" associated with the moved item, while none of Rizzi's do. Thus to support our analysis, we need a suitable theory of Italian Clitics.

Finally, we have not precisely specified the derived constituent structure to be created by @g(f)-Movement. The fronted element appears in some cases to land just to the right of the triggering Complementizer, but there are unsolved problems here.

Thus considerable research into Italian syntax must be carried out before we can thoroughly justify our theory. Nevertheless, we will assume for the remainder of this dissertation that our general A/A-Priority analysis of Wh-Islands is adequate, so that the alternative complications of Opacity/Subjacency are unnecessary and probably incorrect accounts of the same data. This will come in especially handy in §5.3 *infra*, where we suggest

that WH Subjects do not move vacuously into the adjacent Complementizers; for if true, this Vacuous Application Prohibition rules out the Opacity/Subjacency explanations as insufficiently general.

4.3 Appendix: Filters and Mistakes

Many syntacticians have proposed rules of the form:

- (1) If a string Z with constituent interpretation K' is analyzable at the level L by a restricting class Q, then assign * to Z,

where K' is a P-marker in the extended sense (i.e. a "derived interpretation" in the sense of LSLT, SS87,93.1), and where L ranges over every known level of grammatical description (but is normally restricted to one such level). It is routinely observed that such "filters" have "transformational power" in that they rely on an understanding of the concept of a proper analysis (LSLT, S82) for their correct application. Less appreciated is the fact that the hypothetical elementary transformation "assign *" is not defined. Moreover, the restricting classes Q of these filters violate virtually every conceivable characterization of Minimal Factorization.

Chomsky (1976) presents two versions of Minimal Factorization:

(2) Strong Minimal Factorization

A restricting class Q cannot mention a categorial term unless it is satisfied by a factor changed by the rule

(p. 312)

(3) Weak Minimal Factorization

A restricting class Q cannot mention two successive categorial terms unless one of them is satisfied by a factor changed by the rule

(p. 312) Notice that (2), given the standard assumption of minimal change, implies that constant categorial terms may always be separated by variable material. That is, (2) introduces "implicit variables". In what follows, we assume that (3) is the correct version of Minimal Factorization. In any case, both versions presuppose that much of the descriptive power of standard transformational grammar is eliminated from the theory--in particular, that full-scale Boolean conditions on analyzability cannot be effected. Of course, the filters used in descriptive practice require all the latitude of the Peters and Ritchie (1973) formalism, so that they violate Minimal Factorization a fortiori.

We conclude that no rules of the form (1) figure into

the general theory of grammar underlying the (bio-)logically correct description of natural languages. To sustain this verdict, given the formal defects of filter theory, we need only show that any plausible argument for an "assign *" transformation can be naturally explained away. With this in mind, we trace the history of the controversy, concluding that all filters so far proposed can be eliminated under one or more of the following independently motivated principles (4)-(9):

- (4) Each utterance must be represented by a normal string on every grammatical level, and these representations must be properly linked by the mappings $\bar{\alpha}$.

This is nearly the definition of the concept "generate" (see LSLT, *passim*).

(5) Spec-Spec Constraint

A rule mapping a (minor) Specifier into morphophonemic representation must mention (adjacent) categorial terms on both sides of the changed element in its restricting class.

Taken in conjunction with another condition (not expounded here), (5) has (5) as a limiting case:

(6) Stuttering Prohibition

Local repetition of a (minor) Specifier is ungrammatical (up to the effects of morphophonological reduplications).

Many languages have special "gliche" rules to evade this constraint--cf. the fusion of Plural and Possessive in English, and Hale, Jeanne, Platero (1977) on the Hopi Obviative.

The Spec-Spec Constraint and its limiting case, like Minimal Factorization, form part of the definition of the concept of a restricting class. They are thus "conditions on rules" in the narrow sense, and only indirectly "conditions on application", or "on derivation", or "on representation", or "on binding". As examples of the latter kind of condition, we count the components of the definition of a proper analysis:

- (7) Conditions that affect all sorts of factorizations such as A/A, Strict Cycle, etc.
- (8) Conditions eliminating factorizations w.r.t. the assignment of the associated transformation to a stated component of $\bar{\Phi}$ such as Recoverability, Freezing, etc.
- (9) Conditions defining derived constituent structure via the construction of derived proper analyses and roots, such as Structure Preservation, Upgrading, etc.

In the historical sketch, we would consider:

- o How filters in the bad sense were anticipated, but avoided, in the LSLT discussion of "dummy carriers" (SS91.6, 109.4, n.28).
- o How Aspects attempted to clarify the "filtering

effect of transformations" by appeal to the ad hoc diacritic #, viz., "by defining the relative clause transformation in such a way that it deletes the boundary symbol # when it applies. Thus if its application is blocked," e.g. by conditions guaranteeing the recoverability of deletions, "this symbol will remain in the string. We can then establish the convention that a well-formed surface structure cannot contain internal occurrences of #..." (Chapter 3, esp. p. 138)

- o How Ross (1967) extended (modulo a striking qualification w.r.t. the grammaticality-acceptability distinction) the Aspects approach in his (3.27) "Internal NP over S" constraint and his (3.41) "Output Filter on Post-verbal Constituents".
- o How Perlmutter (1968) consummated the foregoing efforts by offering the first general theory of surface filters qua rules; how he attempted to provide an analogous explanation of "obligatory control" via the device of "deep structure filters".
- o How Chomsky and Lasnik (1977) reanalyzed Perlmutter's work, moving to eliminate deep structure filters and to use "negative statements" in place of his "positive statements".
- o How REST advocates are trying to give an artificially unitary account of the notion of "misgeneration" by stating a homogeneous list of "filter on logical form".

To illustrate our proposal for the reanalysis of the data for filter theory, we consider the *that-t filter, reserving other cases for our presentation of the completed paper. Our treatment of the relevant evidence

is guided by the observation that each structure starred by the filter is analyzable not only as ~~that~~-t, but also as ~~that~~-Aux, that is, as a sequence of adjacent (minor) Specifiers. But such a sequence, by the Spec-Spec Constraint, can be spelled only by a special component of the morphophonemic mapping particularly designed to do so. Hence, in the unmarked case, the structures in question will not meet (4), and are thus not generated in the first instance, much less "misgenerated". English, of course, is typical in this respect (up to the action of a marked spelling rule for the sequence Rel-~~that~~-Aux). Observe that this approach assimilates the *~~that~~-t phenomena to an analogous filter of Chomsky and Lasnik's, namely *for-to.

Now, in a language such as Italian with a rule of Subject Pronoun Deletion (control of empty subjects by agreement markers, or whatever), the marked rule spelling ~~that~~-Aux will be independently motivated by the occurrence in the corpus of sentences in which Pronoun Deletion manifestly applies to subordinate Subjects. This is how we capture Perlmutter's generalization (as reformulated by Chomsky and Lasnik).

There may be some independent evidence for our

analysis, in the form of a contrast in the acceptability of various that-t patterns in Dutch. As is well-known, speakers of this language differ in their evaluation of strings like

(a) (*)Wie denk je dat gekomen is?

Maling and Zaenen (1978) assert that there is a true dialect split here, tying the acceptability of forms like (a) to the optionality of dummy er in non-initial position. As was pointed out by George (1978b), this claim is hardly credible in the context of their general approach, since it suggests the possibility of an er-drop rule, thus undermining their efforts to show that (contra Chomsky and Lasnik) the *that-t filter is not universal. Perhaps even more telling is the fact that many native Dutch linguists dispute their factual claims. According to Koster (1977), "It is easy to make [(a)] much worse by deeper embedding and a slightly different word order":

(b) *Wie denk je dat Bill zei dat t is gekomen?

while such examples "are much better when the subject-trace is followed by a (pronominal) object NP...":

(c) Wie denk je dat t het gedaan heeft?

Koster also cites

(d) *Ik vraag wie of t is vertrokken,
 noting that the *~~that~~-t filter "works here correctly".
 But he fails to note that

(e) Ik vraag wie of vertrokken is,
 to which Chomsky and Lasnik must also "assign *", is
 perfectly grammatical. (We wish to thank Hans den Besten
 for this grammaticality judgement.)

To us, these contrasts suggest that our focus on
 analyzability as ~~that~~-Aux rather than as ~~that~~-t is on the
 right track, and that Koster's evidence, though crucial,
 really has little to do with the other factors he cites
 (depth of embedding, pronominal status of intervening
 objects). To substantiate this, we must consider a
 variety of other Dutch examples, like

(f) (*)Wie denk je dat is gekomen?
 We predict that this, though perhaps not as bad as (b), is
 considerably worse than (a), (c) and (e), probably for all
 speakers, and at least for all speakers that reject (a),
 (b) and (d).

5. Coordination and Ellipsis

In this section we defend the claim put forth in S2., that conjunctions are introduced in the base only by rules of the form $X_1 \rightarrow X_1 \text{^} \text{Conj} \text{^} X_1$ (S5.1). If this is right, then we need elliptical analyses of a broad range of cases currently treated as involving only "phrasal conjunction." We justify therefore the necessary ellipsis rules and certain of their crucial properties on independent grounds, giving in particular an explicit account of the "internal parallelism" required between the two conjuncts in elliptical coordinate constructions (S5.2). The central concept of this theory we call "congruence," that sense of analogy which S1.0 offered as the main technical achievement of this dissertation. Its definition, in tandem with the corner conjunction hypothesis just repeated, explains a new observation about the distribution of so-called "Across the Board" WhI Extractions (S5.3). And finally, congruence helps us to describe the "Comparative Subdeletion" constructions of Bresnan (1975), confirming Chomsky's (1977b) suggestion that C-Sub must be divorced from certain less restricted comparative constructions, which appear, unlike C-Sub, to involve Wh-Movement. (S5.4).

5.1 Restricting Phrase Structure Conjunction (Back to the Corner)

The grammar of coordination has always been a centerpiece of generative studies, so much so that a history of the treatment of conjunction could make a vivid summary of the vicissitudes of linguistic theory. In particular, one of the most straightforward of the early arguments for transformational analysis ran

One extremely serious deficiency of this [phrase structure--LNG] grammar, and of the conception of grammar on which it is based, is that we really have no good way to introduce the rule for conjunction as a statement of the grammar. But the simplification of the conjunction rule was one of the fundamental criteria for the determination of constituent structure. Hence if the grammar cannot incorporate this rule, the proposed approach to a demonstration of validity is undermined, and there is considerably less justification for the particular form that the grammar has assumed. Of course, inability to state the conjunction rule is itself a serious defect, irrespective of the fundamental character of this rule as a criterion for the establishment of constituents.

Roughly, the conjunction rule asserts that 184 is an optional conversion:

184

...X...-->...X[^]and[^]X..., where X is a prime

But our framework has no place for such a statement as 184. It will be recalled that each rule of the grammar operates on the forms as they appear in the last step of the derivation in question, at the point when the rule is to be

applied. Our characterization of the process of producing derivations from grammars does not provide for the possibility of taking into account the history of the elements that appear in this last step of the derivation in question. But if 184 were introduced into the grammar as a statement, it would be necessary to know the history of the elements that appear in the last step of the derivation at the point where 184 is to be applied. That is, we should have to know which substring is represented by a single prime, i.e., is a constituent, and in fact, what sort of constituent it is. (Alternatively, we should have to introduce one statement of the form 184 for each prime.)

>[S74.1, p.290]

This paradigm argument from missing generalizations is substantiated in S115.1, where the rule for conjunction is formalized as a family of generalized transformations. Chomsky observed immediately that a theory incorporating his 184 could not ~~ohne weiteres~~ account for examples like

- (1) they elected John president and Bill vice-president
- (2) John was elected president, and Bill, vice-president
- (3) he let the dog in and the cat out
- (4) Bill bought candy, and Jim, pretzels

These are from item 449 of S115.1. "Unless certain subsidiary transformations are given" beyond those of S115.1, Chomsky notes, the grammar fails to generate such examples of what he later (Syntactic Structures, p. 38, §5

n.2) called "conjunction crossing constituent boundaries," where

and does not at first sight link constituents. The distinction drawn here is so vital to later discussion of coordination that we must find a handy terminology for it. Let us therefore convene to speak of "(direct) conjunction" in clear cases of the form "...X^{and}X...", where X is a prime"; elsewhere we will say not "conjunction" but "coordination crossing constituent boundaries" or "indirect coordination". This usage capitalizes on the etymological connotations of adjacency inhering in "conjunction".

Under this convention, we can isolate one of the recurring claims of generative analyses of conjunction as follows

Format((5) Uniform Direct Conjunction Conjecture)

All cases of direct conjunction are generated in the same way, without use of the special rules creating coordination crossing constituent boundaries.

In the classical theory, this idea was embodied in the distinction we have just seen between the unitary formalization of §115.1 and the "subsidiary transformations" not given there.

