

Where (if anywhere) is transderivationality located?*

Pauline Jacobson
Brown University

in Peter Culicover and Louise McNally (eds.), *Syntax and Semantics 29: The Limits of Syntax*.
New York: Academic Press, 1998. pp. 303-336.

1. Introduction

In the early 1970s, a number of researchers (especially within the tradition of "Generative Semantics") suggested that the grammar contains rules/principles/constraints which block certain derivations on the basis of the existence of other derivations (e.g., Perlmutter, 1972); we will refer to any such device as a transderivational one. This idea has, of course, resurfaced in a good deal of recent literature, and has been examined much more systematically under the so-called "Minimalist Program" (Chomsky, 1995). The purpose of this paper is to raise the following question: Are there indeed transderivational processes within the grammar - or is their appearance, rather, an artifact? Obviously I cannot supply a full answer to this question here, and this paper should be taken in the spirit of *remarks* on issues requiring further investigation. Nonetheless, I do want to suggest that the grammar indeed does not contain transderivational processes as such. Many of the cases which appear to necessitate this device seem likely to be the result of processing and/or production principles, and others - as I will argue in Sec. 3 - should simply be reanalyzed in other ways.

Let me qualify this claim at the outset: the concern here is not on all transderivational processes, but only on those which are driven by some notion of "economy" or "minimality". Put differently, most cases which have been argued to require transderivationality involve situations in which one derivation is blocked because of the existence of some "simpler" option, and these are the cases of concern here. The reasons for being suspicious of the existence of such cases is quite straightforward: If some device computes a set of alternatives and then selects the simplest one, it is a complete accident that the selection is for the simplest. Once the full set of alternatives (the "candidate set", in current parlance) has been computed, it is just as easy to locate the most complex, the third simplest, or any other possibility¹ (including

possibilities which have nothing to do with simplicity). Thus, if there are indeed cases where the simplest candidate appears to "win out" over other possibilities, then this becomes unmysterious only if we can view the system as being driven in some way to that alternative rather than selecting the simplest from among a set of fully computed alternatives. By being "driven" to some alternative, I include the possibility of a "race" model - the system computes more than one candidate in parallel, but stops once the simplest (i.e., shortest) candidate is fully computed. (Indeed I will suggest that this is the view which makes the most sense.) Note, then, that a race does involve a kind of selection but - quite crucially - it does not involve selection among fully computed alternatives; here, the "non-winning" candidates are in fact never fully computed. However, as we understand formal grammatical principles, there is no obvious way to locate this kind of process in the grammar itself. There are, however, ways to think of processing and production mechanisms in this light. I will elaborate on these points below. Note, incidentally, that the basic strategy of reanalyzing grammatical constraints as the result of processing effects is also explored in detail in Kluender's paper (this volume).

First, though, let me stress again that the goals of this paper are modest. I will not attempt to provide any model of processing and/or production which truly explains away the numerous cases in the literature which appear to require transderivationality. My aim is simply to suggest first of all that a productive research strategy is to look for explanations along these lines and/or to reanalyze the cases more carefully. The conclusion that the grammar itself includes economy-driven transderivational mechanisms would seem to be a conclusion of last resort.

The remainder of this paper, then, has two parts. In Sec. 2 I give some background assumptions, and then elucidate further the remarks above. In particular, I will give examples of a few cases of apparent transderivationality which seem to lend themselves to processing and/or production explanations. But not all cases which have been proposed as instances of economy-driven transderivationality in the grammar easily lend themselves to a processing and/or production reanalysis. Sec. 3 then considers just such a case which has been discussed in

recent literature (Fox, 1995; Reinhart, 1995). I argue that here the basic generalizations leading to Fox's transderivational analysis are not correct - and the effects of this case probably arise from the interaction of several factors - some of which having to do with the fine-grained semantics of the case at hand, and some of which have to do with the interaction of this with discourse considerations.

2. Transderivational Effects as an Artefact of Processing and/or Production Principles

2.1. Background Assumptions

It seems uncontroversial that speakers are equipped with some kind of "encoding mechanism" - a mechanism allowing meaning (or, more generally, communicative intent) to be mapped into linguistic expressions - ultimately, into phonetic forms. Similarly, speakers are obviously equipped with some kind of "decoding mechanism" - which allows phonetic forms to be mapped into meanings (or again, more generally, into communicative intents). These mechanisms are of course complex - involving a combination of language-particular knowledge (surely, they must draw on language-particular grammars) and more general principles of perception as well as principles of cooperative discourse such as Gricean principles. In addition to encoding and decoding mechanisms, however, just about all work in modern linguistic theory assumes that speakers also know a "grammar" - a system of statements (call these rules, principles, or whatever) which specify a set of well-formed expressions (i.e., phonetic representations) and which pair each with a meaning. More broadly, the grammar of any particular language can be taken as a set of statements specifying a set of well-formed pairs of phonetic representations and meanings. I will make the usual simplification of thinking of "phonetic representations" as a string of words, and will take "meanings" to be some kind of model-theoretic objects (not a logical form, but an interpreted logical form - it is an open question as to whether or not grammars make use of a "logical form" as part of its apparatus for pairing phonetic representations with meaning). I will use the term derivation here in an extremely broad way - any sequence of steps licensed by the grammar which predicts a certain phonetic form-meaning pair to be well-formed is a derivation. Thus I am concerned here not just

with the question of what are the limits of syntax, but more broadly with the question of what are the limits of grammar.

Before continuing, let me clarify one point concerning the status of the "grammar" as opposed to the production and processing mechanisms. Most work in modern linguistic theory takes it as axiomatic that there is indeed a "grammar" (a system of rules specifying a set of well-formed phonetic representation - meaning pairs) which is separate from these other mechanisms. But since the relationship between the grammar and the production/perception mechanisms is at this point not very well understood, one might argue that the notion of a "grammar" is just a convenient fiction - it is nothing more than some subpart of these other mechanisms or some abstraction derived from them. Should this turn out to be the case, then the main point of this paper will be trivially true. I will be arguing that transderivational constraints should not be located in the grammar, and that their appearance in certain cases is just an artefact of production and processing principles. Obviously if there is no grammar, then this conclusion follows immediately (and is not a particularly interesting one), and the question as to the "limits of grammar" of course becomes moot. Nonetheless - as noted above - the adoption of the hypothesis that one can isolate a system of rules called "grammar" (independent of their implementation in real-time processing and production) has at the very least been an extremely fertile research strategy - and I will continue to adopt this hypothesis here.

We can now return to the question of whether or not the grammar itself contains any rules or principles which refer to a set of possible derivations.² Such principles would apply to the output of some computation of a set of alternatives, where each of the derivations within the candidate set would itself be licensed by other principles/rules in the grammar. It has often been noted that such a computation is itself quite complex; this is thoroughly discussed in Johnson and Lappin (to appear). Hence, given the general assumption that a desiderata for a theory of grammar be that one can find a reasonably simple algorithm for determining the

well-formedness of some derivation according to the grammar, there may well be computational reasons to doubt the existence of transderivational principles.

But let me leave aside issues of computational complexity; my concern here is not directly with this issue but rather with a closely related point which nonetheless has a slightly different spin on it. That is, as noted in the introduction, any time that a system can access a set of possibilities, there is no particular reason why it should choose the "simplest" one over any other member of the candidate set.³ Before continuing, we need some working definition of "simplest". As a working definition, we will adopt the obvious one: one derivation is simpler than another if the first involves fewer derivational steps.⁴ (Recall that by a "derivation" I mean nothing more than a sequence of steps which specifies that some pair <phonetic form, meaning> is well-formed.) Thus once the system has computed more than one derivation, there is no reason why the derivation with fewer steps should have any privileged status. As noted above, the overall system would be just equally simple if what were picked was, say, the derivation with the most number of steps. For that matter, the number of derivational steps should have no privileged status over any other conceivable parameter by which we might distinguish among derivations.

A more straightforward scenario, then, is that there is never actually selection of the simplest alternatives - instead, some system is driven automatically to the simplest or, perhaps, it "selects" the simplest derivation before it even has a chance to fully compute a more complex one. But as I will argue below, it is difficult to think of this sort of scenario as being located within the grammar. The reason is twofold. First, the grammar has no notion of "success", "failure" or "goals", and second, grammatical derivations are not located in real time. But encoding and decoding mechanisms do take place in real time and are goal oriented, and thus cases of the relevant sort can coherently be relocated there. I will clarify this point in Sec. 2.2.

Consider then the following three kinds of cases (I use P for a phonetic form, and M for a meaning)

Case A. Suppose we have two derivations D_1 and D_2 as follows:

D_1 pairs s_1 with M_1

D_2 pairs s_1 with M_2

where each step of both D_1 and of D_2 appear to be licensed by the grammar, but where

D_2 involves more steps than D_1 .

Case B. Suppose we have two derivations D_1 and D_2 as follows:

D_1 pairs s_1 with M_1

D_2 pairs s_2 with M_1

where each step of both D_1 and of D_2 appear to be licensed by the grammar, but where

D_2 involves more steps than D_1 .

Case C. Suppose we have two derivations D_1 and D_2 as follows:

D_1 pairs s_1 with M_1

D_2 pairs s_1 with M_1

where each step of both D_1 and of D_2 appear to be licensed by the grammar, but where

D_2 involves more steps than D_1 .

Suppose further that we find cases in which the output of D_2 indeed does not seem to exist (obviously this will be difficult to determine in Case C; we return in Sec. 3 to a potential case along these lines discussed in Fox (1995)). In each such case one might appeal to "economy" considerations within the grammar itself - the grammar rules out D_2 on the basis of the existence of D_1 . But in each such case another explanation suggests itself: Examples of Case A abound in the literature (and, some of the earliest proposals for transderivational rules were based on examples of this type), but many of these examples are now generally agreed to be the result of processing principles - something which drives the decoder to the meaning associated with D_1 . Instances of Case B can be taken to be the other side of the coin: these can hopefully be seen as the result of some principle driving the encoder to the form associated with D_1 . (Cases of the type in C could be due either to the encoder or the decoder.) But an important caveat is in order: merely relegating these cases to the realm of processing and/or production does not in and of itself solve the problem. If either of these systems gets to "see" more than one alternative and then choose, the puzzle simply reemerges - for this again involves selection

from a set of fully computed alternatives. We return to this momentarily, but first let us make the discussion more concrete by turning to an actual example of Case A.