By now, however, virtually everyone has forsaken Chomsky's theory that *and* is introduced exclusively by transformations--for a number of reasons that we take up later--going for the claim that at least some direct conjunction is base generated as such. But no one that I know of rejects the claim that coordination crossing constituent boundaries involves crucial use of ellipsis transformations, and it is indeed hard to imagine treating examples like ((1)-(4)) entirely in the base. Now this shift raises the question, "How much direct conjunction is base generated *as is*, and how much depends on the independently motivated ellipsis system?"

The Conjecture (5), of course, forces on us a particular answer: "All direct conjunction is base generated if any is." Now Chomsky has not accepted this conclusion, because he has not maintained (5) in his base conjunction theories (see, for example, *Aspects* Ch.3, n.11, p.225). Banfield (1973a, Ch.II), however, explicitly adopts the claim that all direct conjunction is base generated, relying on Dougherty (1970, §14, p.892; 1968, VI.1, p.239ff).

But the latter's commitment to the Conjecture is

weaker than Banfield's, inasmuch as he analyzes certain direct conjunctions of VP as crucially derived from Sentence conjunction via a "Coordinate Conjunction Transformation." This ellipsis in effect creates coordination by moving one of the VP's to be coordinated across the boundaries of the directly conjoined Sentences--cf. Dougherty (1970, S2, p.852f; cf. 1968, III.9, p.89ff). The evidence for Dougherty's rule, though not his formulation of it, is vital to our thesis, and we return to it directly.

In any case, to defend our limiting of base conjunction to X1, we must reject the Uniform Direct Conjunction Conjecture, or return to the classical theory that all conjunction is transformational. In the latter case, our thesis would be vacuously true, since there would be no base conjunction at all, even of X1. But we will not pursue this course, so we must detail the arguments for base conjunction, seeking evidence that the UDCC does not "draw the line" in the right place.

In preparation for this inquiry, it will be helpful to examine the role of the UDCC in the classical theory. This is neatly illustrated by the contrast between:

(6) My friend liked the play and enjoyed the book

(7) *My friend enjoyed and my family liked the book

LSLT (S65.1, p.224ff, items 13 and 14) took the latter string to be ungrammatical, so that the obvious grammaticality of (6) argued for putting the "major constituent break" in (8) as shown:

(8) My friend | enjoyed the book

[loc. cit., item 10]

Syntactic Structures (loc. cit.) hedged on the status of (7), admitting that it might be grammatical. But as noted there, this concession does not destroy the elegant demonstration just reviewed of the existence of the Verb Phrase. For even if grammatical, (7) "is much less natural than the alternative" My friend enjoyed the book and my family liked it, while the analogous alternative to (6), My friend enjoyed the play and he liked the book, is not preferable to it. Thus "our conclusion that the rule for conjunction must make explicit reference to constituent structure therefore stands, since this distinction [between (6) and (7)] will have to be pointed out in the grammar".

Such examples as (7) have more recently been accepted

syntactic structure.

As far as I know, no one has previously identified the intonation patterns of Right Peripheral Ellipsis with that of Heavy NP shift. Heaviness, in fact, is usually taken not as a prosodic factor but as either some direct measure of syntactic complexity, or as an extra- or quasi-grammatical variable like "information value." LSLT, for instance, observed that in V-Particle constructions

...the separability of the preposition is determined by the complexity of the NP object. Thus we could hardly have

179 the detective brought the man who was
accused of having stolen the automobile in

It is interesting to note that it is apparently not the length in words of the object that determines the naturalness of the transformation, but rather, in some sense, its complexity. Thus "they brought all of the leaders of the riot in" seems more natural than "they brought the man I saw in." The latter, though shorter, is more complex on the transformational level since it has the infixed sentence "I saw." We will see below that this is a transformational construction. A good deal of further study is needed here to determine the nature of this process and to define properly the relevant sense of complexity of the object.

As the object becomes more complex, then, the naturalness of the transform decreases. This is systematic behavior, and we might expect that a grammar should be able to state it. But it may turn out to involve probabilistic considerations for which our system has no place as it now

stands...

[S100.2, p. 477ff]

In a discussion rather reminiscent of this, Ross (1967 (3.41), p. 36) proposes an "Output Condition on Post-Verbal Constituents" in connection with his formulation of "Complex NP Shift." This move toward confining mention of heaviness to a surface filter disarms, as Kiparsky has pointed out to me, one kind of argument that the affected rules are "stylistic"--namely, that defining the complexity of the moved elements exceeds the expressive power of "core" transformations. At any rate, the move makes sense in the context of our suggestion that heaviness is defined prosodically. Ross, however, sticks to Chomsky's original "structural complexity" view, invoking "performance" in place of Chomsky's "probabilistic."

Also suggestive in this connection is the claim

that the issue is not so much the "heaviness" of the...NP, but rather the relative heaviness of it and the constituents it has moved across... (cf. Fiengo (1977, 49).

[(Kayne and Pollock (1978, p.619, n. 34)]

Again, this is quite natural if metrics provides the relevant measure (cf. Arabic), although neither Fiengo nor Kayne and Pollock reach this conclusion. As Hoffman

(personal communication) has noted, the idea of "relative heaviness" would allow us to use a single filter for Complex N1 Shift and Particle and Dative Movement, dropping pronoun restrictions from the statement of the latter; cf. Emonds (1976).

Thus we conjecture that some late prosodic rules are responsible for the gradations of acceptability involving heaviness, the entirely indirect influence of syntactic complexity, pragmatics, etc., being mediated by their rather loose correlation with intonation. See Hoffman (1980) for discussion.

But now if the "special phonemic features" are not introduced by ellipsis transformations as such, then we conceive of elliptical structures lacking them (this maneuver is typical of surface filter theories). If such exist, the Uniform Direct Conjunction hypothesis finds no support from phonemics, and certain direct conjunction may after all be elliptical.

To resolve this issue we need a good metrical theory, one that could tell us on partially independent grounds whether prosody supports the UDCC. Perhaps it is not too optimistic to hope that such a theory may be had soon--see

Kiparsky (1979) and the references cited there. I'm confident that the theory in question would not sanction the argument from prosody to the UDCC, for intonation is a notoriously unreliable indicator of syntactic structure, as noted by (references?). Citing the effect of morphophonemic "readjustment rules," Koutsoudas (1971, p. 354 n. 14) applied this observation specifically to coordination, to refute arguments that Conjunction Reduction effects "regrouping."

Another consideration that casts doubt on the phonemic argument for UDCC is that "special phonemic features" don't even seem to constitute a unitary phenomenon. The need for them varies with the construction: the requirement is very weak, if not lacking, in Left Peripheral Ellipsis, somewhat stronger in Gapping, and most striking in Right Peripheral Ellipsis. Chomsky seems to have sensed this early on, for he says "the more completely we violate constituent structure by conjunction, the less grammatical is the resulting sentence." But how does Right Peripheral Ellipsis "more completely violate constituent structure" than Gapping, or Gapping than Left Peripheral Ellipsis? There is probably a clear answer to this, but I doubt that it would support

the claim that "special phonemic features" are a necessary condition for ellipsis. Again, I think that a good metrical theory would back me up here.

This leads us back to the justification for base conjunction alluded to above. Reduced to essentials, the argument consists of two steps, both taken in the mid-1960's. One is Chomsky's (1965) observation that the elimination of generalized transformations destroys a crucial aspect of the classical treatment of coordination, so that at least Sentence conjunction must occur in the base. The other side of the argument was Carlotta Smith's claim that certain NJ conjunctions were base generated as such.

We will not now challenge the elimination of generalized transformation, but we reserve the right to do so in the future--see George (1975) for a critique of the *Aspects* arguments, and some evidence that a more restrictive theory of generalized transformations might be helpful. In any case, we want to add a qualification to the *Aspects* analysis of Sentence conjunction in the base. Namely, we will write $VJ \rightarrow VJ^{\wedge} \text{Conjunction}^{\wedge} VJ$ instead of $VJ \rightarrow VJ^*$;

that is, we exclude the "star schema" from our implementation of the base conjunction analysis of Sentences. One reason is that the use of the schema disrupts the symmetry of the base; another is that, in tandem with the plurality argument, it fosters an appeal to the more general schema $X \rightarrow X^*$, (Dougherty-Banfield), which we reject in favor of the Corner Conjunction Hypothesis.

The original (and, as far as I know, the only) argument for the "star" notation has no force, in my opinion. Chomsky (1965, Chapter 1, n. 7, p. 196ff; Chapter 3, n. 9, n. 11, p. 224ff) adopts it only for Sentences (thus abandoning the UDCC, as noted above). In favor of the notation, he reasons as follows: the binary rule can generate multiple coordination only via right-recursive or left-recursive phrase structure; but there is no evidence for choosing between the two possibilities. "In fact, there is no grammatical motivation for any internal structure" (loc. cit.) in such cases. He concludes that the conjuncts must all be daughters of the same node.

But who forces us to use the left-recursive structure

to the exclusion of the right-recursive, or conversely? The binary rule will generate both kinds automatically; moreover, Chomsky's schema will, too, along with the intended multiple branching. This suggests that the absence of evidence for internal structure cited is due to an independent property of grammatical theory, say, the conditions behind the Coordinate Structure Island phenomena (*infra*). If so, Chomsky's observation does not differentiate the two theories.

For the most part this issue will not affect my arguments below. Hence the reader who still thinks Chomsky's conclusion correct may put XJ^* in place of our $XJ^{\wedge} \text{Conjunction}^{\wedge} XJ$, without materially affecting the logic, as long as he resists the temptation to go beyond XJ .

Now let us treat the other base conjunction argument. After constructing his family of generalized transformations E , Chomsky says (LSLT, S115.1, and p. 556ff):

We have not yet accounted for the fact that the subject $NP^{\wedge} \text{and}^{\wedge} NP$ takes a plural verb...; it is necessary to add a special further transformation to effect this change of the verb from singular to plural, when $NP^{\wedge} \text{and}^{\wedge} NP$ is the subject. This can be given either as a component of the mapping $\bar{\phi}P$ or as a transformation added to the T-Marker for conjunction after E .

and

But this adjustment is ad hoc, and its formal character in doubt, as Chomsky noted. Worse still, it is empirically inadequate, because Auxiliary agreement is not the only place where the classical approach would need ad hoc adjustments like this. There are in fact a wide variety of grammatical relationships that depend on the number of an NJ, and they all treat singular NJ's conjoined with and as forming a larger plural NJ:

CERTAIN PREDICATE NOMINALS AGREE IN NUMBER
WITH THEIR SUBJECTS

- (11) John and Mary are linguists
*John are (is) linguists [Gleitman (1965: (61))]
*John and Mary are a linguist

REFLEXIVE PRONOUNS AGREE WITH THEIR ANTECEDENTS

- (12) John and Sam saw themselves (*himself)
[cf. Gleitman (1965: (98))]
*John saw themselves

RECIPROCAL PRONOUNS REQUIRE PLURAL ANTECEDENTS

- (13) Red and green complement each other
*Red complements each other
[Smith (1969: (10))]

SOME LEXICAL ITEMS SELECT PLURALS IN CERTAIN
POSITIONS

- (14) John and Mary met [Gleitman (1965: (123))]
*John met

- (15) Ice cream and cake are a popular combination
*Ice cream is a popular combination.
[Smith (1969: (11))]

- Jimmie and Tommie are a pair of fools
*Jimmie is a pair of fools
[Smith (1969: (12))]

John and Mary are alike

*John is alike...

[Lakoff and Peters (1966: (2, 2'))]

Daphne was lodged between the devil and
the deep blue sea

*Daphne was lodged between the devil

(As intimated above, the first effort to show base conjunction of NJ is that of Smith (1969); despite its late publication, the ms is cited by Gleitman (1965, n.34), so that it probably antedates other papers on the same hypothesis. Another quick historical note: Dougherty (1968, V.3 (42-5), p.235) gives examples including between to support his argument from respectively; I do not know whether anyone has used between to argue the more general selectional point before. We return to the history of NJ-conjunction later on.)

Now we obviously "miss the generalization" if we attach a rider like LSLT's to the statement of each of these dependencies. Instead, we need somehow to label the larger NJ itself as plural, despite the singularity of its conjuncts.

We might, for example, identify the grammatical formative and with the formative Plural, relying on the

Fundamental Law of Features (S2, supra) to assign the correct number to the compound. But this does not complete the argument for NJ conjunction in the base. To carry it further we must defeat the proposal that conjoined NJ's come together transformationally, the resulting NJ being labelled with plurality by a further transformation (which, unlike Chomsky's adjustment, would not directly touch the Auxiliary or any of the other items dependent on the number of the big NJ), or by percolation, as just suggested.

One defect of the latter proposal is that it requires the NJ-conjoining transformation to be extrinsically ordered before any of the rules sensitive to the number of its output. Otherwise the grammar would generate singulars where plurals are required in examples like ((11)-(15)). Now if this ordering is arbitrary--imposed only to get the right derivations in this domain--then of course it is subject to the usual objections against extrinsic ordering. Why, for example, don't we find a language with the opposite configuration of data (as we surely will not)?