2.2. Case A: The "Lazy" Decoder

One of the first proposals for the existence of transderivational rules centered on just this sort; this concerns "free" word order in Russian. Perlmutter (1972) points out that word order in Russian is relatively free, except in cases where neutralization in the morphology would cause ambiguity. Thus the words for "son" and "father" are morphologically distinct in nominative and accusative: *syn* (son-nom) vs. *syna* (son-acc) and *otec* (father-nom) vs. *otca* (father-acc). The case distinction is, however, neutralized for "daughter" and "mother": *doc'* (daughter-nom or acc) and *mat'* (mother-nom or acc). In (1), then, either SVO or OVS order is possible: both (1a) and (1b) mean "The son sees the father". In both (2a) and (2b), on the other hand, only the SVO interpretation is possible.

- (1) a. *Syn vidit otca.* "The son sees the father"
 b. *Otca vidit syn.* "The son sees the father"
- (2) a. *Doc' vidit mat'.* "The daughter sees the mother" (unambiguous)
 b. *Mat' vidit doc'.* "The mother sees the daughter" (unambiguous)

This is surprising, since whatever allows OVS order in (1b) should do the same in (2b). Thus Perlmutter suggested that there is a transderivational principle in the grammar: the derivation that pairs (2b) with the OVS meaning is blocked on the basis of the existence of another derivation pairing this with the SVO meaning. Perlmutter does not directly cast this in economy terms, but this seems implicit in his discussion. That is, given an explanation relying on transderivationality in the grammar, we need an account of why the SVO derivation blocks the OVS derivation rather than the other way around. Lurking in the background, then, is the assumption that SVO order is "unmarked" - i.e., that it involves fewer steps than the derivation of OVS order. Hence this is a straightforward instance of Case A: the derivation pairing (2b) with the OVS meaning is longer, and so it is blocked on the basis of the existence of a shorter derivation which pairs (2b) with a different meaning.

The literature is replete with cases of this general sort - including various cases which have gone under the rubric of "Minimal Attachment" (Frazier and Fodor, 1978; Frazier, 1985). It has, however, become relatively standard to think of these types of cases as being the result of a "lazy processor" - one which does the minimal work possible so as to pair a form with a meaning. Before continuing, one important point should be noted: many of the Minimal Attachment cases differ from the above in that they involve the preference of one reading over another, rather than one reading absolutely blocking another. (Actually, I do not know whether one can get the missing reading of (2) in the right context.) If both types of cases are relegated to processing strategies, then the question arises as to why in some cases the "simpler" path results in a preference, while in other cases it (apparently) blocks the other path - or at least makes the other path far more difficult to access. In the ensuing discussion, I lump these two types of cases together - for it is not clear that there really are just two distinct types of cases. Even if we were to confine our attention purely to cases where one reading is *preferred* (rather than absolutely blocking the other) there is a continuum as to how available the second reading is. I assume, then, that there really is no reason to draw a line and say that cases where one reading appears to absolutely block another are an entirely different sort of beast rather than being the extreme end of a continuum.

But what does it really mean to say that the processor takes the shortest route? First, some terminology and assumptions. The input to the processor is some phonetic form, and the job of this system is to output a meaning. Since the processor presumably does so by going through some series of steps, I will use the term *decoding derivation* to mean any series of steps which begins with a form and ends with a meaning. Note further that a processing explanation for Case A effects assumes that - in the relevant cases - there is some correlation between the number of steps in the grammatical derivation which pairs with M and the number of steps in the decoding derivation mapping to M. Put differently: the cases of concern here are ones where the grammar appears to contain a D₂ yielding the pair $\langle 1, M_2 \rangle$ and a shorter D₁ yielding the pair $\langle 1, M_1 \rangle$. If we place the burden on the processor to compute only M₁ as

output due to processing economy considerations, then we are assuming that the processor also goes through fewer steps to arrive at M_1 from 1. This assumption is, in fact, implicit or explicit in much of the work on sentence processing, and in the remarks below I will continue to assume that for at least the cases at hand, this is correct.⁵

But we are still left with an important question: what does it really mean to say that the processor makes the minimal number of moves possible so as to arrive at a meaning? How does it know in advance what path will result in the minimal number of moves? Suppose that at any given stage in the computation, the decoder has available to it a set of local choices. It has no way to know which local choice will lead to the globally shortest path. And we do not want to conclude that it gets to "see" ahead in order to determine this - after all, "*seeing*" the paths means *computing* the paths, and so this merely recreates the paradox.

There is one concrete way to resolve this and perhaps a second - I begin with the latter first. The idea here would be that - although there is a set of possible "moves" available at each step in a decoding derivation - the moves are ordered in such a way that the system can only see the move at the top of the ordering and it is forced to make this move. Thus given a particular phonetic form, then, there is initially a unique path available to the system. (The issue is of course complicated by the fact that on-line processing takes place before the entire sentence is heard, but we will ignore this important complication for the sake of exposition.) However, should the processing system fail (i.e., not come up with any meaning) it can then backtrack to some earlier point in the computation and try a different path. Note that this involves keeping track of the prior moves. This is the view taken in much of the work by Frazier (cf., Frazier, 1985) and is often assumed to be the explanation for garden path effects. (Although note that garden path phenomena differ from the effects discussed here in that - by definition - these involve cases of local ambiguities which disappear downstream.)

To illustrate the basic idea with the Russian case: the idea would be (roughly) that the processor is forced to initially consider some path where the sentence is analyzed as having SVO order, and then it checks case-marking. For the case of (2), SVO order and case marking

are not mutually consistent, and so the processor will stop, and will not come up with any alternative analyses. In the case of (1b), however, these two are not, and a processor which has embarked on that path will fail. It can therefore then go back, and recompute using OSV order.

Although it is quite common to assume that the decoder is driven to the shortest route in something like the way sketched above, there remain open questions about this. In particular, this leaves open the question of why and how it is that the ordering among alternative (local) choices just happens to be the one leading (in general) to the shortest path.

An alternative scenario which avoids this difficulty is a "race" or "competition" model - whereby the processor computes all possible analyses in parallel but stops as soon as one succeeds. This in fact is suggested in Frazier and Fodor (1978). It thus follows that derivations with more steps will never be computed. Thus assume again that the decoding derivation which pairs OVS meaning with (2b) involves at least one more step than the pairing of SVO meaning with this form. The two will be computed in parallel, but the decoder will stop as soon as a meaning is computed, and will thus "stand pat" with the SVO meaning. In that case, it never fully computes the OVS meaning. In (1b), on the other hand, it cannot stand pat with SVO and must keep going. Notice that it becomes automatic on this view that the shortest path is the selected one. Moreover, the selection for the shortest here is rooted in the fact that the computation does, indeed, take place in real time. Note too that - as pointed out in Frazier and Fodor (1978) - garden path effects are quite compatible with a race model. Consider the well-worn garden path: *The horse raced past the barn fell*. Suppose that both the main clause analysis and the NP analysis of *the horse raced past the barn* are being computed in parallel, but where the computation of the NP analysis involves more steps. At the point that *fell* is reached, the NP analysis will still be lagging behind the main clause analysis of *the horse raced past the barn*, and so it will still take the processor extra time to get "back on track", as it were.

The choice between these two alternatives (and perhaps yet others) obviously depends on detailed models of processing and detailed on-line experimental results.⁶ But my aim here is

not to provide a detailed processing model, but rather to suggest that apparent Case A effects can coherently be seen as artefacts of processing principles. One might at this point ask: What is the difference between the grammar and the decoder (and also the encoder) which allows us to coherently relocate economy principles in the latter but not in the former? There are two key differences. First, both of the alternative models explored above ("driven to the shortest route" and "race") crucially rely on a notion of success or failure. The important point to observe about transderivational effects is that some path is blocked only when there is a shorter available path. This means that when the longer path succeeds, it is because the attempted shorter path fails (in the case of the decoder, for example, the attempted shorter path does not result in a meaning). In that case, the system can backtrack (under the first alternative), or continue on with the longer path (in the race model). But the grammar itself has no analogous notion of success or failure, and so there is no obvious way to translate this into a model of grammatical derivations. Second, the race model exploits the fact that decoding (or, encoding) takes place in real time - and so it makes sense that time pressures would force one path to stop as soon as another succeeds. (Again this point is stressed in Frazier and Fodor (1978).) But grammatical derivations do not take place in real time, and so again it is difficult to see any way to mimic the race model within the grammar itself.⁷

2.3. Case B: The Lazy Encoder

The literature also contains a number of familiar instances of Case B - where a shorter derivation $\langle M_1, 1 \rangle$ wins out over a longer derivation $\langle M_1, 2 \rangle$.⁸ Once again these lend themselves to explanations based on "the lazy encoder" rather than explanations based on derivational selection within the grammar. In other words, we can again assume that in the relevant cases the derivational steps involved in mapping a meaning into a phonological form has some non-arbitrary connection with the number of derivational steps licensing that pairing by the grammar, and we can assume that the encoder is set up so as to automatically compute - or, to stop when it finds - the shortest route.

Let us again illustrate with an actual example - this particular example has, to my knowledge, never actually been described in transderivational terms, but it is quite parallel to others which have. Moreover, this case will require an extended notion of "meaning": simple propositional meaning will not be enough. We need to consider the fact that the speaker's job is actually to begin with a "communicative intent" and pair it with a phonetic form.

The example in question centers on Postal's "Crossover Constraint" Postal (1971). Thus Postal noted that sentences like (3), at best, questionable (see also Jackendoff, 1972):

(3) ?*John was seen by himself.

Postal's analysis relied on the assumption that Passive moves the underlying subject into the by-phrase while at the same time moving the underlying object to subject position; hence it switched the order of the subject and object. This, then, was subsumed under Postal's more general Crossover Constraint which blocked the movement of an NP over a coreferential (or, coindexed) NP.

A transderivationalist, however, might be tempted to reanalyze (3) as follows. Assume that the derivation of a passive sentence includes at least one more step than the derivation of a corresponding active. Despite the plethora of accounts of passive, this assumption does indeed follow under almost all. If passives involve movement, then obviously they contain the additional step (or steps) of movement. If, instead, passives are the result of a word-level rule mapping an active verb into a passive verb with a morphological and meaning change, then there is also the additional step of this mapping. Finally, early GPSG contained an account of passive via a metarule - a phrase structure rule introducing passives was derived from a rule introducing actives. All of these accounts, however, share the property that an extra step is involved in the grammatical derivation of a passive (recall that the term "derivation" here is being used broadly to designate any sequence of steps licensing a well-formed pair $\langle S, M \rangle$).⁹ Note further that while actives and corresponding agentive passives as in (4) (generally) have the same truth-conditions, they arguably have different communicative

intents. If subject position is associated with some discourse function, then (4a) and (4b) differ in at least that respect:

- (4) a. Mary saw John.
b. John was seen by Mary.

But now consider the pair in (5):

- (5) a. John saw himself.
b. *John was seen by himself.

Since *John* is in subject position in both, it is difficult to see any difference either in the propositional meaning or - in most circumstances - the communicative function of the two. One can therefore imagine a proposal which blocks the derivation in the grammar of (5b) on the basis of the existence of a derivation yielding (5a): (5a) has a shorter derivation and yet the two are paired with the same meaning (in the broadest sense).

I do not believe that anyone would actually propose this transderivational explanation (although a case discussed in Reinhart (1995) is reminiscent of this in some ways). The more plausible alternative is to let the grammar be silent on the subject of (5b) and to relegate the effect to the encoder. Although I don't know that this has actually been proposed anywhere in the literature, I believe that this idea has certainly been in the air, and suspect that most people indeed assume something like this as an explanation for the oddness of (5b). The rough and ready idea is that a speaker will not go to the trouble of encoding some communicative intent into a passive unless s/he has to - the assumption being that passives are, indeed, "more trouble" than actives. Since (5a) and (5b) are truth conditionally equivalent and are equivalent in terms of any discourse functions associated with the subject, (5b) will not be chosen.

There are many other well-known cases which lend themselves to reanalysis along these lines. Most notable are cases where the (apparent) violation of some (apparent) grammatical principle is amnesty just in case there would otherwise be a violation of some other grammatical principle (Perlmutter (1972) also discusses cases of this type). For example, it is well-known that resumptive pronouns in English are odd in general but quite natural in

islands - i.e., they are (comfortably) allowed just where their failure to occur would result in an ill-formed output (see Maling and Zaenen, 1982; see also Shlonsky 1992¹⁰). But a plausible production-based account would be one in which the principles of English grammar allow resumptive pronouns everywhere but where it was also ensured that an encoding derivation using a resumptive pronoun is longer than one without. If the speaker (encoder) is forced to take the shortest path yielding a well-formed output, then the only time the resumptive pronoun will "win" is when the other path dead-ends in an ill-formed output. (Notice incidentally that this kind of extragrammatical explanation for the distribution of resumptive pronouns interacts in an interesting way with the issue of whether (some) island constraints are themselves just the result of extragrammatical principles. Suppose that the latter is true. Then there is no obvious way to let the grammar allow resumptive pronouns just where island constraints would otherwise be violated. Even if we were to endorse economy-based transderivational principles in the grammar, the grammar could not block derivations with resumptive pronouns on the basis of the fact that it allows shorter derivations without them while still allowing resumptive pronouns in islands, because the grammar itself also allows for island violations.)

But again we need some more detailed notion of how it is that the encoder is sure to take the shortest path - merely saying the the encoder will "go to the least trouble" is somewhat mystical since it cannot know in advance which path will be longer. Again the same two basic scenarios as were discussed above suggest themselves here. On one, the encoder is driven to this path - at each step in its computation it can see only one possible move, but it can backtrack if the path fails. In the case where a passive is good, for example, the active must be tried first, but the the encoder can backtrack and try a passive if the active fails. (In this case the failure would be that the discourse functions associated with the active version are inappropriate.) But in the "Crossover" example in (5b) the passive will never be computed - in any situation (or, at least, most situations) where the passive path could have succeeded, then so could the active - but the active will have been tried first. Of course this general model is

subject to the same question that arose earlier: it is not clear why the local choices available to the system just happen to be ordered in such a way that the shortest path will win. The alternative is again a race model - both forms are computed in parallel, but the encoder stops once one path succeeds. If the active does the job, it will win out, and the passive will never be fully computed. Either way, though, the appearance of transderivationality is arguably an artefact of encoding principles.

3. A closer look at an apparent transderivational case

Above I briefly discussed two cases of apparent transderivationality, and suggested that both are rather typical of cases which can be relegated to processing and/or production mechanisms. In the remainder of this paper, I will shift gears a bit, and will examine more closely an intriguing case discussed recently in detail in Fox (1995) (see also Reinhart, 1995); this involves the interaction of wide scope objects and VP Ellipsis. The interesting thing about this case is that it will turn out to resist reanalysis in terms of processing or production mechanisms - the reason for this will be elucidated later - and at first blush it appears to provide striking confirmation for economy-driven transderivationality in the grammar. However, I will argue that a closer look reveals that this is not so: the transderivational explanation makes several incorrect predictions. I unfortunately do not have a full-fledged reanalysis accounting for all of the facts, but in Sec. 3.6 I will present the broad outlines of a possible reanalysis. The suggestion here is that the relevant effects arise from an interaction of the fine-grained semantics with discourse considerations.

3.1. Background: the analysis of VP Ellipsis and the analysis of wide scope objects

Discussing this case is somewhat complex, because there are a variety of available analyses both of wide scope object quantifiers and of VP Ellipsis. For the sake of exposition I will narrow these down quite a bit: in fact none of the packages which I will look at in detail is exactly the one assumed by Fox. However, as far as I can determine, this does no damage to the main points.

Thus note first that there are a variety of competing analyses of (6):

(6) Lucille read *The Brothers Karamazov*, and Susan will too.

(1) The earliest accounts (cf., Ross, 1967) took (6) to be the result of a rule which deleted the second VP under (some notion of) string identity with the first VP. But it has often been noted that string identity is insufficient - the "missing VP" in the second conjunct must have a meaning and/or LF which is the same as that of some other VP. (2) In view of this, Sag (1976) proposed that the phenomenon involved deletion of some target VP provided there is another VP whose LF is identical (in some sense) to that of the target's. This is essentially the view also taken by Fox (although his definition of "identity of LF" is a bit different from Sag's). (3) Note that one can slightly recast this. A third conceivable view is that (6) is the result of the deletion of a target VP, provided there is some other VP with the same meaning (in the model-theoretic sense) as that of the target VP. Unlike the proposal in (2), this proposal is, of course, compatible with a theory of grammar in which there is no significant level of LF. Call this the Deletion/Meaning Identity view. (4) A fourth view is the one which is probably the most common, and derives from Williams (1977). This is that there is no deletion of a VP in the syntax. Rather, the rules mapping surface structures to LFs copy in an LF representation which is identical to the LF representation of some other VP. (5) Finally, we can revise (4) in the same way that we did (2) so as to be compatible with a theory of grammar with no level of LF. That is, assume again that there is no deletion in the syntax, and that what is picked up here is the meaning of some VP rather than some LF representation. More concretely: assume that an auxiliary like *will* can, in the syntax, be "missing" its complement; this would follow from a rule operating on lexical items to map auxiliaries which want complement VPs into homophonous items which occur without their complement VPs. Assume further that the meaning of the new (complementless) auxiliary is *will'* applied to some free variable over properties. The value of that free variable is ultimately supplied by the meaning of some other (overt) VP in the discourse context. Call this the "Supply a Meaning" view.¹¹ Each of these views has certain advantages and certain disadvantages (and there are still other proposals in the literature; see, e.g., Dalrymple, Shieber, and Pereira (1991) and Lappin

(1996)), but since a thorough treatment of VP Ellipsis is beyond the scope of this paper I will narrow the discussion down to two, chosen partly for expository convenience. Since the interaction of VP Ellipsis with scope actually becomes more complex once we introduce a level of LF, the exposition will be simplest if we concentrate primarily on the theories in (3) and in (5) - the Deletion/Meaning Identity view and the Supply a Meaning view.¹²

Similarly, there are a variety of competing analyses of scope ambiguities as in (7):

(7) At least one woman read every book.

To be consistent with the choices above, we will need to adopt one which does not make essential use of LF as a significant level of representation. Moreover, in order to capture the essence of Fox's proposal, we need to consider an account in which the derivation resulting in wide scope on the object is more complex than the derivation resulting in object narrow scope.¹³

One account satisfying both desiderata is based on proposals in Kempson and Cormack (1983) and Hendriks (1987); these in turn derive from work on type-shifting by Partee and Rooth (1983). (The account sketched below is in fact essentially that of Kempson and Cormack, who are also concerned with giving an account of wide scope object readings in the context of their interaction with VP Ellipsis.)

Thus assume (following Partee and Rooth (1983)) that an ordinary transitive verb like *read* is listed in the lexicon as having its "lowest type" meaning. That is, *read'* is a relation between individuals, and hence of type $\langle e, \langle e, t \rangle \rangle$ (here and throughout I will ignore intensions). However, something is needed to allow quantified NPs to occur in object position. Rather than making use of any Quantifying In rule (Montague, 1974), Quantifier Lowering (McCawley, 1971; Lakoff, 1972) or its descendant Quantifier Raising (May, 1977), we will assume instead that there is an Argument Lift process - this shifts the meaning of *read* into a new meaning such that it wants an expression of type $\langle \langle e, t \rangle, t \rangle$ in object position. We will not spell out the rule in full generality yet; suffice it to say here that it maps *read* into a new verb (call it *read*₂) which is homophonous with the input, but whose meaning is as in (8):

(8) $read_2' = P[x[P(y[read'(y)(x)]]]$

In prose: $read_2'$ is that relation holding between an individual x (the subject) and a set of properties of individuals such that for all individuals x and property sets P , $read_2'(P)$ maps x to true just in case the property of "being-read'-by- x " is in the set of properties mapped to true by P . (Notice that - stripping away the intensions - this is the meaning that Montague (1974) assigned to $read$ in the lexicon.) Thus $read_2'(\text{every-book}')$ will map x to true just in case "every-book-was-read-by- x " is true.