On the other hand, principled extrinsic ordering is

unobjectionable, and in fact it is used throughout EST: lexical insertion before reordering rules before bound anaphora "construal," for instance. So if any of the reflexive, reciprocal, auxiliary, or predicate nominal number agreements are checked by "shallow interpretation" rules, as commonly assumed, then the theory does in fact motivate their ordering with respect to N1 conjunction, taken as cyclic.

Notice, however, that we have not by this maneuver disposed of all of the number dependencies illustrated in ((11)-(15)). The remaining cases furnish us with a more serviceable argument for base conjunction. The reasoning here is (i) these are cases of selectional rules requiring plural N1's in certain positions; (ii) selectional rules apply in deep structure; (iii) hence certain and-N1's must already be conjoined and labelled as plural in deep structure.

The affected lexical items have been called "symmetric predicates" by Lakoff and Peters (1966), but this is misleading inasmuch as it corrupts a technical usage already well established in logic and mathematics. Many items, to be sure, are "symmetric predicates" in both

senses, like similar or meet. But hug, for instance, is a "symmetric predicate" for Lakoff and Peters (being syntactically like meet), but not on the standard definition, since "x hugs y iff y hugs x" does not express a tautology. And conversely, equal (or resemble) is a symmetric predicate in the logical sense--x equals (resembles) y iff y equals (resembles) x--but not for Lakoff and Peters, since we do not have

(16) *The price of chicken and the price of rice
equal

This example turns out to have more than terminological importance. But before we get into that, let's consider that more than courtesy urges us to observe the usage with priority. For one thing, it is more accessible to the beginner, in that it is more faithful to the ordinary language meaning of "symmetric"; the newer usage is moreover misleading, in suggesting that the items designated form a natural class from a semantic point of view, while no one has offered any evidence that they do so. Also, many better expressions come to mind for the syntactic category; "silent reciprocal predicates" is more vivid, but still somewhat misleading, as we shall see. Perhaps the most accurate term would be "plural

selectors," if anything like contemporary theories is correct.

The problem with "silent reciprocal" is that it suggests an each other-deletion analysis of items like hug, such as that proposed by Gleitman (1965, E.5). The idea is to derive examples like (14) from their each other paraphrases; this approach is intuitively appealing, and fits naturally into Gleitman's insightful study. On EST assumptions, however, we cannot adopt such an account, and further inquiry supports these assumptions by showing Gleitman's theory inadequate to the data.

The starting point is equal, our specimen of a symmetric predicate that does not take "silent reciprocals"--cf. (16). This element is typical, in that very few predicates, symmetric or not, resemble hug in admitting a "silent reciprocal" use, and, as far as I know, they don't form a natural class from any point of view--syntactic, logical, or morphophonemic. Hence, unless I'm overlooking something, the reciprocal deletion rule would need an idiosyncratic exception feature.

Gleitman's solution has to be even more complex. Besides cases like (14), where only each other would have

to be deleted, there are forms in which various Prepositions are missing along with it, for example:

- (17) This problem and that are similar (to e.o.)
[cf. Lakoff and Peters (1966 (77))]
- (18) Daphne compared trace theory and generative semantics (to each other)
- (19) Daphne and de la Trachette bickered shamelessly (with each other)
- (20) A handsome and well-meaning bouncer made the mistake of trying to separate Daphne and de la Trachette (from each other)

Hence the deletion rule for silent reciprocal sentences must be further complicated to obligatorily remove certain Prepositions, and the diacritic theory must explain how the rule keeps from eliminating the wrong Prepositions. This would be more difficult than it seems at first glance. Consider (18) alongside of

- (21) Daphne and de la Trachette compared him
*(to each other)

Now we have already marked each other deletion to apply in the context of compare, and required it to remove to when it does so. But (21) shows that all this is not enough: the deletion has to know where the reciprocal antecedent is, too!

Thus the diacritic solution, like the nominalization

theories attacked in "Remarks," "requires devices of great descriptive power which should, correspondingly, be very 'costly' in terms of a reasonable evaluation measure" (op. cit., p. 31); whence the appeal to selection, the natural habitat of lexical idiosyncrasy, in premise (i) of the above figure.

But is the indicated use of this rule feature too costly? Probably yes. Consider, for instance, the Sentences (22); these cases, unlike the earlier ones, have Plural selectors which cannot occur in the same frame with each other:

(22) Ice cream and cake are a popular combination
with each other

*Jimmie and Tommie are a pair of fools
with each other

*John and Mary are alike unto (?) each other

*Daphne was lodged between the devil and
the deep blue sea with (?) each other

Here we seem to have the Plural selector in its pure state; see Dougherty (1968, VII.3, p. 280ff) for more items of this kind. They show Gleitman's rule, even if complicated as indicated earlier, to be inadequate. But if we have $N^{and}N$ in the base and the selectional apparatus needed for these cases, we should extend them to the

"silent reciprocal" items as well, dropping each other deletion and its diacritics.

Now as we have already noted, almost everyone nowadays agrees on this much. Combining this with the *Aspects* claim that clauses are base-conjoined, we have widespread acknowledgement of at least as much base conjunction as would be generated by

$VJ \rightarrow VJ^{\wedge}Conjunction^{\wedge}VJ$, $NJ \rightarrow NJ^{\wedge}and^{\wedge}NJ$

Now what people disagree on is: how much more base conjunction is there? What has to be transformational?

We can distinguish four reasonable positions on this issue, as conditions on the schema

(23) $X \rightarrow X^{\wedge}Conj^{\wedge}X$, where

I $X = VJ$, or $X = NJ$ and $Conj = and$

II $X = NJ$ or VJ

III $X = PJ$, where P is a major category

IV X is a major category

The conditions appear in the order of their utility to a defender of elliptical direct conjunction. Anyone who accepts the Extended Standard arguments for base conjunction admits at least as much phrase structure

conjunction as is induced by I, as noted. Most other investigators, like Dougherty and Banfield, have gone straight to IV. In earlier work, reported in George and Kornfilt (1978) and in George (1978a), I proposed to draw the line at II, so as to broaden the scope of coordinate ellipsis. I now think that I provides just the right amount of base conjunction, even though the latter condition appears simpler. I'll say why later on. For now, it suffices to keep in mind that my key thesis is that direct conjunction of V^i is freely created by ellipsis of V^i conjunction. This is easier to show if conjunction of V^i of lower rank cannot be base generated. But all I really need is the free-ellipsis result, not an east proof of it.

Thus my theory could even survive under condition IV, though uncomfortably, because of Dougherty's dilemma, alluded to above. This turns on examples like

(24) John lived in Boston and was beaten
by hoodlums [Dougherty (1968, III.9 (1), p. 89)]

As Dougherty observes, this sentence has to be elliptical because the Passive VP is not generated as is in the base. He therefore creates the "Coordinate Conjunction Transformation" mentioned earlier to generate this form of

transformational direct conjunction, which for him is exceptional, since he elsewhere subscribes to the Uniform Direct Conjunction hypothesis as embodied in condition IV.

Now as Dougherty notes, his theory systematically assigns two structural descriptions to active VP conjunction, one involving the ellipsis rule and one not. But there is no intuitively ambiguity here, and Dougherty seems a little disturbed over this discrepancy. In defense of his schema, he says

At present no factual arguments are known to support the contention that a semantically unambiguous surface structure may not be derived from two or more synonymous deep structures.

[ibid., n. 21, p. 324]

Maybe so, but what of the classical methodological principle that structural ambiguity should correspond to multiplicity of intuitive analysis? This maxim was crucial to a wide variety of arguments for introducing transformations: if we use only phrase structure description, we find

...examples of sentences with dual interpretations, but where we have no grounds for establishing constructional homonymity in the intuitively relevant manner.

The opposite deficiency would occur if we were sometimes led to assign several P-markers to sentences that have, intuitively, only one analysis. We can also find cases of this kind...where we have too much constructional

homonymity...

[LSLT S&6.1-2, p. 298ff]

Dougherty himself relies on the classical methodology to support his transformation in other cases--Dougherty (1968, III.9, p. 94ff)--so his repudiating it here seems unfair.

Of course this instance of "too much constructional homonymity" does not arise if we prohibit base conjunction of VP, i.e., if we adopt one of the corner conjunction hypotheses I-III, though perhaps only the most restrictive I eliminates all analogous discrepancies from the grammar of coordination. And there are further indications that I may be correct.

First we observe that the only strong arguments in the literature for conjoining anything but V involve N¹ and N¹. Unless there is as yet unknown evidence of more general base conjunction, we must suppose condition I linguistically significant; otherwise, the configuration of data would represent a metatheoretical accident.

Readers steeped in EST literature will now have grumbled that there is evidence of base conjunction beyond that allowed by I. But I'm talking strong evidence, and as

we will see the arguments in question are rather flimsy.

But first, let us consider another argument for condition I. The idea goes back at least as far as Koutsoudas (1971, p. 373). While and-NJ conjunction regularly shows plural number, as we have seen, the correct generalization is different in or-NJ conjunction,

...correct agreement relations holding in all such cases between the verb and its closest subject. For example:

- (108) The boy or the girl is running
- (109) *The boy or the girl are running
- (110) The boy or the girls are running
- (111) *The boy or the girls is running

Now we reason that the material before or represents a deep structure Sentence conjoined with the string after or, which is also a sentence, ellipsis having rendered invisible everything in the first conjunct but its subject. The NJ's linked by or thus do not form a constituent acting as subject of the following VP. Rather, the NJ before or has no direct relation to the Verb after it, whose subject is the second NJ, wherefore the agreement shown.

Explanations of this form are unavailable, as far as I know, in theories sanctioning or-NJ base conjunction.

If so, then even II allows too much base conjunction, not to mention III-IV.

Now it remains to consider the possible arguments against I. As far as I know, there are just two: one from Dougherty (1963, 1973) based on respectively, and one from Jackendoff (1977) based on "symmetric predicates". We deal first with the more recent, since the older argument will lead us into some tangential historical issues.

Jackendoff (1977: (7.62), p. 194) observes case like

(25) The same man got drunk and was arrested
by the cops...

and goes on to say

Dougherty invokes Conjunction Reduction specifically in order to deal with conjoined derived verb phrases [recall our discussion supra--LMG]. But the underlying source he needs is ungrammatical here:

(7.63) *The same man got drunk and the cops
arrested the same man...

(Actually, Dougherty (1973, p. 853, abstract) takes pains to deny that his "Coordinate Conjunction Transformation" is a reduction transformation, and might take issue with Jackendoff's wording here. Nevertheless, his rule does effect a deletion, so we won't dwell on the accuracy of

Jackendoff's reading, which is correct on one essential point, that Dougherty admits transformational direct conjunction for just these cases.)

Now I seriously doubt that Jackendoff's (7.63) is really ungrammatical; compare the same man got drunk and the same man was arrested by the cops. In any case, a condition ruling it out would no doubt be a "late" constraint on anaphora. If this is true, however, the deep structure is available as a source in Dougherty's theory, so Jackendoff's objection collapses.

Another way of looking at this point is to consider that Dougherty's theory could easily be modified to take (26) instead of Jackendoff's (7.63) as the needed deep structure:

(26) The same man got drunk and the cops arrested him

This move would leave Jackendoff only one recourse, if he still wanted to deny Dougherty's appeal to transformational conjunction: he would have to claim that (25) has "the wrong reading".

But the most prominent reading of (25) is surely found in the proposed source, and is paraphrased by

- (27) The same man got drunk and he was
arrested by the cops

This refers to a man previously mentioned (or otherwise drawn to the attention of both interlocutors). The second possible reading, the one crucial for Jackendoff, seems to me rather forced; it would be the one where no particular man was singled out before, paraphrased by the same man got drunk as was arrested by the cops. Judgements are subtle, to be sure, but I myself would invariably prefer the as construction if that were what I meant. In any case, it is unlikely that there is a structural ambiguity here; if not, there is no argument against deriving (25) from (26) via (27).

Jackendoff raises this problem in the context of his theory of relativization, after taking Vergnaud, whose theory he rejects, to task for using "symmetric predicate" arguments! He gives examples like

- (28) John whistled and Mary hummed at equal
volumes [from op. cit. (7.59), p. 182]

which, though they feel like Right Peripheral Ellipsis structures, haven't any straightforward source:

- (29) *John whistled at equal volumes and Mary
hummed at equal volumes [from loc. cit. (7.60b)]

Now I'm not sure how to generate (23)--perhaps from something like John whistled and Mary hummed, and at equal volumes--but if Jackendoff really thinks this kind of case vitiates the argument from "symmetric predicates", he has no business pulling the same argument against Dougherty in the next breath.

Actually, Jackendoff seems confused in general over the role of conjunction in syntactic analysis. He says

Such arguments based on conjunction have been considered strong evidence for structure throughout the history of transformational grammar. We have seen, however, that this type of argument leads ultimately to conclusions which contradict the most fundamental and best accepted notions of phrase structure.