Now consider how this combines with the subject position in (7); first we consider the case where the object has narrow scope (call this the NSO reading). Here things are rather simple. The subject denotes a set of properties (thus of type $\langle\langle e,t\rangle,t\rangle$), and so we can derive this reading by simply taking the subject to be the function taking the VP as argument (this is essentially Montague's analysis). But what about the case where the object receives wide scope? Here we will follow Hendriks (1987) in generalizing the argument lift rule: any argument position can lift from something of type e to something of type $\langle\langle e,t\rangle,t\rangle$. In other words, any verb wanting an individual in some argument position can shift to want a generalized quantifier instead in that position. The generalized rule is given in (9) (note that although (9) looks like a rule manipulating translations into λ -expressions, it can and should be seen as a rule mapping model-theoretic objects into others):

(9) Notational conventions: Let A^* stand for a sequence of λ 's each followed by a variables (where the variables can differ in their types). Similarly, let A^* be a sequence of variables (without the λ 's) matching those in the λ -expression. Finally, let X^* and Y^* be sequences of semantic types.

Rule: Let α be an expression whose meaning is of type $\langle X^*, \langle e, \langle Y^*, t \rangle \rangle \rangle$. Then there is a homophonous expression of type $\langle X^*, \langle \langle e, t \rangle, t \rangle, \langle Y^*, t \rangle \rangle$ whose meaning is:

$A^*[P[B^*[P(\lambda y[(A^*)(y)(B^*)])]]]$. (The cumbersome looking notation here simply allows the other argument slots of the original item to remain intact and occur in the same order that they normally would. The individual argument slot of the input verb switches to a generalized quantifier slot; the new meaning is such that the generalized

quantifier which ultimately occupies this slot is applied to a property formed by -
abstraction over the corresponding individual argument slot of the original verb.)

It turns out that if (9) applies first to the subject position and then to the object position, the result is the object wide scope reading (hereafter, the WSO reading). We leave it to the interested reader to work out the full details, but we can note that the meaning of our new verb - call it *read*₄ - is as shown in (10):

(10) $P [Q [P (z [Q (y [\text{read}'(z)(y)])])]]]$

This can then combine with the object and then with the subject, yielding WSO. The two derivations are sketched schematically in (11) where each is followed by a first-order representation of the final meaning. Since we will be concerned with the meaning of the VP, I represent this in boldface:

(11) D₁: (NS0)

read; *read*' ==> argument lift on object position:

read; *read*₂'

read every book; *read*₂' (every-book')

a woman read every book; *a woman*'(*read*₂' (every-book'))

the meaning: $x[\text{woman}'(x) \ y [\text{book}'(y) \ \rightarrow \ \text{read}'(x,y)]]$

D₂ (WSO)

read; *read*' ==> argument lift on subject position:

read; *read*₃' ==> argument lift on object position:

read; *read*₄'

read every book; *read*₄'(every-book')

a woman read every book; *read*₄' (every-book') (*a-woman*)

the meaning: $y[\text{book}'(y) \ \rightarrow \ x[\text{woman}'(x) \ \text{read}'(x)(y)]]$

Here D₂ is clearly more complex than D₁ as it involves one additional step. Incidentally, there is also a third possible derivation: (9) can apply first to the object position and then to

the subject position; in this case both subject and object are arguments of the new verb. This derivation will not concern us further; it yields the same meaning as does D_1 (i.e., NSO). (It is also more complex than D_1 , and Fox's proposal therefore makes the interesting prediction that it will be blocked on the basis of the existence of D_1 . However, this particular consequence of his proposal will not be examined here.)

Now what about the case where the subject NP is an ordinary, individual-denoting NP (such as a proper name or definite description)? For expository simplicity, let us follow Montague (1974) in assuming that all NPs (always) denote generalized quantifiers; hence the meaning of *Lucille* is $\underline{P[P(l)]}$.¹⁴ Consider, then, (12):

(12) Lucille read every book.

The most straightforward derivation for this is the one parallel to D_1 in (11); it is given in (13). However, there is also a derivation parallel to D_2 ; this too is shown in (13). But in this case - unlike in the case of (11) - D_1 and D_2 ultimately result in the same meaning.

(13) D_1 :

read; $read'$ ==> argument lift on object position:

read; $read_2'$

read every book; $read_2'$ (every-book')

Lucille read every book; $Lucille'(read_2' (every-book'))$

the meaning : $y [book'(y) \rightarrow read'(l,y)]$

D_2 :

read; $read'$ ==> argument lift on subject position:

read; $read_3'$ ==> argument lift on object position:

read; $read_4'$

read every book; $read_4'$ (every-book')

Lucille read every book; $read_4' (every-book') (Lucille')$

the meaning: $y[book'(y) \rightarrow read'(l,y)]$

3.2. The puzzle and the transderivational explanation

We are now in a position to turn to the interaction of VP Ellipsis (VPE) and wide scope object quantification. The story begins with Sag (1976), who noticed that while the first conjunct in (14) is ambiguous when it occurs alone, the ambiguity (apparently) disappears when followed by a VPE sentence - the first conjunct appears to have only the NSO reading. (15) and (16) are similar and have been discussed in subsequent literature:

(14) Someone hit everyone, and then Bill did. (modification of the S in Sag, 1976)

(15) A nurse examined every patient, and Lucille did too.

(16) At least one critic from the Times admires every movie, and Pauline Kael does too.

(Fox, 1995)

Sag's particular account of this need not concern us here,¹⁵ but we can note that under the accounts of VPE under consideration here combined with the account of wide scope quantification discussed above, this is surprising. (Actually, it is equally surprising under just about any theory; see fn. 15.) The reason is that the derivation of the first conjunct resulting in wide scope is D₂. This gives the VP the meaning *hit*₄(*everyone*'). Consider first the Deletion/Meaning Identity view of VPE, whereby a VP may be deleted under phonological and meaning identity with some other VP in the discourse context. Consider further the sentence (17), which would be the pre-VP-Deletion version of the second conjunct in (15):

(17) A nurse examined every patient, and Lucille examined every patient (too).

Here the VP *examined every patient* in the second conjunct can also have the meaning shown in (13) under Derivation 2, and so it should be able to delete. The facts are equally mysterious under the Supply a Meaning view. Here, we simply have a "missing" VP complement of *do*, and what is picked up is the meaning of some other VP. There is no reason why this can't be the meaning of the VP in the second derivation (D₂) shown in (13).

The next installment of the story is due to Hirschbuhler (1982), who showed that indeed Sag's generalization was incorrect. One can, in fact, get analogous cases in which the

first conjunct (as well as the second) has the WSO reading; Hirschbuhler's example was (18), and (19) and (20) are similar:

- (18) An American flag flew over every stadium, and a Canadian flag did too.
 (19) A doctor examined every patient, and a nurse did too.
 (20) At least one critic from the *Times* admires every movie, and at least one critic from the *Post* does too. (Fox, 1995)

What, then, explains the difference between the examples in (14) - (16) and those in (18) - (20)? One might take the key difference to have something to do with the "referentiality" of the subject (see Cormack (1984), Diesing (1982)), and indeed I will argue in below that this is part of the story.¹⁶ Fox, however, proposes a rather different explanation. Consider a theory in which VPE involves deletion (as in, e.g., the Deletion/Meaning Identity view), and where in (15) and (16) there is therefore a stage of the derivation at which the second conjunct has the full representation shown in (21):

- (21) and Lucille examined every patient too
 ... and Pauline Kael likes every movie too

Now in all cases the WSO reading will be possible if the meaning of its VP could be that shown in D₂ (modulo, of course, the difference in lexical items). But in the cases in (14) - (16), D₁ and D₂ result in the same meanings - note that this is not the case for the Hirschbuhler-type examples ((18) - (20)). Suppose then that the grammar blocks one derivation if there exists a shorter one with the same meaning - in other words, it includes an economy-driven transderivational principle for Case B and Case C. (This particular example happens to be an instance of Case C.) Then the mystery would be explained. Thus Fox argues that the effect here gives support for economy-driven transderivationality in the grammar.

3.3. Initial arguments against the transderivational explanation

There are, however, several reasons to doubt the transderivational explanation - intriguing though it may be¹⁷ - and some of these also militate against a recasting of the explanation in processing/production terms. My first objection is admittedly a world-view

dependent one and will be convincing only if one believes in either the Supply a Meaning theory or its analogue in (3) above (i.e., where the "missing meaning" is supplied by copying in an LF). This objection is that Fox's story crucially requires a deletion view of VP ellipsis. Note that the story makes sense only under the assumption that the second conjunct in (15) does indeed have a derivational stage where it is represented as the full S in (17) - after all, it is the derivation of the full S which is being blocked. Under the Supply a Meaning view, for example, Fox's story has no analogue. Here the syntax simply derives *Lucille did* with the meaning did'(P)(I) and the value of P is contextually supplied. Since the first conjunct can have as its meaning the VP shown in D₂ in (11), then this meaning is available to be picked up as value of P (similar remarks would hold if an LF is copied in).

But there are also more theory-neutral reasons to be suspicious of the transderivational explanation. The main point that I will try to motivate below is that the pattern is actually quite complex: the Sag effect arises from a confluence of factors, and if we tease apart these factors then we see that the general pattern favoring NSO interpretations is not the expected pattern under Fox's explanation. Thus, note first that this explanation for the difference between (14) - (16) and (18) - (20) relies crucially on the different semantics of existentially quantified NPs and proper names (and definite descriptions) - the latter are such that D₁ and D₂ yield the same meaning. But this difference does not actually appear to be the relevant parameter. Notice that, for example, the WSO reading (of the first conjunct) is indeed quite odd in (15) if our main interest is in the actions of Lucille as a character in the discourse context. But consider a scenario in which Lucille happens to be the Chief Resident at some hospital, and where our interest is not in her, but in the quality of patient care in that hospital on a given day rather than in the doings of Lucille per se. Under this scenario, (15) is perfectly alright with the WSO reading for the first conjunct:

(15) A nurse examined every patient, and Lucille did too.