Really? "Such arguments"? No! Only one argument from conjunction has been appealed to "throughout the history of transformational grammar", and this is NOT the "symmetric predicate" argument attacked (and then used) by Jackendoff; it involves rather the conjunction criterion discussed at length above. Jackendoff's conception of phrase structure does not challenge THIS form of argument at all--on the contrary, I know of no better measure of phrase structure to this day, and just where Jackendoff's general theory is plainly correct, the conjunction criterion backs him up.

"Such arguments" is evidently supposed to mean the argument from "symmetric predicates", but this has NOT been used "throughout the history" of generative studies. Its possible relevance was certainly not considered in LSLT or SS, and, as far as I know, was not suspected until C. Smith's mid-sixties paper, also mentioned above. It was precisely the relatively problematic nature of this more recent argument that impelled us to isolate the "plural selectors in the pure state" before.

Thus whatever the value of Jackendoff's theory of relatives, I see no merit in defending it on the plea that the theory of conjunction has "always been a troubled area in transformational grammar". For what aspect of natural language has not? "Gapping is ever problematical," indeed! But is Passive any less so, or, for that matter, Relativization?

Now we turn to the respectively argument against the Corner Conjunction Restriction. The simplest kind of case to consider is the following:

- (30) (*)Daphne and de la Trachette are peeling
the oranges and will boil the potatoes,
respectively

(31) (*)John boiled and fried the eggs and the
bacon, respectively

Now Dougherty (1968, VI.3 (5), p. 250) reports the latter example as grammatical, and there is a widespread (though not unanimous) tendency among linguists I have consulted to sustain his judgment, and also to accept the other example. But my better lay informants mostly reject such forms out of hand. Though of course perfectly intelligible, they simply "sound bad" alongside of typical respectively cases like John and Mary fried the bacon and eggs, respectively, which, to be sure, are awkward enough to start with.

I am therefore amply disposed to call strings like (30) and (31) "derivatively generated". If this is right, then of course they show nothing directly about the grammar. On the other hand, if they are grammatical, then surely such forms as (32) are, too:

(32) (*)John and Mary were hunting lions and
were frightened by snakes, respectively

Consistently enough, Dougherty (1970 S13 (364), p. 837) also accepts this as grammatical. But on his own admission, the VP's cannot be conjoined in the base here, by the argument from Passive discussed above. He is thus

forced to posit an entirely ad hoc transformation for just these cases, deriving them from structures like

(33) *[[John and [_{NP}]]] [[was hunting lions]
and [_{VP}]]] respectively and snakes frightened
Mary

by first applying Passive in the second clause and then substituting Mary for the dummy NP and was frightened by snakes for the dummy VP. But of course we cannot possibly admit such a complex transformation for such a narrow range of "exotic" forms. If they are grammatical at all they must surely be generated by the apparatus already needed for tamer respectively data.

The only way I can figure to meet this requirement while maintaining the grammaticality of (30) and (31) is to give up the assumption that the distribution of respectively is stated exclusively over deep structure. But this is an indispensable premise in the argument from these examples to base V(P) conjunction, the implication being that some rule must see the paired coordinate constituents "already there" to sanction the deep structure (although Dougherty is not explicit about this). If we drop this supposition the argument evaporates, so

that even if grammatical, ((33)-(32)) show nothing against the Corner Conjunction Restriction.

The reader may have noticed that we have moved to take up again the historical vantage point of the first paragraph of this section, after having left it to pursue a purely theoretical exposition. Before concluding this section, then, we take the opportunity to mention a few issues which, though important in the historical development of conjunction theory, appear tangential in hindsight, and were therefore glossed over in our preceding discussion.

As we noted before, the first effort to show base generation of $NJ \text{ and } NJ$ seems to have been that of C. Smith, *op. cit.* Actually, Smith held that this form was generated both in the base and by transformation. Her argument was that a theory deriving $NJ \text{ and } NJ$ exclusively by transformation, like LSLT or SS, "is deficient, since it gives no formal basis for the systematic ambiguity discussed below," namely, that between the distributive and non-distributive readings of VP predicates with $NJ \text{ and } NJ$ subjects.

The term "distributive" is taken from traditional

grammar to designate the reading in which the predicate is understood as applying separately to each conjunct--roughly, where $N1 \wedge_i N1^V$ may be paraphrased as $N1^V \wedge_i N1^V$; this represents Smith's transformational conjunction. The other reading, where the reading is applied as it were "collectively" to the whole subject, does not have the indicated paraphrase. It is characteristic of our "plural selectors"; this is what led Smith to offer what Lakoff and Peters called "symmetric predicates" as evidence of base N1 conjunction.

Smith's argument turns out to be flawed, even though it introduced the data base for later, sound, arguments for her conclusion that base N1 conjunction exists. The reason her own analysis fails is that the relevant ambiguity shows up in Sentences whose plural subjects contain no conjunction at all. *They came* may mean either *they each came* or *they came together*; this is a structural ambiguity, since we don't get **they each came together*. But if some mechanism over and above the theory of conjunction must provide for ambiguity as to distributive sense here, then there is no direct argument from this ambiguity for a split treatment of N1 conjunction.

The only obvious proposal to save Smith's analysis in the face of this objection does not work--namely, deriving all plurals from NJ-conjunction. This was suggested soon after the problem was noticed. We refer the reader to Chomsky (1970, p. 76ff), McCawley (1968), and Dougherty (op. cit.) for full discussion. Suffice it to say that, besides using "referential indices" in a way illegitimate in a reasonable metatheory (cf. Chomsky (1970), especially n. 11) this approach cannot account for cases in which the sense of the NJ to be derived is a concept of infinite extension.

Thus we have to find a more general analysis of Smith's ambiguity. But it is worth remembering that the failure of her theory does not diminish the importance of her observation that Plural selectors provide clear cases of base and-NJ conjunction. Now the most reasonable substitute for Smith's approach in the literature is essentially that of Dougherty (op. cit.): in our terms, assume general and-NJ conjunction in the base, and assign Sentences a feature dictating for the presence or absence of distributive sense. A positive value for this feature is forced by the appearance of elements like each, and a negative value by the choice of a Plural selector in the

head of the Sentence.

I qualified the last reference to Dougherty with "essentially" because he admits more base conjunction than we, as noted earlier; moreover, his feature proposal is embedded within a more general theory of selection which we may elsewhere need to alter--see George and Hoffman (in preparation).

One final historical note. Until now we have ignored the rule of Conjunct Movement offered by Lakoff and Peters (op. cit.) in their "Phrasal Conjunction and Symmetric Predicates"; see Dougherty (op. cit.) for critical discussion. As far as I can tell, the issue of the existence and nature of this rule has no bearing on the scope of "phrasal conjunction", so we will not go into it here.

5.2 Grammatical Ellipsis

In the preceding section S5.1, we sketched a history of the analysis of conjunction in generative grammar, giving special attention to the arguments that motivated the introduction of coordination in the base. Although we provisionally accepted Chomsky's (1965) proposal to base generate all sentential (V1) conjunction, together with an

amended version of Smith's (1969) argument for and-NJ conjunction in deep structure, we questioned the familiar inference that this decision justifies providing the base rule schema $X \rightarrow X^{\text{Conj}} X$ for general deep structure coordination of all major categories.

Not only did we find the positive evidence for direct recursion in the base limited to the two original cases; we also noted that Dougherty's observations about VP coordination with a Passive member show that at least some cases of direct conjunction must be generated by the transformations motivated independently by coordination across constituent boundaries. Moreover, we may even have evidence here against the general base schema, insofar as we want to avoid assigning structural ambiguity to examples with active coordinate VP's.

Further support for our suggestion came from Koutsoudas' analysis of (n)or-NJ conjunction as exclusively derived from VJ conjunction by ellipsis, insofar as this is justified by his claims about Subject-Verb Agreement.

We thus reject the base rule schema for conjunction in favor of two non-canonical phrase structure rules:

(34) $VJ \rightarrow VJ^{\wedge} \text{ and } ^{\wedge}VJ$
 (n)or

(35) $NJ \rightarrow NJ^{\wedge} \text{ and } ^{\wedge}NJ$

These rules generate a narrow range of cases of direct conjunction. All other coordination, whether direct or not, is derived from these by ellipsis. As we proceed to characterize the needed transformations, we will run into new evidence for our limitation of base conjunction, which are predicted to be grammatical by the base schema theory, but ruled out by independently motivated properties of our more conservative theory.

The general picture, then, supports our claim of S2 that direct recursion is limited to corner conjunction. In fact, if the rules given above do cover the whole range of direct recursion in the base, an even narrower restriction holds. We are thus proposing to "simplify" (constrain the descriptive power of) the base component. As usual, this possibility argues for transformational analysis only in tandem with an explicit presentation of a system of transformations provided by the metatheory to reconstruct our intuition in a more revealing way than a pure base theory.

So how are ellipsis rules written, and how do they work? One strategy for answering these questions would be to consider the extensive literature on the controversy over the structural change effected by ellipsis rules: against the classical appeal to the deletion elementary, we have the idea current in EST circles that the supposedly deleted material is in fact a string of "empty categories" that an "interpretive" rule (binding transformation) links to a "controller" in the other conjunct.

Unfortunately, I cannot go into this question in detail; in any case, as far as I know, the discussion in the literature has been inconclusive. For the time being, then, I will focus attention on the restricting class of the transformations, since I think that the most interesting claims about the scope of these rules and about the nature of parallelism stand regardless of the choice of elementaries. In the final analysis, of course, the two issues must be decided by a single general theory, so from time to time it will prove convenient to suppose a definite choice of elementaries in expounding our arguments. Thus I will henceforth assume the deletion solution. My only reason for this somewhat arbitrary

decision is that I believe that the classical account works more smoothly with an adequate theory of selection; this is because the interpretive theory must require matching lexical items to agree in "contextual features" (cf. Aspects), which George and Hoffman (1979) argue are invisible to all but lexical insertion rules. I want to emphasize, however, that the arguments concerning parallelism and the limitation of base conjunction would translate straightforwardly into the analogous "empty category" theory.

With this, we turn to the question of the restricting classes of the ellipsis transformations. We begin with Left Peripheral Ellipsis, since it is both the easiest case to motivate (because of the complete acceptability of its outputs) and the most important (by virtue of its interaction with the rest of the theory--especially the Corner Conjunction Restriction and its ramifications). Here are some typical examples:

(36) The bouncer took Daphne to the bar and
de la Vain to the beach

(37) Daphne fell into the hands of evildoers and
was rescued by the handsome bouncer

(38) Daphne assaulted de la Vain and
kissed the bouncer

(39) Daphne's arms and face were severely scratched
 Recall that a transformational rule is needed anyhow for the first case, even in a theory generating all direct conjunction in the base. But such a theory, as we have noted, is untenable, because of the second case seen in the light of the transformational analysis of Passives. The fundamental idea of our analysis is that the simplest formulation of the rule needed for the first construction will automatically generate the remaining ones, a fact which renders superfluous the base conjunction of non-corners.

And what is the simplest formulation of this rule? Given our interpretation of the A/A Condition (S4), we must write

(40) str Conj cat cat str

(3-->0)

For if we tried to get along with only one constituent variable cat in the restricting class, every proper analysis would take the whole second conjunct as the corresponding factor. But then, if we treated this factor as the delendum, the output of the rule would always be missing the right conjunct; similarly, if we took cat as

the context, nothing could be deleted. Either way, none of the cases motivating the rule would be derived. Hence, by the "logic of markedness", we pass to the more complex formulation (40).

Since the delendum must now be a constituent, we must apply the rule "iteratively" in order to create structures like that of (35), where the totality of the missing material is not a constituent.

Now consider that our rule analyzes the left conjunct under a string variable--it does not constrain the controller of the deletion, so to speak. This is because we do not want to complicate the rule to guarantee "recoverability". Rather, we must rely on the metatheory for this. At issue are such facts as the following. First, the rule cannot derive

(41) *Generosity frightens Daphne and loves
de la Vain

from

(42) Generosity frightens Daphne and the bouncer
loves de la Vain

This presumably follows from the well known constraint that rules deleting anything but a specified terminal

string must find in the context a string "structurally identical to" or "non-distinct from" the delendum; see Lees (1962) Aspects. As usual, however, the "controller" can't be just anywhere in the context. In the case of Left Peripheral Ellipsis, it must occupy the same structural position in the left conjunct that the delendum does in the right. Thus we don't get

- (43) *His diary proves that the bouncer dates Daphne,
and has, of course, visited de la Vain
evidently

from

- (44) His diary proves that the bouncer dates Daphne,
and he (the bouncer) has, of course, visited
de la Vain evidently

The deletion fails because the target is the main clause Subject in the right hand conjunct, corresponding to the subordinate clause Subject in the left. (I included ~~(evidently, of course)~~ to eliminate an irrelevant reading, in which the material after and is subordinate to prove; cf. *His diary proves that he has, ~~(evidently, of course)~~ visited de la Vain. But the intrusion of ~~(evidently, of course)~~ is not what blocks the intended derivation, since we have The bouncer dates Daphne and has, ~~(evidently, of course,)~~ visited de la Vain.