Points along these lines are made in Shieber, Pereira and Dalrymple (1996) and in recent discussion by Hah on LINGUIST List (1996).

Before continuing, I should point out that Fox is aware of this fact: he claims that his generalization holds only when the first conjunct contains an individual level predicate rather than a stage level predicate. (I will instead use the terms *stative* vs. *event* predicates to distinguish these.) The difference, he argues, has to do with the addition of an existentially quantified event variable in the case of the event predicates, but not in the case of the stative predicates. (I return to this in Sec. 3.x.) Thus Fox is careful to use primarily sentences like (16), which involve stative predicates. And it is indeed true that the effect is more robust for statives - (16) resists the WSO reading (of the first conjunct) much more than does (15). Nonetheless, the fact that even in the case of event predicates there is (without additional context) a strong preference for the NSO reading might lead us to conclude that the two cases are not essentially different. In other words, I would suggest that whatever accounts for the strong (out-of-context) preference for the NSO reading in (15) can hopefully generalize to the case of stative predicates as in (16). Moreover, it seems to me that even in the case of statives, one can construct a context to bring out the WSO reading.¹⁸ Suppose that in order to be considered a successful film festival, each film in the festival must be nominated for an Oscar. Suppose further that in order to be nominated for an Oscar, each film must be admired by at least one critic from *The New York Times* and at least one from *The New Yorker*. Moreover, at the moment in question, Pauline Kael happens to be the only critic for *The New Yorker*. Speaker A asserts that she believes that the film festival is quite successful - each film will indeed be nominated for an Oscar. Speaker B disagrees. Speaker A can then respond:

(22) No I really think each film will indeed be nominated for an Oscar. After all, at least one critic from *The New York Times* admires every film, and Pauline Kael does too.

Notice too that if we replace *Pauline Kael* in (22) with *the only critic from The New Yorker* (or *the chief critic from The New Yorker*) it becomes even better. (I find this even easier if the second conjunct is replaced with *and so does Pauline Kael* and, similarly, *and so does the only critic from The New Yorker*; I have no idea why this should be so.)¹⁹

Third, Fox notes a very interesting asymmetry in these. Notice that if the antecedent (the non-VPE) sentence contains the proper name or definite NP in subject position, then the effect is lessened - the second conjunct can have the WSO reading with greater ease:

(23) Lucille examined every patient, and a nurse did too.

(24) Pauline Kael admires every movie, and at least one critic from the *Times* does too.

(Of course WSO is always more difficult than NSO, but in (23) - (24) the WSO reading is much more accessible than in (14) - (16).) On the face of it, this appears deadly for the transderivational story - for the transderivational principle should also block D_2 for the first conjunct in (23) and (24). And yet, under almost any account of VPE, D_2 must indeed be possible for the first conjunct. The reasoning here is that since the second conjunct in, for example, (23) can have the WSO reading, this means that the "missing", "deleted", or "picked up" VP has as its meaning *examine₄(every-patient')*. But then this must also be the meaning of the VP in the first conjunct. (I have phrased this in the terms of the scope account assumed here, but analogous remarks would hold under a view where LF identity is crucial rather than meaning identity, and analogous remarks would hold under other views of WSO readings.)

Fox's explanation for this asymmetry relies on cyclicity. The essential idea assumes that the second conjunct is subordinate to the first, and therefore in an earlier cycle. Thus the checking of the competing derivations for the second conjunct in (23) and (24) happens at a point in the derivation where we are considering only the second conjunct. It has no access to the first conjunct, since that will be on a higher cycle. But now consider the case where the tables are turned. On the higher cycle, the first conjunct does have access to the derivation of the second conjunct, and the requirement that the two VPs must be identical forces D_2 for the first conjunct. In other words, although D_1 for the first is simpler than D_2 and yields the same meaning for that conjunct alone, it is not a competitor here because it itself is ill-formed. The reason is that it will not yield a VP whose meaning is the same as that of the second conjunct.

But this explanation is highly dubious, for we find the same asymmetry even when the VP Ellipsis is in a separate sentence from the target. Indeed, the asymmetry appears even across speakers in a discourse. Thus compare the discourses in (25) to those in (26):

(25) Speaker A: At least one doctor examined every patient. Speaker B: Lucille did too.

Speaker A: At least one critic from the *Times* admires every movie. Speaker B:

Pauline Kael does too.

(26) Speaker A. Lucille examined every patient. Speaker B: At least one doctor did too.

Speaker A. Pauline Kael admires every movie. Speaker B: At least one critic from

The Times does too.

In (25) it remains extremely difficult to get WSO for Speaker A's sentence, but it is easier to get this reading for Speaker B's sentences in (26). However, I know of no view of cyclicity which would allow the derivation of the first S in such discourses to "access" the derivation of the second S, and which also would not allow the derivation of the second S to access the first. We would need to extend the notion of cyclicity to provide cyclicity for an entire discourse. While this may not be impossible, suffice it to say that no such view has, to my knowledge, ever been put forth. I return to the asymmetry and to further problems with the transderivational analysis in Sec. 3.5, but first let us pause to consider the implications of these facts for a processing and/or production reanalysis.

3.4. Can this be reformulated in terms of processing/production principles?

Before continuing, let me stop at this point to pick up one issue which has been left hanging until now. Sec. 2 of this paper argued against transderivationality in the grammar - but suggested that some cases appearing to warrant this device could be the result of the fact that the speaker and/or hearer is driven to the shortest route, quite possibly by computing more than one in parallel via a race model. We can now ask whether the case at hand could conceivably be reanalyzed along these lines. That is, could Fox's basic proposal be maintained - but recast in production or perception terms? In fact, I think it is quite likely that perceptual

principles to give rise to the strong preference for NSO in general, but the question here is whether we can tell a perception/production based story for why the NSO reading is so much harder in the VPE context. Interestingly, the asymmetry shown in (25) and (26) suggest that the answer is "no" - these facts are equally problematic for a production/perception translation of Fox's proposal.

To elucidate, let us phrase this in terms which are neutral between an encoding and a decoding explanation. Recall, moreover, that the basic proposal crucially assumes a level of representation where the second conjunct has its full VP representation - and the "action" takes place in the pairing of this representation with a meaning. Hence, let us imagine the following two discourses, where the second conjunct in both cases is given in its more abstract representation (with a full VP):

(27) Discourse 1. A. At least one critic admires every movie. B. Pauline Kael admires every movie (too).

Discourse 2. C. Pauline Kael admires every movie. D. At least one critic admires every movie (too).

Our interest, then, is in the computation (either encoding or decoding) of B versus the computation of C. Now in order to give a perceptual and/or production explanation which mimics Fox's, we have to assume that the encoder and/or decoder has access to two different derivations which correspond (in some way) to the two grammatical derivations D_1 and D_2 - we will call these D_1' and D_2' . Moreover, such an explanation will have to assume that D_1' involves fewer steps. There are two further facts to keep in mind. One is that some principle ensures that the meaning (or LF) of the VP in B is the same as that of A, while this principle also ensures identity between C and D. And second - and most important - recall that if we are placing the burden on an encoder or decoder, *then the computation takes place in real time*.

And that is exactly the rub: the real-time order of the two sentences in each discourse is the opposite of the cyclic order of the corresponding sentences in (15) and (23). In other words, consider the computation of C. Since the system has no way to know that this sentence will be

followed by D, we would have to assume that in the initial computation, the encoding or decoding derivation is D_1' and not D_2' . This means that in the computation of D, we should only be able to get the derivation which pairs this with the NSO meaning. Since this is not the case, we must conclude that the system can at this point go back and recompute C. But then we are left with a mystery regarding B. Again, we might assume that the system is initially forced to compute B via D_1' . But since parallelism forces the VP meaning of B to be the same as that of A (and since A by itself can have the WSO meaning), why can't B then be recomputed? The bottom line is that since the real time order of the two sentences is the reverse of their cyclic order, it is difficult to see any motivated encoding or decoding story which "mimics" the cyclic explanation.

3.5. More problems with transderivationality

The asymmetry between (25) and (26), then, is not easily explained by a consideration of which S has access to which other (whether cast in grammatical terms or in processing/production terms). This in turn casts doubt on the transderivational principle, as it should block D_2 for the first S in (25) as well as for the elided S in (26). What then is responsible for this asymmetry? The obvious hypothesis is that the asymmetry is a fact about which sentence is full and which is elided - WSO (of the other conjunct) is not possible if the sentence with the proper name contains the ellipsis site, but is more accessible if the proper name subject occurs with a full VP.²⁰ The hypothesis that the location of the ellipsis site is the relevant parameter gains confirmation from the following:

(28) At least one critic from the *Times* admires every movie. Pauline Kael does too, and at least one critic from the *Post* does too.

I find it just as difficult to get the "missing" reading in (28) (as do several informants whom I consulted). The cyclic account fails to predict this. The *Pauline Kael* sentence should have the missing derivation D_2 (and hence so should the first S) because the third S can have D_2 . (Note that the second conjunct in the second sentence here (the *Post* conjunct) should be able to have D_2 because this results in a different meaning from D_1 .) But if we assume - as does Fox - that a

sentence in a higher cycle can access a lower cycle (cf. Fox's explanation for the ambiguity in (24)) then the *Pauline Kael* sentence in (28) should have both derivations. This in turn means that the first sentence in (28) should allow WSO just as easily as it would if this sentence were in isolation. But this prediction is incorrect: it is as difficult to get WSO for this sentence as it is for (16).