This is typical of a wide range of structures where ellipsis fails because it requires the conjuncts to be "parallel" to a certain degree in order to apply. The general observation was known to traditional grammar, and has been presupposed in all generative studies of conjunction. But as far as I know, the operative concept has never been carefully defined, and as promised in S1, an attempt at this will be the central technical contribution of this dissertation.

Consider now that the constraint extends beyond the relation between the controller and the delendum; the material to remain in the right conjunct must also be parallel in some sense to the corresponding parts of the left. For example, the following are bad, despite the fact that the needed ellipsis satisfies the recoverability requirement as usually understood:

(45) *The bouncer was muscular and a guitarist

(46) *The bouncer took Daphne to a bar and de la Vain

(47) *He kissed Daphne tenderly and de la Vain
yeerday

Now if we put, say, to the beach at the end of (46), we get the good sentence (36) already noted; similarly, if we substitute passionately for yesterday in (47), or last

week for tenderly, the star again vanishes. Thus it seems that ellipsis, in order to be recoverable, must match the "remnants" (term due to Sag (1975)) as well as the delendum with antecedents in the first conjunct. Curiously, however, this further matching, unlike the delendum-controller link, is not contingent upon the internal structure of the correspondents. Intuitively, then, we can say that the recoverability constraint projects the requirement of conjunct agreement (S4) "down from the top" of the coordinate structure onto "corresponding" constituents of the two conjuncts, till it reaches the remnants of the ellipsis on the reduced side.

As already noted, data motivating this requirement were observed long ago, though no one has as yet offered a precise statement of the requirement itself. Obviously, we need to say that certain constituents in the right conjunct must have the same categorial labelling as the "corresponding" constituents of the other; this was seen as early as LSLT (S115.1, p. 557ff). In formalizing the condition, we must say (i) which constituents count and (ii) what "correspond" means here.

We can view the problem as one of constructing a

mapping from a certain set of constituents of the right conjunct into the set of constituents of the left. The requirement that corresponding constituents have the same labelling then appears as a condition that our mapping preserve constituency relations. Now the question is, what is the domain of the mapping?

The only serious effort that I know of to face this issue is the informal statement of Tai (1969, p. 73ff):

Identity Deletion can apply to a pair of conjuncts with identical constituents, if and only if the highest constituents preceding or following the highest identical constituents in one conjunct have the same labelling as the corresponding constituents in the other constituent.

Though not phrased in terms of our mapping, this is plainly on the right track, as far as it goes, from the point of view of our guiding intuition that parallelism projects conjunct agreement "down into" a coordinate structure. Tai does not, however, number constituents containing remnants or delenda among those that have to match. For simplicity's sake, and for empirical reasons to appear later, we will change this in our formalization. Moreover, Tai's statement presupposes as separate determination of identity for delenda, which is used to fix the domain of the correspondence. This we will drop

in favor of a uniform "recoverability" interpretation, keyed to the categorial terms of the proper analysis.

More formally now, we say that Left Peripheral Ellipsis applies only on a proper analysis that is congruous, where we define this concept as follows:

- (48) Let F be a proper analysis of Z w.r.t. K , Q
 (K being a phrase marker of Z , and Q the
 restricting class of an ellipsis
 transformation.)
 Let
 $D = \{\text{constituents of the right conjunct that}$
 $\quad \text{the categorial terms of } F \text{ are in}$
 $\quad \text{construction with}\}$
 and
 $C = \{\text{constituents of the other conjunct}\}$
 Then F is congruous iff there is a function
 $\delta: D \rightarrow C$ such that, for all $A, B \in D$
 (i) if A is properly included in B , then
 $\delta(A)$ is properly included in $\delta(B)$
 and
 (ii) if P is a category such that $E(A, P, Z, K)$
 then $E(\delta(A), P, Z, K)$.

This definition, to be fully serviceable, must be generalized to ellipsis in the left conjunct and in non-coordinate constructions, like certain comparatives; the reformulation needed is straightforward if somewhat hairy, so we omit it here. Assuming this extension, we can write the

Format((49) Law of Congruity)

Ellipsis transformations apply just on congruous proper analyses.

Any model of transformational grammar consistent with our minimal assumptions (the Corner Conjunction Restriction; our treatment of conjunct agreement as extended feature percolation; and our "logic of markedness" interpretation of Minimal Factorization, A/A, and allied Conditions) will, under the Law of Congruity, incorporate our guiding intuition of elliptical parallelism. In particular, ((45)-(47)) above, and comparable misapplications of Gapping, Right Peripheral Ellipsis, and Comparative Subdeletion, are eliminated by this principle.

To see how the formalism works, consider the derivation of (36) from the intermediate stage The bouncer took Daphne to the bar and took de la Vain to the beach. The needed proper analysis is as:

(50) ...bar--and--took--de[^]la[^]Vain--to[^]the[^]beach

str--Conj--cat-- cat -- str

The function we must find has the domain $D = \{\text{took}, \text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}, \text{to}^{\wedge}\text{the}^{\wedge}\text{beach}, \text{took}^{\wedge}\text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}^{\wedge}\text{to}^{\wedge}\text{the}^{\wedge}\text{beach}$ (and $\text{took}^{\wedge}\text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}$, if this is a constituent)); the range C consists of all constituents of The bouncer took Daphne to the bar. Now consider the function $\phi:D \rightarrow C$ such

that

(51) $\delta(\text{took}) = \text{took}$

$\delta(\text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}) = \text{Daphne}$

$\delta(\text{to}^{\wedge}\text{the}^{\wedge}\text{beach}) = \text{to}^{\wedge}\text{the}^{\wedge}\text{bar}$

$\delta(\text{took}^{\wedge}\text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}^{\wedge}\text{to}^{\wedge}\text{the}^{\wedge}\text{beach}) =$
 $\text{took}^{\wedge}\text{Daphne}^{\wedge}\text{to}^{\wedge}\text{the}^{\wedge}\text{bar}$

and, if $\text{took}^{\wedge}\text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}$ is a constituent here

$\delta(\text{took}^{\wedge}\text{de}^{\wedge}\text{la}^{\wedge}\text{Vain}) = \text{took}^{\wedge}\text{Daphne}$

By inspection, this δ preserves inclusion and E relations as required, so that the indicated proper analysis is congruous. Hence, deletion of the second took produces a grammatical string.

Contrast this case with the bad ellipsis leading to (45). Assuming as above that The bouncer was muscular and was a guitarist is grammatical, we now ask why we may not use Left Peripheral Ellipsis to delete the delete the second was. The answer is that the needed proper analysis would appear as

(52) ...muscular -- and -- was -- a guitarist -- e

str --Conj -- cat -- cat --str

so that we would need to construct $\delta:D \rightarrow C$, where D would have to contain $\text{was}^{\wedge}\text{a}^{\wedge}\text{guitarist}$, $\text{a}^{\wedge}\text{guitarist}$, and the

second was. Now to preserve the constituency relations we must put $\delta(\text{was}) = \text{was}$, and $\delta(\text{was}^{\text{a}^{\text{guitarist}}}) = \text{was}^{\text{muscular}}$. But since a^{guitarist} is properly included in was^{a^{guitarist}}, $\delta(\text{a}^{\text{guitarist}})$ must be properly within $\delta(\text{was}^{\text{a}^{\text{guitarist}}})$ --namely, $\delta(\text{a}^{\text{guitarist}})$ must be either was or muscular. Now since $E(\text{a}^{\text{guitarist}}, N, Z, K)$, we must have $E(\delta(\text{a}^{\text{guitarist}}), N, Z, K)$; but neither was nor muscular is introduced under a subcategory of N , so no δ has the required properties. Hence, by the Law of Congruity, the intended ellipsis fails.

This approach seems adequate for a broad range of cases of ellipsis, both Left Peripheral and otherwise. Its most remarkable virtue, from the viewpoint of our earlier discussion, is that it makes possible a natural statement of the classical generalization "...X...-->...X^{and}X, where X is a prime" in a theory that separates direct conjunction into a base and a transformational component. Furthermore, it does so by utilizing a principle motivated independently, in the theory of indirect conjunction; this invalidates a potential argument for the general base schema that we rejected under the Corner Conjunction Restriction. We will go on to find cases that, like our earlier

consideration of (p)qr and Passive VP conjunction, seem to favor Corner Conjunction positively. We will also see the theory profitably applied to difficulties in the analysis of "Across the Board" extractions and Comparative Subdeletion.

Before we leave the task of defining "congruity", however, we must note two problems that could cause us to modify the formalism already offered.

First is the fact that in cases where the construction of a function δ under the definition (48) leads to a grammatical ellipsis, the resulting δ always appears (in English, at least) to preserve "linear order" as well as the inclusion and E relations. As far as I know, this follows from known properties of English Phrase Structure, which prevent any relevant δ from meeting the conditions of the definition (48) without also preserving linear order. The question is, of course, whether in English or in any other language, an ellipsis could be sanctioned under our reading of the Law of Congruity but fail because the needed δ broke linear order. If so, we'd have to add a third homomorphism restriction to the choice of the congruity function.

The second problematic aspect of our construction is that it admits a certain range of cases solely because of our assumption (of S4) that the constituency relation \bar{E} is not sensitive to the rank of categories, as illustrated by

(53) Daphne kissed the bouncer and (she) fled

The function needed to guarantee congruity here must take fled into kissed^the^bouncer; but the usual theory of constituent structure makes fled here both a Verb and a Verb Phrase, while it makes $\bar{\Phi}(\text{fled}) = \text{kissed}^{\text{the}^{\text{bouncer}}}$, a Verb Phrase only. Hence, if we refine \bar{E} to distinguish ranks, our interpretation of the Law of Congruity will prohibit the ellipsis in (53), unless we modify the standard bracketing of VP.

Of course, this line of reasoning is not too alarming, since we should in any case resist the crucial refinement of \bar{E} on grounds of restrictiveness (S4). In all fairness, however, we must note the following indication that such a refinement may be called for: while Verbs seem to map freely to Verb Phrases under $\bar{\Phi}$, as in (53), N is not carried over so comfortably to N^1 . For example, we have ellipsis in:

- (54) Daphne's indolence and (her) cruelty to her
subjects exceeded Nero's

But we do not have, reversing the order of the conjuncts

- (55) *Daphne's cruelty to her subjects and indolence
exceeded Nero's

alongside of the grammatical

- (56) Daphne's cruelty to her subject and her indolence
exceeded Nero's

This pattern is predicted by the Law of Congruity only if we make E sensitive to rank; otherwise, the elliptical versions of ((54)-(56)) should be acceptable regardless of the order of conjuncts within the coordinate Noun Phrase. Since we are not prepared to accept the indicated re-definition of E, we must assume that there is some unknown explanation for the bad flavor of (55). Chomsky (personal communication) has suggested that the relative heaviness of the conjuncts plays a role, as in our remarks on Right Peripheral Ellipsis in §5.1. This seems plausible, but it remains to explain why the VP reduction analogous to (55) is not ruled out on the same grounds.

Nevertheless, we henceforth rely on the coarser construction of E in our interpretation of the Law of Congruity, disregarding the problems just noted. With

this, we return to the Corner Conjunction Restriction vs. the General Base Schema. We have drawn nearer to an explicit statement of a transformational analysis whose formal advantages will justify the proposed "simplification of the base". To complete the transformational theory, we must construct a family of rules comparable to the Left Peripheral Ellipsis already formulated.

Of these, the only one directly relevant to the issue of base conjunction is Right Peripheral Ellipsis. This is in fact the mirror image of the previous transformation:

(57) str cat cat Conj str

(3--->Ø)

Again by the "logic of markedness", the A/A Condition forces us to use Cat as a context predicate. Also as before, the formulation is motivated by cases of coordination crossing constituent boundaries, but extends to some cases of "direct conjunction" under our limitation on the base:

(58) \$The bouncer rescued, and de la Vain consoled,
the Siamese kitten that Daphne had
treed

(59) The bouncer rescued and consoled the kitten

On our theory, the $V^1 \text{ and } V^2$ construction of (59) must derive from VJ conjunction, i.e. from The bouncer rescued it (the kitten), and he (the bouncer) consoled the kitten. Left Peripheral Ellipsis deletes the Subject of consoled, and Right Peripheral Ellipsis deletes the Object of rescued.

It is here that we come upon the independent evidence promised earlier against the General Base Schema for conjunction. Since $V \rightarrow V^1 \text{ and } V^2$ is an instance of the Schema, a grammar incorporating it should freely conjoin Verbs in the base. In fact, however, the $V^1 \text{ and } V^2$ construction is fully acceptable only when no more than one VP constituent follows it. Thus we do not have (60) alongside of (59):

(60) *The bouncer takes and mails toys to Daphne's kitten

The base theory has no natural means of ruling out forms such as this. The ellipsis theory with the Corner Conjunction Restriction, in contrast, explains the data ohne weiteres. For, to generate (60), Right Peripheral Ellipsis would have to apply twice. Starting with The bouncer takes toys to Daphne's kitten and (he) mails toys to Daphne's kitten, it would first eliminate the Indirect

Object of takes to give The bouncer takes toys and mails toys to Daphne's kitten; then it would reapply to eliminate the Direct Object of take, producing (60). But there are grounds independent of the $V \overset{\emptyset}{\text{and}} V \overset{\emptyset}$ form for prohibiting such multiple application of Right Peripheral Ellipsis. For example, we do not get

(61) *The bouncer took, and de la Vain mailed,
toys to Daphne's kitten

Now it is an interesting question why Right Peripheral Ellipsis is so constrained; given our treatment of Left Peripheral Ellipsis, the solution cannot be to prevent Ellipsis rules in general from applying iteratively. The key is probably in the special intonation pattern associated with Right Peripheral Ellipsis outputs. Whatever the reason, however, if we can say that Right Peripheral Ellipsis never applies twice in a single construction, then our theory will avoid generating (60), which the base schema creates directly.

There are a couple of considerations that appear at first glance to argue in the opposite direction, that is, for the General Base Schema, but turn out on closer inspection to be neutral.

The first is the fact that we have to rule out an elliptical derivation of the and-NJ construction. Otherwise, given our assumption that Tense is in VP, we will generate

(62) *Daphne and de la Vain is here
 from Daphne is here and de la Vain is here. One might hope to avoid the problem by making Tense a sister of the Subject and appealing to our prohibition of iterative application of Right Peripheral Ellipsis. This will not work, however, if we are right in following Koutsoudas on the issue of (n)or-NJ coordination. In any case, we can rule out (62) by the principle of Filtering by Analogy (S6), which would block the elliptical derivation because it is strictly more complex than the "competing" base analysis.

More serious, then, is the second problem of accounting for the distribution of the Coordinate Quantifiers both and (n)either. Dougherty (1968, 1970) proposed that the base schema for conjunction introduce these directly. The schema we have so far considered ($X \rightarrow X^{\wedge} \text{Conj}^{\wedge} X$) can be modified quite naturally to do so: $X \rightarrow (\text{Quantifier})^{\wedge} X^{\wedge} \text{Conj}^{\wedge} X$; such an account would appeal

to feature percolation to insure the matching of the Quantifiers with the right conjunctions: both with and, and (n)either with (n)or. (See Dougherty loc. cit.)

Of course, we could modify our Corner Conjunction rules analogously. But this would not immediately account for examples like

(63) Daphne both delighted the bouncer and
treed the kitten

Our theory must treat this as elliptical, but the source is a little weird:

(64) ?Daphne both delighted the bouncer and
she treed the kitten

Furthermore, this both could not be placed by our base conjunction rules. One might hope to utilize here the rules producing so-called "Quantifier Float" structures, like

(65) The bouncers both delighted Daphne

Some care would be needed here, however, since some of the potential sources are ungrammatical:

(66) *Both Daphne delighted the bouncer and
she treed the kitten

Of course, this latter string is a problem for the base

schema, too. Compare the good

- (67) Either Daphne delighted the bouncer,
or she treed the kitten

There is some indication, in fact, that working out the Quantifier Shift approach could provide a positive argument against the base schema theory--namely, bad cases that the modified base schema would generate automatically but a Corner Conjunction theory wouldn't necessarily derive at all, like

- (68) *Daphne's either dog or kitten fled the kitchen

- (69) *Daphne is quite both outrageous and kind

I am not now prepared to offer a specific theory of the Coordinating Quantifiers, but the overall pattern certainly does not on the face of it favor the General Base Schema.

Before we end the discussion of Right Peripheral Ellipsis, we examine its interaction with the Law of Congruity. Consider

- (70) *Daphne wants the bouncer to rescue, and de la
Vain loathes, the Siamese kitten that she treed

This would come from a non-elliptical sentence like Daphne wants the bouncer to rescue it (the Siamese kitten...) and

de la Vain loathes the Siamese kitten that she treed. The required proper analysis would be as

(71) ...to--rescue--it--and--de la...

str-- cat --cat--Conj--str

Now D must contain all constituents of the first conjunct that properly include the Object of rescue, and there are obviously at least four of them. The corresponding Object of loathes, however, is properly included in only two constituents of the second conjunct. Hence, no δ can preserve proper inclusion as required. The proper analysis is thus not congruous, so the ellipsis fails.

Observe that this kind of example justifies in part our modification of Tai's idea of parallelism. A further interesting property of our analysis is that it correctly predicts the grammaticality of the Sentence obtained by switching the conjuncts of (70):

(71) De la Vain loathes, and Daphne wants the bouncer
to rescue, the Siamese kitten that she treed

Here we have $D = \{\text{loathes, it, loathes it, De la Vain, De la Vain loathes it}\}$. Given this domain, one congruity function δ with the needed properties puts

(72) $\delta(\text{loathes}) = \text{wants}$

$\delta(it)$ = the Siamese kitten...treed

$\delta(loathes\ it)$ = wants the bouncer to rescue...
treed

$\delta(De\ la\ Vain)$ = Daphne

$\delta(De\ la\ Vain\ loathes\ it)$ = Daphne wants...treed

The existence of this δ makes the indicated proper analysis congruous, and the ellipsis good. Thus congruity involves a certain asymmetry not previously noted in discussions of coordinate parallelism.

With this we end our efforts to justify directly the Corner Conjunction Restriction and the associated ellipsis rules, assuming henceforth that cases of direct conjunction not appearing to fall under the Corner Conjunction rules of the base are actually generated by Left and Right Peripheral Ellipsis applying to true Corner Conjunction. Thus in one sense we move back to a theory of coordination more traditional than what is now in vogue. On the other hand, we have seen no evidence for the "Raising" and "Pruning" operations traditionally presupposed in the "Conjunction Reduction" literature (Ross, 1967, and much allied work)--that is, we have not been led to assume "derived conjunction" as the output of our ellipsis rules. Thus, transformationally derived

"direct conjunction" of VP's, for example, may well appear only in the configuration $[[\text{NP}^{\text{S}} \text{VP}]^{\text{S}} \text{and} [\text{VP}]^{\text{S}}]$ where the VP's in question are crucially not sisters.

Now the serious arguments for Raising-Pruning in the standard theory were only two, and these both pale in the light of well known meta-theoretical counter-indications. The first "derived conjunction" argument is from intonation, which is anyway a notoriously bad test of constituent structure; the second, from "across the board" violations of the Coordinate Structure Constraint, is disposed of in the next section (§5.3).

Observe in this connection that the "ever problematical" rule of Gapping, which we can't formulate here (!), is obviously an ellipsis rule in our sense, being subject to the same parallelism constraint as the other coordination transformations--cf. Tai (1969), Hankamer (1971), Sag (1976), and Koster (1978). But no one has proposed that Gapping creates derived conjunction, which shows that this is at least not a necessary property of ellipsis rules. Thus, by Occam's Razor, we conclude that it is never a property of them--see also George and Kornfilt (1980).

This is why in S5.1 we left open the possibility, which we now take to be realized, of direct conjunction "where the sequence 'X^{and}X' does not form a constituent of the type X." For example, in the case cited just above, the sequence "VP^{and}VP" is not a VP, on our theory; in fact, it is not a constituent at all.

5.3 Wh-Movement "Across the Board"

Consider now that Ross's (1967, 4.89) Coordinate Structure Constraint

In a coordinate structure, no conjunct may be moved, nor may any element contained in a conjunct be moved out of that conjunct.

follows in effect from the Law of Inclusion (A/A) and our interpretation of Subjacency, under the Corner Conjunction Restriction. But Ross (op. cit., 4.2.4) observes sentences like the following

- (73) Students who fail the final exam or
do not do the reading will be executed
[his (4.1.22)]

offering to qualify the Coordinate Structure Constraint by adding the following exception condition

- (74) unless the same element is moved
out of all the conjuncts.

Thus he seems to assume that each conjunct of the Relative

clause in (73) starts out with a Relative Pronoun that is extracted from it by the Wh-Movement rule; apparently, the transformation is supposed to fuse the two Pronouns outside of the whole coordinate structure.

Ross does not attempt to formalize the complex operation supposedly effected by Wh-Movement; still less does he provide any evidence for his presupposition that the Relative Pronoun is "moved out of [both] conjuncts" in (73). But I now claim that the Coordinate Structure Constraint holds in its unconditional form, and that there is no such thing as "movement across the board"; that the examples Ross brought under this rubric are actually generated by simple Wh-Movement within each conjunct, as in

(75) Students who fail the final exam or
 who do not do the reading will be executed.

plus Left Peripheral Ellipsis of the second Relative Pronoun. The same kind of option is available in other WH Movement constructions: Questions, Free Relatives, and certain Comparatives (§5.4 *infra*).

In support of our analysis, we observe first that the independently motivated constructions of our theory of

ellipsis support our proposed derivations, so that Ross's is at best redundant. But given that his view requires us to complicate both the Coordinate Structure Constraint and WH Movement, it seems not only superfluous but harmful.

Our theory has another, more concrete, advantage, illustrated by some contrasts like

(76) They removed the prisoner, who(m) the
the judge has sentenced and (who(m))
the warden will execute.

(77) They removed the prisoner, who has lost
his appeal and *(who(m)) the warden
will execute.

(78) They removed the prisoner, who(m) the
judge has sentenced and *(who) will now
appeal.

The generalization seems to be roughly that we get "across the board" wh-Movement phenomena only when the relevant WH's come from "parallel positions" in their conjuncts.

There are two possible explanations for this observation within our general framework, both assimilating the "across the board" parallelism constraint to the Law of Congruity (§5.2), where Ross, in contrast, can offer only an *ad hoc* account, as in Williams (1978). Perhaps the more obvious of our alternatives is to claim that the Nominative and Objective Cases are so

distinguished in English that we can appeal to a Case conflict between the "antecedent" and "anaphor" Relative Pronouns to explain why the deletions in (77) and (78) are bad.

The other possibility, more promising in my view, is suggested by Chomsky's (1973, p. 254) proposal that the theory of cyclic transformations

be so constrained as to forbid operations that never change the terminal string of a phrase marker, but only its structure, as in the original formulations of Subject Raising to Object position,

so that rules that result exclusively in restructuring, never changing the terminal string, should be confined "to the readjustment rule component of the grammar..." (*ibid.*). We call this the Weak Vacuous Application Prohibition; if we strengthen it so as to prevent cyclic transformations from ever applying vacuously, departing now from Chomsky's formulation, then we can explain the "across the board" parallelism requirement without appeal to "abstract Case", as follows. The WH Objects in (77) and (78) must be fronted to COMP by WH-Movement, while the WH Subjects cannot be, by the Strong Vacuous Application Prohibition. Hence the Pronouns are not in "corresponding positions" for the Law of Congruity, so the ellipsis

fails.

The latter approach is superior in several respects. First, it preserves our hypothesis that syntactic Case is normally lost when it is morphologically neutralized (§2.2 supra). Secondly, it leads to correct predictions about certain more obscure "across the board" examples, such as

(79) Here is the prisoner who ratted on the
punk and *(who) Foley said___was torched.

(80) Here is the prisoner who(m) you saved
and (who) Foley said should be torched.

In the first of these, both Pronouns are Nominative in abstract Case, yet the deletion is bad; in the second, there is an abstract Case conflict, but the ellipsis is (relatively) good. Chomsky (personal communication) reports a proposal of Sjoblom to the effect that the second Pronoun in (80) is adjusted to Objective Case by a rule giving the hypercorrected (*) Here is the prisoner whom Foley said should be torched. but notes that this would not help where said is replaced with got the idea and similar sequences that should not "assign Case", even in the abstract sense. Also, Sjoblom's suggestion would not help on (77) and (78).

Now as far as our generalization of Chomsky's Vacuous

Application Prohibition, we should note that Ross (1967, 4.2.4.2) anticipated it, though he specifically exempted Wh-Movement from it. In any case, the Vacuous Application Prohibition, besides its value in the "Raising" controversy, may be grounded theoretically in the construction of "derived interpretations", or in the principle of Filtering by Analogy (S6 *infra*).

One more historical note before we end this discussion. Ross (*loc. cit.*) apparently sensed the connection between "across the board" movement and the transformations for coordination (his "Conjunction Reduction"), though I can't quite tell what he had in mind. Postal seems to have developed a clearer view of the same connection in his (1971-1972) lectures at MIT. This was first brought to my attention by Hankamer (personal communication), and it was verified for me in an unpublished letter from Postal (17 September 1979). The essence of Postal's proposal is that "across the board" extraction reduces to Conjunction Reduction, via Raising-Pruning, feeding ordinary WH Movement. That is, Conjunction Reduction is supposed to release the WH Pronoun from the Coordinate Structure Island.