Yet another observation casting doubt on the transderivational account is that a completely analogous pattern holds when there is a negative in the VP Ellipsis sentence - but here the scoping of the object over the negative does affect the truth conditions. Thus note first that the WSO reading is possible in a sentence in isolation as in (29):

(29) *Pauline Kael doesn't like every movie.*

Of course we cannot determine the scope of the object with respect to *Pauline Kael*; by WSO, then, I mean the scoping of the object with respect to the negation. But now consider what happens in a corresponding VP Ellipsis sentence:

- (30) a. *At least one critic from *The Times* likes every movie. But Pauline Kael doesn't.*
 b. *Bill likes every movie. But Pauline Kael doesn't.*

In both cases in the second sentence, it is very more to get the reading where the (understood) object has scope over the negative (the meaning where for every movie, Pauline Kael dislikes it). The transderivational analysis of the earlier facts, however, is silent on these. Recall again that Fox's story crucially assumes that what is under consideration is the analysis of the full S *Pauline Kael doesn't like every movie*. We have already seen that when this surfaces with a full VP (as in (29)) then WSO is possible. This is not surprising, since this is not logically equivalent to the NSO reading. But then this should also be allowed for the second conjunct in (30). Once again, the crucial parameter seems to be whether or not we have VP Ellipsis.²¹

Strikingly, the general pattern here is the same as we have seen above. For example, this reading is also odd in (31) in a context where the focus of our interest is Lucille - but under

the scenario where Lucille is the Chief Resident and our interest is in patient care, then the WSO reading is much more natural:

(31) At least one nurse examined every patient. But Lucille didn't.

Similarly, Hirschbuhler's *Canadian flag* example easily allows the WSO reading (in both sentences):

(32) An American flag flies over every stadium. But a Canadian flag doesn't.

Finally, in support of the fact that "referentiality" is at issue here rather than scope interactions, consider the following variant of Hirschbuhler's example:

(33) A/The American flag flies over every stadium. The Canadian flag does too.

Here *the Canadian flag* is not existentially quantified, but is instead a definite description referring to a type (or, kind), not a token. Whether a definite description refers to a token or a type, it will not show scope interactions with a universally quantified NP. In other words, (34) should yield logical equivalence under both derivations, and so D₂ should be blocked:

(34) The Canadian flag flies over every stadium.

And yet, (33) is as natural as Hirschbuhler's original sentence.

In sum, then, the effect first noted by Sag seems to depend on the complex interaction of several factors. First, it is more noticeable if the subject of the VP Ellipsis sentence is more "referential". Second, it is more noticeable if the predicate is a stative one. Third, the surprising absence of the WSØ reading shows up far less when the sentence with the "referential" subject contains a full rather than an elided VP. We now turn briefly to the sketch of an alternative way to view the effect.

3.6. An alternative explanation

The intuition that I want to pursue here is that scoping a universally quantified object over a proper name (or definite NP) subject does in fact result in a different meaning from the NSO meaning (the meaning difference is sufficiently subtle that we are not normally aware of

an obvious ambiguity here). The basic idea is that there are always other existentially quantified variables involved (time variables and/or perhaps event variables). If we further assume that these are always bound "lower" than the subject, then it follows that scoping the object over the subject will necessarily effect the relative scope of the quantifier binding these variables with respect to the object - yielding a different meaning for NSO and WSO (regardless of what the subject is). The hypothesis, then, is that the WSO meaning is a rather unnatural one with a highly "referential" subject (especially with a stative predicate), although it can also be made a bit more accessible when the universally quantified object NP is overt (hence the asymmetry). (It should be noted that Fox pursues a similar line for some of the apparent counterexamples to his generalization but - quite crucially - he does not extend this to the case of statives, and thus uses the latter as key evidence for the transderivational principle.)

Now to make the above remarks a bit more concrete. As mentioned above, there are at least two obvious sources for an additional existentially quantified variable which will show scope interactions with the object. One is to assume that all sentences contain existentially quantified event variables. The other does not rely on event variables, but rather takes the scope effects to always be the result of the object interacting with an existentially quantified time (or, interval) variable. I will pursue the first option here (although there of course also can be time variables, but these will be ignored); but will leave open the possibility that the effect can ultimately reduced to the interaction of quantifiers with tense - hence removing the necessity for event variables.

Consider then a sentence like (35):

(35) At least one doctor examined every patient.

There are six logically possible scope relations, which I represent schematically below:

- (36)
- a. one doctor, x [event, e [every patient, p [x examined p at e]]
 - b. one doctor, x [every patient, p [event, e [x examined p at e]]
 - c. event, e [one doctor, x [every patient, p [x examined p at e]]
 - d. event, e [every patient, p [one doctor, x [x examined p at e]]
 - e. every patient, p [event, e [one doctor, x [x examined p at e]]
 - f. every patient, p [one doctor, x [event, e [x examined p at e]]

Of course just how we would get these various scope relations depends on the account of scope in general, and on how it interacts with the binding of the event variable. The interaction of quantifier scopes with event variables is, in fact, potentially quite complex (see, e.g. Bayer, 1996 for thorough discussion) and is thus far beyond the scope of this paper. Suffice it to say here that if we take the general line on WSO discussed in Sec. 3.1, the idea will be that the meaning of the VP is such that the full Ss can ultimately show various scope effects. Note further that (a) and (c) are equivalent as are (e) and (f) - leaving us with four possible meanings: (i) (a) and (c) = single doctor, single event ; (ii) (b) = single doctor, multiple events (b); (iii) (d) = multiple doctors, single event; (iv) (e) and (f) = multiple doctors, multiple events.

Let us make two additional assumptions. The first is that while a single event can be complex (and have multiple subevents as its parts), a single complex event is not the same model-theoretic object as multiple events. But note that the distinction is a subtle one - and I would suggest that it is precisely because it is difficult to differentiate multiple events from a single complex event that it is not always easy to determine the full range of relevant meanings. This assumption would ultimately require further justification, but we can note that it is analogous to the distinction between "groups" and "sums" in the literature on plurals (see, e.g., Landman, 1989) and that the latter distinction does appear to have independent motivation. The second crucial assumption is, at this point, purely stipulative: this is that the semantic combinatorics works in such a way that the event variable is necessarily bound below the subject. In other words, the scope relations shown in (c), (d), and (e) are all impossible.²² (This stipulation is, admittedly, rather unpleasant, but it may well be that if the account is

recast in terms of time variables then this would follow from the fact that tense is "packaged" within the VP.) Crucially, then, this rules out the meaning in (d) - where there are multiple doctors and a single event.²³ The claim, then, is that scoping the quantifier over the subject necessarily also scopes it over the event quantifier - giving rise to a multiple event rather than a single event reading.

With this much in mind, let us return to (15):

(15) At least one doctor examined every patient, and Lucille did too.

Again we assume that the missing VP here must have a meaning identical to that of the VP in the first conjunct (whether VP Ellipsis involves deletion, or the supplying of a meaning). The first VP in turn can have one of two meanings - one ultimately gives rise to the multiple doctor/multiple event scenario (i.e., the scopes in (f)) and one to the single doctor/single event scenario (the scopes in (a)). Thus if the missing VP is understood as the first (which is the NSO reading) we again have a single event; if it is understood as the second (WSO reading) we have multiple events. In other words, the difference in scope of the object vis-a-vis the subject does affect the meaning - even when the subject is a proper name.

Now recall that when our interest is in the activities of Lucille it is difficult to get the WSO reading - but this is much easier when our interest is in patient care. Suppose that when we are talking about some character in a discourse situation - and our interest is in the doings of that character - we are generally expecting some statement about their participation in a single event. (In fact, this need not always be the case, as long as the discourse functions associated with VPE favor this.) Then it follows that the multiple event reading will be unnatural - and this in turn causes a heavy preference for the single event (and hence single doctor) reading for the first conjunct. But if our interest is in patient care, then there would be no particular reason to look for a single event in which Lucille participated - and hence the multiple event (and thus WSO) reading is much more natural.

Fox actually suggests something along these lines for the case of event predicates²⁴ - he too points out that these can contain an additional existential quantifier over events which

will interact with the scopes. But he also points out that the effect is far more robust in the case of statives (as in the *Pauline Kael* sentences), and he attributes this to the lack of an event variable in these. As was shown earlier, however, here too it is possible with enough context to bring out the WSO reading for the first conjunct. Moreover, I have argued above that the general pattern is the same for both events and states - making it likely that the same basic mechanism is at work in both cases. Thus we can take the term "event" in its broadest way to include not only events (in the usual sense) but also states, and assume that there is an existentially quantified variable (perhaps over states) in the stative sentences as well. (This is the view in, e.g., Parson, 1990.) But this leaves us with the question of why the effect is so much more robust with statives than with event predicates. It does not seem unreasonable here to assume that it is even more unnatural to access multiple states when our interest is in a single character. In other words, the WSO reading of the first conjunct in (16) is one which forces not only multiple critics but multiple states of affairs:

(16) At least one critic from *The Times* admires every movie, and Pauline Kael does too.

Thus if the first conjunct has WSO, the ultimate understanding of the second conjunct is one in which we are saying something about multiple states involving *Pauline Kael* - arguably a highly unnatural reading. But if we construct a context in which we are not really interested in the predilections of *Pauline Kael* as a focus of our interest but rather in her role as a potential nominator for an Oscar, the multiple states reading is more natural.

Finally, we have yet to provide an account for the asymmetry discussed earlier - the WSO reading (for the second conjunct) is much more accessible if it contains the "non-referential" NP and the first conjunct contains the referential one:

(24) Pauline Kael admires every movie, and at least one critic from *The Times* does too.

There are two possible lines to pursue here. One would have to do with the pragmatics and discourse conditions surrounding VP Ellipsis - roughly this line would say that these conditions interact with "referential" subjects in such a way as to enhance the expectation that we are talking about a single event (or state) in the second conjunct. The second possibility would

have to do with the presence or absence of overt material which individuates the events or states - when the universally quantified NP is overt it is easier to access the reading in which the events or states are multiple (and are individuated by the object). Either of these lines (if they could be fleshed out in fuller detail) would predict that the WSO meaning (multiple states) is more natural for the first conjunct in (24) than for the second in (26), and would thus account for this asymmetry.

Needless to say, the above account is sketchy at best, and contains several promissory notes. Nonetheless, it is hopeful that something along these lines can be fleshed out. The important point, then, is that the effect appears to be due to a combination of factors ("referentiality" of the subject; stativity of the predicate; overtness of the quantifier; and the fact that WSO is always harder) and we find the effect to greater and lesser degrees. My hypothesis, then, is that the effect is due to a subtle interaction of the fine-grained semantics of the sentences (they contain additional existentially quantified variables and hence WSO is always different from NSO) with discourse considerations. If this is correct, then this case provides no evidence for economy-driven transderivationality in the grammar. In any case, though, we have seen in the previous sections that the transderivational explanation fails to account for the full range of facts.

To conclude, I have argued that economy-driven transderivationality in the grammar is highly suspect. For some of the cases that appear to motivate these, a processing and/or production explanation would appear to be in order. In other cases, a closer look at the facts suggest that the transderivational explanation simply does not slice up the pie correctly.

Footnotes

* For helpful discussion of the material in this paper, I would like to thank to Tanya Reinhart, Rolf Noyer, Mark Johnson, Michael Rosen, and the audience at the Ohio State University conference on "The Limits of Syntax". Much of this paper started life as a discussion of the paper given by Reinhart at the "Limits" conference, and some of the issues herein are discussed by her (although in a rather different light) both in that paper and in Reinhart (1995). Finally, I would especially like to thank Chris Barker for reading a draft of this and providing extremely illuminating comments on incredibly short notice.

¹ This assumes that the set is finite. But even if it is infinite, if it can be ordered in some way such that there is a "simplest" element then one could equally easily locate the *n*th simplest element.

² When I use the term "the grammar" I mean the grammar of any given language. In other words, the question here is really: For any language, does its grammar contain transderivational rules and/or principles? If the answer were yes, then we could ask the further question as to whether these are language particular or are, instead, general principles in UG. I will, however, claim that the answer is no - and so the distinction between language-particular principles and those given by UG need not concern us here.

³ Strictly speaking, this remark is not quite true: if the output of the system in question is sent to some other system, one might imagine a scenario whereby the first system chooses the simplest alternative in order to make the job of the next system easier. In particular, one might imagine that the grammar chooses the simplest derivation out of some candidate set in order to make the job of the processing system and/or the production system easier. But this kind of scenario makes sense only if we embrace two further assumptions: (1) the output of the grammar (i.e., a set of phonetic form/meaning pairs) is "sent" in some straightforward way to the encoder and/or decoder, and (2) that this state of affairs (i.e., that the grammar incorporates a mechanism for choosing the simplest derivation) arose from some sort of evolutionary pressure.

This, in fact, seems to be roughly the position of Chomsky (1995) - the notion that the grammar is the "perfect" computational system for its interfaces with other cognitive systems embodies some assumption along the lines of (1). Both of the above assumptions are possible, but since we understand so little at this point either about the evolution of grammatical mechanisms and about the interplay between the grammar and the encoder/decoder, the assumptions in (1) and (2) are not obviously warranted ones. See Johnson and Lappin (to appear) for detailed discussion of these points. See also fn. 7.

⁴ This definition is not always the one which has figured in to specific proposals. Sometimes the notion of a "simpler" derivation assumes that certain operations are more costly than others, and/or assumes that certain applications of a rule will be more costly than other applications of the same rule (cf., "shortest move", Chomsky, 1995). However, unless there is some independent criterion according to which some rule (or some application of a rule) is more costly than another rule (or another way in which it could have applied), then the claim that it is always the "simplest" derivation which wins is circular.

In some cases, however, it appears that one can break down the competing rules (or, rule applications) into substeps according to which one has more substeps than the other. But in any case this will not affect the basic point at hand - I will ultimately suggest that the simpler derivation is one whose corresponding encoding and/or decoding takes less time (in real time), and so a "more costly" rule (or a more costly application of one rule) can be taken to correspond to something which - in the corresponding encoding and/or decoding - also takes more time.

⁵ It may strike the reader that this is nothing more than the Derivational Theory of Complexity. However, I am not claiming here that there always needs to be a direct correlation between the length of grammatical derivations pairing $\langle \cdot, M \rangle$ and the length of a decoding derivation mapping \cdot to M . We need only claim that this correlation exists in those situations where Case A effects are found.

⁶ The issues are further compounded by several facts. First, it seems clear that Case A effects are not always found - there do seem to be sentences which are genuinely ambiguous (it is conceivable all that such cases involve decoding derivations of equal length, but this seems unlikely). A second and related point is that - as discussed in fn. 5 - there is no reason to assume that there is always a correlation between the length of grammatical derivations and the length of decoding ones; the decoder might well have strategies available to it which change this correlation. (This may explain the absence of Case A effects in certain instances.)

⁷ Chris Barker (personal communication) raises the following question about the line of reasoning here. I am assuming that Case A effects arise in those situations in which there is a correlation between the number of steps in the competing grammatical derivations and those in the corresponding processing derivations. For example, I am assuming that in the Russian case, the processor would have to go through more steps to pair (2b) with OVS than with SVO, although the grammar itself allows both pairings. Given this, why not assume that indeed there are actual transderivational principles in the grammar (and that in this case the derivation pairing (2b) with OVS is indeed blocked), where these have become grammaticized in order to simplify the job of the processor? This question is similar to the issues raised above in fn. 3 - is it possible that there are indeed transderivational processes in the grammar per se, and that these have been incorporated into the grammar because of their function in simplifying the job of the encoder and/or decoder?

This is a question which I obviously cannot supply a definitive answer to, but the notion that certain functionally-based principles have become "grammaticized" is one which I think should be approached with caution. The difficulty with this notion is that it seems to be based on some kind of loose analogy with evolutionary principles of biological systems - and "grammars" themselves are not biological organisms. Of course grammars are rooted in (or, located in) biological organisms - and so if it could be shown that this grammaticization took place at the level of the evolution of the species then it is possible that UG would contain a

transderivational mechanism. As noted in fn. 3, though, this kind of explanation for the existence of grammatical transderivational principles also makes sense only if one assumes that the results of the grammar are "sent" in some straightforward way to the processor. Moreover, it is crucial to assume that the transderivational computation which rules out the more complex derivation is not sent in any way to the processor. But I see no obvious way to conceive of the relationship between the grammar and the processor such that the grammar has access to competing derivations, contains a principle to "block out" some of these, and then sends the result to the parser.

But the most important point to note about this kind of reasoning is that in any case where one processing (or production) derivation is more complex than a competing one, then there simply is no need to posit a transderivational mechanism in the grammar. In other words, the assumption that the processing/production explanation is the sole reason why we do not seem to get the more complex derivation is the null hypothesis. To posit that the grammar contains an additional principle which blocks the competing derivation is a more complex explanation of the phenomena.

⁸ In fact, Fox and Reinhart claim that all instances of transderivationality in the grammar are of this type; although they also allow for the subcase where α_1 and α_2 happen to be the same, which is the case described above under Case C.

⁹ The only theory which does not have the consequence that Passives involve an extra step is one which claims that actives and passives are both listed in the lexicon and that speakers simply know both forms separately. The discussion below assumes that this is incorrect. It further assumes that in this case there is indeed a straightforward correlation between the complexity of grammatical derivations and the complexity of the corresponding encoding derivation.

This assumption is not universally accepted; some theories do in fact posit some Passive rule in the grammar, but assume further that the processing and/or production models

are such that passives have been "precomputed". Under such a view, then, there is no correspondence in this case between the complexity of the grammatical derivations (Passives are more complex according to the grammar) and the complexity of the encoding/decoding derivations (Passives are not more complex according to the latter). I am assuming that this "have your cake (rule) and eat it too (don't use it)" is incorrect in this case.

¹⁰ Note, though, that Shlonsky (1992) phrases this in terms of an economy-driven transderivationality principle in the grammar.

¹¹ I am glossing over one important question in this discussion: in this view and in the Deletion/Meaning Identity view there is no obvious reason why the meaning must (in general) be supplied by some other overt VP rather than by any contextually salient property. In other words, why is this (in general) an instance of what Hankamer and Sag (1976) called "Surface Anaphora" rather than their "Deep Anaphora"? In fact, however, the issue is somewhat complex: it is well-known that one can find instances where the "missing property" is just some contextually salient one (whose meaning is not that of any overt, surface VP); but it also appears to be true that there is a strong preference for an overt VP to supply the missing meaning. I will not attempt to address these issues here.

¹² Again, I stress that neither of these is Fox's actual view; his view is the theory in (2). But his points can be recast without any damage by recasting his view into (3), and doing so removes certain irrelevant complicating factors from the discussion.

¹³ Not all accounts actually do have this property. Consider, for example, an account in which wide scope object readings are the result of Quantifier Raising of the object, while in the narrow scope reading the object is interpreted "in situ". But what does it mean to say that it is interpreted in situ? Suppose that *read* is listed in the lexicon with a meaning of type $\langle e, \langle e, t \rangle \rangle$ (that is, it expects an individual in object position), but that it can get a generalized quantifier in object position by undergoing a type-shifting rule mapping this to a new meaning of

type $\langle\langle\langle e,t\rangle,t\rangle,\langle e,t\rangle\rangle$. (This type shift rule is detailed below in (9).) Then both the wide scope and the narrow scope object interpretation involve additional steps.

¹⁴ Thus the final system that I am adopting here is a kind of odd and arguably unnatural mix of Partee and Rooth (1983) and Montague (1974). Recall that Partee and Rooth (1983) are assuming that ordinary transitive verbs like *read* are listed in the lexicon as having their "lowest type" meaning - i.e., they are of type $\langle e,\langle e,t\rangle\rangle$, and can take a generalized quantifier (rather than an individual) in object position only via argument lift on the object position. To be consistent with this basic idea, Partee and Rooth further assume that proper names (and ordinary definite NPs) also are "born" with their lowest type meanings, and hence *Lucille* just denotes the individual \underline{l} . It can, however, type-raise to denote the generalized quantifier $\underline{P[P(\underline{l})]}$. It turns out, however, that by allowing *Lucille* to denote either \underline{l} or $\underline{P[P(\underline{l})]}$, we introduce yet additional possible derivations which simply serves to complicate the exposition. Thus although the Partee and Rooth proposal is actually much more natural than the kind of hybrid I am considering here, I do this simply to ease the exposition and ignore additional possible derivations.