Now Postal himself evidently no longer accepts this proposal, evidently because of some counter-arguments (now lost?) presented by A. Andrews during Postal's (1971-1972) lectures. Unfortunately, however, I have so far been unable to piece together Andrew's original reply to Postal. But there are several indications that Postal was right to forsake his Conjunction Reduction theory. Aside from the fact that Raising-Pruning is virtually unmotivated, there is the observation of Le Sourd (personal communication), who notes that

(31) *The warden told the snitch about, and
Foley questioned the screw about, the
enormous blue-white diamond in the lock-up

is considerably worse than the typical Right Peripheral Ellipsis

(32) \$The warden stole, and Foley fenced the
enormous blue-white diamond in the lock-up.

Le Sourd's point is that (31)* is the immediate source of

(83) The enormous blue-white diamond
that the warden told the snitch about
and Foley questioned the screw about...

, which, however, is much better than (31)*, while

(84) The enormous blue-white diamond
that the warden stole and Foley fenced...

is comparable to (32).

Similarly, we have Grosu's (1977a) example showing that Right Peripheral Ellipsis violates the Complex Noun Phrase Constraint:

- (85) Bob believes that it's possible that
his brother bought, and I maintain that
there is little doubt that his sister-
in-law received, an enormous
blue-white diamond [his (1)]

But if Right Peripheral Ellipsis ("Right Node Raising") could feed Wh-Movement à la Postal, we would also have the ungrammatical

- (86) *The enormous blue-white diamond that
it's possible that his his brother
bought and I maintain that there
is little doubt that his
sister-in-law received...

. Note finally that our reconstruction of "across the board" data in effect reverses the feeding order (cf. Kiparsky, 1968) proposed by Postal between "Conjunction Reduction" and Wh-Movement.

5.4 Comparative Subdeletion

Our efforts so far may be seen in part as a contribution to the theory of transformational deletions (resp. "empty node interpretation"). Specifically, we claim that the deletions of Sentence Grammar are confined to locally-governed (or free) erasures of "designated

elements" (minor Specifiers) and ellipses under the Law of Congruity. Thus we support Chomsky's (1977b) conjecture that there are no "unbounded deletions" in the sense of Bresnan (1976b) and much allied work. Such rules are in general to be reanalyzed (i) as outside of Sentence Grammar altogether (for example, "discourse" rules like VP deletion); or (ii) as the result of "unbounded" (i.e. iterated subjacent) movement followed by local deletion (for example certain Relatives and Comparatives); or (iii) as subject to the characteristic parallelisms constraint we have tried to define (§5.2, *supra*).

The English Comparative construction illustrates this strategy nicely. Those Comparatives that plainly involve deletion (not including the form *than-NI*, which might be base generated as is) fall into two general cases: ones subject to the Island conditions in the usual sense, which we will call "Free Comparatives", and ones subject to apparently stronger conditions, namely Bresnan's "Comparative Subdeletion" constructions. Besides "decaying faster with increasing complexity" than Free Comparatives, C-Subs seem to obey a special parallelism condition, as Chomsky points out (*op. cit.*, citing Bresnan on decay). We may add to Bresnan and Chomsky's

observations the following paradigm, which shows that the key notion of parallelism distinguishes Subjects and Objects in a fashion reminiscent of our "Across the Board" data:

- (87) Joan killed more Englishman than the Inquisition
burned (Frenchmen)
- (88) Joan killed more Englishman than (*Frenchmen)
fought the Inquisition
- (89) More Frenchmen revered Joan than (Englishmen)
admired Sir Thomas More
- (90) More Frenchmen revered Joan than Sir Thomas More
converted (*Englishmen)

Now Chomsky argues, convincingly in my opinion, that the whole pattern indicates a WH Movement plus local deletion analysis of Free Comparatives. As for C-Sub, he considers a similar analysis involving the "pure-term" interpretation of A/A, but he notes that it does not explain the special restrictions on this subconstruction. He also suggests that an alternative theory, splitting C-Sub off entirely from the general construction, may be preferable, but he does not formulate one.

We are now in a position to do so. We claim simply that C-Subs, unlike Free Comparatives, are generated by an ellipsis rule, subject to the Law of Congruity:

(91) str than str WH cat str
as

(4-->0)

(Inasmuch as the construction of the needed function $\bar{\theta}$ in these cases is straightforward, and may be done by inspection, we leave this as an exercise for the reader.) Continuing to follow Chomsky on Free Comparatives, we can with this explain at once the C-sub parallelism observations, with the bonus that we can avoid appeal to the mixed-term/pure-term distinction in restricting classes. The decay properties will also follow, either from Subjacency or Opacity, if one of these holds in general for ellipses.

6. Derivative Generation and Filtering by Analogy

Generative studies of natural language have always distinguished themselves by a unique treatment of linguistic intuitions. While the results of native speakers' introspection are crucial to the vitality of the science and accepted as such (page the Structuralist allegiance to the "spontaneous corpus" and to "behavioral correlates" of intuition), we reserve the right not to take an informant's judgment at face value:

The point of departure...will thus often be intuition. As motivation for the elaboration of linguistic theory, we will cite cases where previously constructed theory leads to counterintuitive results, and will try to develop a natural reformulation which, among other gains, will lead to a significant correspondence with intuition, i.e., will supply formal grounds for intuition. But it is important to keep clearly in mind that this does not mean that linguistic theory itself is based on intuition, that "intuition" and such notions appear in its basis of primitive terms. On the contrary, this basis is composed of the clearest and most objective notions we can find. Only in this way can linguistic theory serve the purpose of explicating our grammatical intuitions.

If a linguistic theory dictates counter-intuitive solutions for a language, we have two possible courses of action. We may disregard the intuition as fallacious (or as an intuition about something other than grammatical form), or we may re-construct the theory. Between these two poles of reliance on the results of a given theory and reliance on intuition, there are many possible positions and attitudes...

[LSLT S15.1-2, p. 123]

In particular, the theory has always allowed for ungrammatical structures that may be judged quite intelligible, comparatively acceptable, or even especially vivid just because of their deviance. Originally couched in terms of "partial grammaticalness", this idea has more recently been expressed in the distinction between "direct" and "derivative generation" (as of Aspects Chapter 4 n. 2, p. 227f). Thus we admit the description of certain forms as "(relatively) acceptable but ungrammatical" when this enhances the overall lucidity of our explanations.

If used to excess, of course, this ploy could trivialize the whole enterprise; we don't want to cry "derivative generation" every time our grammar fails to generate something that is in fact acceptable. Our first order of business in this section, then, is to propose some measures to prevent overuse of this theoretical dodge.

In the classical studies, claims of ungrammatical acceptability were mostly confined to violations of lexical selection rules (cf. Aspects Chapter 2 n. 11,

p. 212f), though the notion is implicit in scattered discussions of certain marginal forms in other domains (e.g. LSLT S61.1, p. 231f; S72.2 after statement 17, p. 281). The first full scale effort to apply the concept of derivative generation outside of the lexicon, as far as I know, appears in "Remarks", where we are urged to

Notice also that although gerundive nominalization applies freely to sentences with verb phrase adjuncts, this is not true of the rules for forming derived nominals. Thus we have (15) but not (16):

(15) his criticizing the book before he
read it...

(16) *his criticism of the book before he
read it...

This too would follow from the lexicalist assumption, since the true verb phrase adjuncts such as before-clauses...will not appear as noun complements in base noun phrases.

The examples (15) and (16) raise interesting questions relating to the matter of acceptability and grammaticalness...If the lexicalist hypothesis is correct, then all dialects of English that share the analysis of adjuncts presupposed above distinguish the expressions of (15), as directly generated by the grammar, from those of (16), as not directly generated by the grammar. Suppose that we discover, however, that some speakers [italics mine--LM3] find the expressions of (16) quite acceptable. On the lexicalist hypothesis, these sentences can only be derivatively generated. Therefore we should have to conclude that their acceptability to these speakers results from a failure to take note of a certain distinction of grammaticalness. We might propose

that the expressions of (16) are formed by analogy to the gerundive nominals (15), say by a rule that converts X-ing to the noun X nom (where nom is the element that determines the morphological form of the derived nominal) in certain cases...

This already suggests one rule of thumb for excluding bad analogies:

Good Informant Law

At least some speakers find derivatively generated forms in fact bad, without showing separate indications of a relevant dialect split with the other speakers.

Emonds anticipates us on this point when he says with reference to the partially acceptable application of root transformations in embedded clauses:

...the general heuristic subscribed to here for...sentences of doubtful grammaticality (acceptability judgments being erratic) is that they are ungrammatical, provided that they are not semantically difficult or of undue length or embedding. The reason for this is that it would be hard to explain even slight unacceptability for sentences that are relatively short and simple, semantically clear, and perfectly grammatical. But it is to be expected that intelligent language users would possess strategies of interpretation to render sentences that are relatively short and simple, semantically clear, and slightly ungrammatical perfectly understandable and nearly perfectly acceptable...

[Emonds (1972, p. 15n)]

Another test of derivative generation was also suggested in "Remarks":

...if the expressions of (16) are directly generated, we would expect them to show the full

range of use and meaning of such derived nominals as his criticism of the book. If, on the other hand, they are derivatively generated in the manner just suggested, we would expect them to have only the more restricted range of use and meaning of the expressions of (15) that underlie them. Crucial evidence, then, is provided by the contexts (17) in which the derived nominal his criticism of the book can appear, but not the gerundive nominals (15) (with or without adjuncts):

- (17) --- is to be found on page 15.
I studied --- very carefully.

The fact seems to be that speakers who accept (16) do not accept (13), even though they accept (19):

- (18) a. His criticism of the book
before he read it is to be
found on page 15.
b. I studied his criticism of
the book before he read it
very carefully.
(19) a. His criticism of the book is
to be found on page 15.
b. I studied his criticism of the
book very carefully.

If correct, this indicates that speakers who fail to distinguish (16) from (15) are not aware of a property of their internalized grammar, namely that it generates (16) only derivatively, by analogy to the gerundive nominal...

We can state the idea of this passage more generally as the

Law of Aggravation

Aggravating the violation that leads to ungrammatical acceptability causes it to blow up for all speakers.

The relevant notion of "aggravating the violation" covers at least two general cases, one where the relevant derivation involves a second violation, as in our gradation of Wh Island violations and allied examples (S4.2 *supra*), and one where the first violation "feeds" (Kiparsky again) a transformation, as in the following case.

It is well known that there-insertion feeds Subject Raising, as in There is likely to be a riot. Also well known is the fact that there-insertion normally operates only on indefinite Subjects:

(1) *There [destressed] is the riot,

but marginally applies to definites in situations where the speaker is reminding someone of the existence of the thing referred to by the Subject:

(2) (*) There is the rutabaga in the icebox.

The interesting point for our purposes is that Raising is not fed by there-insertion on definite Subjects, as observed by Aissen (personal communication):

(3) *There is likely to be the rutabaga in the icebox.

This follows at once from our Law of Aggravation, if marginal forms like (2) are derivatively generated.

Our final restriction on derivative generation is suggested in the "On Wh-Movement" treatment of resumptive Pronouns in English:

It has been noted that English speakers sometimes use a construction with a pronoun where an island constraint would block relativization...; cf. Andrews (1975) for some discussion. I suppose that this is an ancillary process, not to be incorporated, strictly speaking, within the grammar.

The point is that such resumptive forms are

normally not available except to salvage island violations. Thus

- (4) (*)Daphne is the punk that I ain't no
coward but I'll never tangle with him,

is reasonably acceptable--especially
for Texans, apparently--in comparison to the
cognate form with extraction

- (5) *Daphne is the punk that I ain't no
coward but I'll never tangle with___,

while the simpler

- (6) *Daphne is the punk that I'll never
tangle with him

is our of the question alongside of the
acceptable cognate extraction

- (7) Daphne is the punk that I'll never
tangle with___

This suggests the

Direct Competition Law

There is no directly generated cognate to a
derivatively generated form.

It remains, of course, to define the needed notion of
"cognate", but unfortunately we cannot do so here.
The correct definition probably involves some
construction like the mapping Φ we used in
developing the Law of Congruity, or perhaps
some reference to meaning.

The same concept is apparently needed for
another principle, which treats cases like the
Subjunctive-Infinitive alternation in

French. Consider, for example, a verb like vouloir, which occurs in the structures exemplified in:

(8) Je veux qu'il marche plus vite.

(9) Je veux marcher plus vite.

The language uses the Infinitive for the Control construction here and the Subjunctive where the Subjects of the two clauses differ. The problem is that we cannot have the Subjunctive with "coreferent" Subjects:

(10) *Je veux que je marche plus vite.

The standard approach to this problem is to treat (9) as derived from (10)* by an obligatory rule of Equi NP Deletion, which is complicated so as to convert Subjunctive to Infinitive. Besides the fact that this approach merely stipulates what we want to explain, it is inadmissible on EST principles excluding "side effect" compounding and rule-particular specifications or obligatoriness. Similarly, (10)* cannot be ruled out by Disjoint Reference (Chomsky, 1973), not only because the Subjunctive is finite (see George and Kornfilt, 1978), but also because we have the analogous form

(11) Je veux que nous marchions plus vite.