¹⁵ Sag's actual account relied on a stipulation in his definition of "alphabetic variant". It is worth noting that in many other accounts this stipulation has no natural analogue (this will be discussed immediately below, but see also Partee and Bach (1981) for relevant discussion), but it is also worth noting that even in Sag's account this stipulation was simply extra and followed from nothing else. Thus - if the generalization really were as Sag thought it was - this would actually be unexpected under any theory. As we will see below, however, the generalization is incorrect.

¹⁶ Fox takes great pains to argue against this view, and unfortunately space precludes a thorough treatment here of all of his relevant data. I will, however, be presenting some new data in support of this.

¹⁷ Fox's paper is a detailed and closely argued one, and unfortunately I do not have the space to give it full justice. Some of the objections that I raise below are, in fact, anticipated in his paper, and I will try to answer the major counterarguments that he raises, but it would take additional space to thoroughly discuss all of his points.

¹⁸ Johnson and Lappin (to appear) provide the following example involving a stative predicate, where the WSO reading is quite natural (and is the preferred reading)

- (i) At least one natural number other than one divides every prime number, and one does too.

¹⁹ There is a potential confound in this example (and in some of the examples discussed above) in that it may be marginally possible to get a reading where the overt quantifier in the first conjunct is scoped over *and*. To remove this confound, we can consider cases where the antecedent VP and the VP Ellipsis site are actually in separate sentences (the examples from hereon in are constructed so as to remove this confound). Even here, though, I believe that - with enough context - the WSO reading is possible in the first conjunct. Thus consider the following discourse in the context given above:

- (i) Speaker A. No, I don't think that each film will be nominated for an Oscar. It's just not enough that least one critic from *The Times* admires every film.

Speaker B. True, but Pauline Kael does too.

(Again, this is better with *so does Pauline Kael*).

²⁰ Fox anticipates this objection, and constructs a rather complex case to answer it (his (83)). I agree with his judgments for the crucial case and have no explanation of it - except to note that it is sufficiently complex that a number of other parameters may be interfering here. But whatever is the explanation, it is worth noting that it involves event predicates - and so even under Fox's analysis the judgments are, in fact, not actually accounted for.

²¹ Fox briefly discusses the interaction of VPE with negation: he claims that the following is ambiguous, where both conjuncts can have the WSO reading:

(i) Ken Hale doesn't speak more than three languages, and Rob Pensalfini doesn't as well. But I find that this too strongly favors NSO (more strongly than does the first conjunct in isolation), and to the extent that WSO is possible, there is a strong tendency to take it on reading where the object quantifier has scope over *and* (and hence it is the same languages which both of them fail to speak). Note that, as discussed in fn. 19, the possibility of scoping the object over the conjunct should disappear if the two are in separate sentences (as in my examples in (30)), and once we do that I believe that it becomes extremely difficult to get a WSO reading:

(ii) Ken Hale doesn't speak more than three languages. Rob Pensalfini doesn't as well. In any case, though, the issue is not whether these readings are completely impossible (I claim that with enough context they always can emerge), but whether the pattern here is the same as in the cases where we have a universally quantified NP in object position and a proper name in subject position. The fact that WSO is much more difficult in the examples in (30) than in, e.g., (29) supports the claim that the basic pattern has nothing to do with logical equivalence.

²² Actually, there is another assumption which would work just as well for the story I will tell: this is that the subject and event binding must be "together". Either view has the effect of ruling out (d), which is all that is crucial here.

²³ I am assuming in this discussion that the absence of this meaning is the result of the way that the semantic combinatorics work, although I have not said just how they would work so as to rule this out (nor how the other meanings can arise). However, as far as I can tell it is not crucial that (d) be completely impossible - if there is some reason that it is unnatural or difficult to access then the story to be told below could probably still be maintained.

Note further that since the scope relations in (d) are not logically equivalent to any of the others we ought, in fact, be able to test whether this really is an impossible meaning for (35). But this is difficult to determine - precisely because of the above observation that the

distinction between a single complex event and multiple events is a subtle one. It is thus difficult to differentiate (d) from the multiple doctors/multiple events reading.

²⁴ Although Fox's account is not identical to this one; in particular, he assumes that the meaning in (d) is possible.

References

- Bayer, S. (1996). *Confessions of a Lapsed Neo-Davidsonian*. Ph.D. Dissertation, Brown University, Providence.
- Chomsky, N. (1995). *The Minimalist Program*. Cambridge: MIT Press.
- Cormack, A. (1984). "VP Anaphora: Variables and Scope", in F. Landman and F. Veltman (eds.), *Varieties of Formal Semantics*. Dordrecht: Foris.
- Dalrymple, M., S. Shieber, and F. Pereira (1991). "Ellipsis and Higher-Order Unification", *Linguistics and Philosophy* **14**, 399-452.
- Diesing, M. (1992). *Indefinites*. Cambridge: MIT Press.
- Fox, D. (1995). "Economy and Scope". *Natural Language Semantics* **3**, 283-341.
- Frazier, L. (1985). "Syntactic Complexity", in D. Dowty, L. Karttunen, and A. Zwicky (eds.), *Natural Language Parsing*. Cambridge: Cambridge University Press, 129-189.
- Frazier, L. and J.D. Fodor (1978). "The Sausage Machine: a new two-stage parsing model". *Cognition* **6**, 291-325.
- Hah, E.-C. (1996). "Summary: Ambiguous or Unambiguous", *LINGUIST List*, Vol. 7 - 1419.
- Hankamer, J. and I. Sag (1976). "Deep and Surface Anaphora", *Linguistic Inquiry* **7**, 391-428.
- Hendriks, H. (1987). "Type change in semantics: the scope of quantification and coordination", in E. Klein and J. van Benthem (eds.), *Categories, Polymorphism, and Unification*, Edinburgh: Centre for Cognitive Science / Amsterdam: University of Amsterdam Institute for Language, Logic, and Information.
- Hirschbuhler, P. (1982). "VP-Deletion and Across-the-Board Quantifier Scope", in J. Pustejovsky and P. Sells (eds.), *Proceedings of NELS 12*. Amherst: University of Massachusetts GLSA, 132-139.
- Jackendoff, R. (1972). *Semantic Interpretation in a Generative Grammar*. Cambridge: MIT Press.

- Johnson, D. and S. Lappin (to appear). "A Critique of the Minimalist Program", to appear in *Linguistics and Philosophy*.
- Kempson, R. and A. Cormack (1983). "Type Lifting Rules and VP Anaphora", in M. Barlow, D. Flickinger, and M. Westcoat (eds.), *Proceedings of the West Coast Conference on Formal Linguistics 2*. Stanford: Stanford Linguistics Association, 140-152.
- Landman, F. (1989). "Groups I", *Linguistics and Philosophy* **12**, 550-605.
- Lakoff, G. (1972). "On Generative Semantics" in D. Steinberg and L. Jakobovits (eds.), *Semantics*. Cambridge: Cambridge University Press.
- Lappin, S. (1996). "The Interpretation of Ellipsis", in S. Lappin (ed.), *The Handbook of Contemporary Semantic Theory*. Oxford: Blackwell, 145-175.
- Maling, J. and A. Zaenen (1982). "A Phrase Structure Account of Scandinavian Extraction Phenomena", in P. Jacobson and G.K. Pullum (eds.), *The Nature of Syntactic Representation*. Dordrecht: Reidel, 229-283.
- May, R. (1977). *The Grammar of Quantification*. Ph.D. Dissertation, MIT. Cambridge, MA.
- McCawley, J. (1971). "Where do Noun Phrases Come from?", in R. Jacobs and P. Rosenbaum (eds.), *Readings in English Transformational Grammar*. Waltham: Ginn & Co., 166-183.
- Montague, R. (1974). "The Proper Treatment of Quantification in Ordinary English", in R. Thomason (ed.), *Formal Philosophy*. New Haven: Yale University Press.
- Parsons, T. (1990). *Events in the Semantics of English*. Cambridge: MIT Press.
- Partee, B. and E. Bach (1981). "Quantification, Pronouns, and VP Anaphora" in J. Groenendijk, T. Janssen, and M. Stokhof (eds.), *Formal Methods in the Study of Language: Proceedings of the Third Amsterdam Colloquium*, Amsterdam: Mathematisch Centrum, 445-482.
- Partee, B. and M. Rooth (1983). "Generalized Conjunction and Type Ambiguity", in R. Baurle, C. Schwarze, and A. von Stechow (eds.), *Meaning, Use and Interpretation of Language*, Berlin: de Gruyter, 361-383.

Perlmutter, D. (1972). "Evidence for Transderivational Rules." unpublished ms., Brandeis.

Postal, P. (1971). *Crossover Phenomena*. New York: Holt, Rinehart and Winston, Inc.

Reinhart, T. (1995). *Interface Strategies*. OTS Working Papers, Research Institute for Language and Speech, Utrecht University.

Ross, J.R. (1967). *Constraints on Variables in Syntax*. Ph.D. Dissertation, MIT. Cambridge, MA.

Sag, I. (1976). "A Logical Theory of Verb Phrase Deletion", in S. Mufwene, C. Walker, and S. Steever (eds.), *Papers from the Twelfth Regional Meeting Chicago Linguistic Society*. Chicago: Chicago Linguistic Society, 523-532.

Shieber, S., F. Pereira, and M. Dalrymple (1996). "Interactions of Scope and Ellipsis", *Linguistics and Philosophy* **19**, 527-552.

Shlonsky, U. (1992). "Resumptive Pronouns as a Last Resort", *Linguistic Inquiry* **23**, 443-468.

Williams, E. (1977). "Discourse and Logical Form", *Linguistic Inquiry* **8**, 101-139.