I submit that (10)* is ruled out simply because the simpler cognate (9) is grammatical, due to the principle of

Filtering by Analogy

Assign * to any structure if it has a strictly simpler grammatical cognate.

This analysis reconstructs on principled grounds

the guiding intuition by behind the standard account.

A comparable example of this principle appears in the Control system for English gerunds:

- (12) I regret your reading the diary.
- (13) I regret reading the diary.
- (14) *I regret my reading the diary.
- (15) I regret our reading the diary.

Again Disjoint Reference cannot rule out (14)*, this time because of A/A, as shown by the grammaticality of (15). As in the French example, the otherwise mysterious status of (14)* is explained on the assumption of Filtering by Analogy.

The same assumption may solve a potential problem left over from our discussion of phrase structure conjunction (S5.1 *supra*). Recall that we considered the possibility that (a)or-N

coordination was generated by ellipsis in order to explain some observations of Koutsoudas. If our suggestion was correct, we need to explain why an analogous elliptical derivation of

(16) *Bob and Bill is working
does not work. Filtering by Analogy solves this problem, excluding (16)* because of the existence of the strictly simpler derivation of the cognate Bob and Bill are working without ellipsis.

Bibliography

- Akmajian, Adrian, Culicover, Peter, and Wasow, Thomas (eds.) Formal Syntax (New York: Academic Press, 1977)
- Bach, Emmon and Harms, Robert (eds.) Universals in Linguistic Theory (New York: Holt, Rinehart and Winston, 1968)
- Banfield, Ann (1973a) Stylistic Transformations: A Study Based on the Syntax of Paradise Lost, University of Wisconsin, Madison, Ph.D. Dissertation
- Banfield, Ann (1973b) "Narrative Style and the Grammar of Direct and Indirect Speech", Foundations of Language, vol.10, pp.1-39
- Bresnan, Joan (1970) "On Complementizers: Toward a Syntactic Theory of Complement Types", Foundations of Language, vol.6, pp. 297-321
- Bresnan, Joan (1975) "Comparative Deletion and Constraints on Transformations", Linguistic Analysis, vol.1, pp. 25-74
- Bresnan, Joan (1976a) "On the Form and Functioning of Transformations", Linguistic Inquiry, vol.7, pp.3-40
- Bresnan, Joan (1976b) "Evidence for a Theory of Unbounded Transformations", Linguistic Analysis, vol.2, pp.353-393
- Cattell, Ray (1976) "Constraints on Movement Rules", Language, vol.52, pp.18-50
- Chomsky, Noam (1955) The Logical Structure of Linguistic Theory, published in 1975 by (New York: Plenum)
- Chomsky, Noam (1957) Syntactic Structures (The Hague: Mouton)
- Chomsky, Noam (1961) "On the Notion 'Rule of Grammar'", cited from Fodor, J. and Katz, J. (eds.)
- Chomsky, Noam (1965) Aspects of the Theory of Syntax (Cambridge: M.I.T.)
- Chomsky, Noam (1967) "Remarks on Nominalization", cited from Studies on Semantics in Generative Grammar (The Hague: Mouton, 1972)
- Chomsky, Noam (1970) "Deep Structure, Surface Structure, and Semantic Interpretation", cited from Studies on Semantics in Generative Grammar (The Hague: Mouton, 1972)
- Chomsky, Noam (1972) Studies on Semantics in Generative Grammar (The Hague: Mouton)

- Chomsky, Noam (1973) "Conditions on Transformations", cited from
Essays on Form and Interpretation (New York: Elsevier, 1977)
- Chomsky, Noam (1976) "Conditions on Rules of Grammar", cited from
Essays on Form and Interpretation (New York: Elsevier, 1977)
- Chomsky, Noam (1977a) Essays on Form and Interpretation
(New York: Elsevier)
- Chomsky, Noam (1977b) "On Wh-Movement", cited from Akmajian, A.,
Culicover, P. and Wasow, T. (eds.)
- Chomsky, Noam (1978) "On Binding", manuscript
- Chomsky, Noam and Halle, Morris (1968) The Sound Pattern of English
(New York: Harper and Row)
- Chomsky, Noam and Lasnik, Howard (1977) "Filters and Control",
Linguistic Inquiry, vol.8, pp. 425-504
- Cinque, Guglielmo (1977) "The Movement Nature of Left Dislocation",
Linguistic Inquiry, vol.8, pp. 397-412
- Clark, Mary (1974) "An Analysis of EQUI Which Suggests that
Sentences Must Be Put Together from the Bottom Up", in
Kaisse, E. and Hankamer, J. (eds.)
- Culicover, Peter and Wexler, Kenneth (1977) "Some Implications of a
Theory of Language Learnability", in Akmajian, A., Culicover, P.,
and Wasow, T. (eds.)
- Dougherty, Ray (1968) A Transformational Grammar of Doordinate
Conjoined Structures, M.I.T. Ph.D. Dissertation
- Dougherty, Ray (1970) "A Grammar of Coordinate Conjunction: I&II",
Language, vol. 46, pp. 850-898, vol. 47, pp. 298-339
- Emonds, Joseph (1970) Root and Structure-Preserving Transformations,
M.I.T. Ph.D. Dissertation
- Emonds, Joseph (1973) "Constraints on Phrase Structure Configurations",
Stanford Occasional Papers in Linguistics, no. 3, pp. 115-136
- Emonds, Joseph (1976) A Transformational Approach to English Syntax
(New York: Academic Press)
- Emonds, Joseph (1979) "Appositive Relatives Have No Properties",
Linguistic Inquiry, vol. 10, pp. 211-242
- Fiengo, Robert (1977) "On Trace Theory", Linguistic Inquiry, vol. 8,
pp. 35-62

- Fodor, J. and Katz, J. (eds.) The Structure of Language
(Englewood Cliffs: Prentice-Hall, 1964)
- Ford, A., Reighard, J. and Singh, R. (eds.) N.E.L.S. VI, Montreal Working Papers in Linguistics, May 6, 1976
- George, Leland (1975) "Generalized Transformations", manuscript
- George, Leland (1978a) "Ellipsis and Deletion Across the Board", manuscript
- George, Leland (1978b) "On the Universality of a Filter", M.I.T. Lecture
- George, Leland (1979) "Introspection and Interpretation", Brown University lecture
- George, Leland and Kornfilt, Jaklin (1978) "Finiteness and Boundedness in Turkish", to appear in a volume edited by Frank Heny
- George, Leland and Hoffman, Craig (1979) "Restricting the Theory of Lexical Selection", University of Connecticut lecture
- George, Leland and Kornfilt, Jaklin (1979) "Against Shallow Interpretation", Brown University lecture
- George, Leland and Hawkins, William (1980) "Filters and Mistakes", Abstract submitted to G.L.O.W., 1980
- George, Leland and Kornfilt, Jaklin (1980) "Not Quite Razing Raising", to be presented at the 1980 G.L.O.W. Colloquium
- Gleitman, Lila (1965) "Coordinating Conjunctions in English", Language, vol. 41, pp. 260-293; cited from Reibel, D. and Schane, S. (eds.)
- Grimshaw, Jane (1977) English Wh-Constructions and the Theory of Grammar, University of Massachusetts, Amherst, Ph.D. Dissertation
- Grosu, Alexander (1977) "Towards a Realistic Model of the Human Language Faculty: A Review of Noam Chomsky's Reflections on Language", Lingua, vol. 41, pp. 169-182
- Hale, Kenneth, Platero, Paul and Jeanne, La Verne Masayesva (1977) "Three Cases of Overgeneration", in Akmajian, A., Culicover, P. and Wasow, T. (eds.)
- Hankamer, Jorge (1971) Constraints on Deletion in Syntax, Yale University Ph.D. Dissertation
- Hawkins, William (1980) "Binding and Bounding", manuscript

- Hoffman, Craig (1980) The Syntax of Prepositional Phrases in English, forthcoming University of Connecticut Ph.D. Dissertation
- Hornstein, Norbert (1975) "S and the X' Convention", Linguistic Analysis, vol. 3, pp. 137 - 176
- Jackendoff, Ray (1977) X' Syntax: A Study of Phrase Structure (Cambridge: M.I.T.)
- Kaisse, Ellen and Hankamer, Jorge (eds.) N.E.L.S. V, 1974
- Kayne, Richard (1975) French Syntax (Cambridge: M.I.T.)
- Kayne, Richard and Pollock, Jean-Yves (1978) "Stylistic Inversion, Successive Cyclicity, and Move NP in French", Linguistic Inquiry, vol. 9, pp. 595-621
- Kimball, John (ed.) Syntax and Semantics, vol. 4 (New York: Academic Press, 1975)
- Kiparsky, Paul (1968) "Linguistic Universals and Linguistic Change", in Bach, E. and Harms, R. (eds.)
- Kiparsky, Paul (1979) "Metrical Stress Assignment is Cyclic", Linguistic Inquiry, vol. 10, pp. 421-441
- Kornfilt, Jaklin (1980) "Backward Gapping: An Embedded Mystery in Turkish", manuscript
- Koster, Jan (1977) "On the Form of Filters", manuscript
- Koster, Jan (1978) Locality Principles in Syntax (Dordrecht: Foris)
- Koutsoudas, Andreas (1971) "Gapping, Conjunction Reduction, and Coordinate Deletion", Foundations of Language, vol. 7, pp. 337-386
- Kuroda, S.-Y. (1968) "English Relativization and Certain Related Problems", Language, vol. 44, pp. 244-266; cited from Reibel, D. and Schane, S. (eds.)
- Lakoff, George and Peters, Stanley (1966) "Phrasal Conjunction and Symmetric Predicates", cited from Reibel, D. and Schane, S. (eds.)
- Lasnik, Howard and Kupin, Joseph (1977) "A Restrictive Theory of Transformational Grammar", Theoretical Linguistics, vol. 4, pp. 173-196
- Lees, Robert (1960) The Grammar of English Nominalizations (The Hague: Mouton)

- Lyons, John (1968) Introduction to Theoretical Linguistics
(Cambridge: Cambridge University Press)
- Maling, Joan and Zaenen, Annie (1978) "The Nonuniversality of a
Surface Filter", Linguistic Inquiry, vol. 9, pp. 475-497
- McCawley, James (1968) "The Role of Semantics in a Grammar", in
Bach, E. and Harms, R. (eds.)
- Perlmutter, David (1968) Deep and Surface Structure Constraints in
Syntax, M.I.T. Ph.D. Dissertation
- Peters, Stanley and Ritchie, Robert (1973) "On the Generative Power
of Transformational Grammars", Information Sciences, vol. 6,
pp. 49-83
- Reibel, David A. and Schane, Sanford A. (eds.) Modern Studies in
English: Readings in Transformational Grammar (Englewood
Cliffs: Prentice-Hall, 1969)
- Rizzi, Luigi (1977) "Violations of the Wh-Island Constraint in
Italian and the Subjacency Condition", manuscript
- Ross, John Robert (1967) Constraints on Variables in Syntax, M.I.T.
Ph.D. Dissertation; cited from Indiana University Linguistics
Club version
- Ross, John Robert (1973) "Nearer to Vee", manuscript
- Ross, John Robert (1974) "Three Batons for Cognitive Psychology", in
Weimer, W. and Palermo, D. (eds.)
- Sag, Ivan (1976) Deletion and Logical Form, M.I.T. Ph.D. Dissertation
- Sapir, Edward (1921) Language (New York: Harcourt, Brace and World)
- Siegel, Dorothy (1974) Topics in English Morphology, M.I.T. Ph.D.
Dissertation
- Smith, Carlota (1969) "Ambiguous Sentences with and", in Reibel, D.
and Schane, S. (eds.)
- Tai, James (1969) Coordination Reduction, Indiana University
Ph.D. Dissertation
- van Riemsdijk, Henk (1977) The Binding Nature of Prepositional
Phrases, manuscript
- van Riemsdijk, Henk and Groos, Anneke (1979) "Matching Effects in
Free Relatives: A Parameter of Core Grammar", to appear in the
Proceedings of the Pisa Colloquium on Markedness, April 20-22,
1979

- Vergnaud, Jean-Roger (1974) French Relative Clauses, M.I.T. Ph.D. Dissertation
- Weimer, Walter B. and Palermo, David S. (eds.) Cognition and the Symbolic Processes (Hillsdale, N.J.: Lawrence Erlbaum, 1974)
- Williams, Edwin (1971) "Small Clauses in English", published in Kimball, J. (ed.)
- Williams, Edwin (1974) Rule Ordering in Syntax, M.I.T. Ph.D. Dissertation
- Williams, Edwin (1978) "Across-the-Board Rule Application", Linguistic Inquiry, vol. 9, pp. 31-43
- Woisetschlaeger, Erich (1976) "On How Useful a Construct the Left-Branch Condition Is", in Ford, A., Reighard, J. and Singh, R. (eds.)