0. Introductory remarks. I assume that every serious theory of language must give some explicit account of the relationship between expressions in the language described and expressions in some interpreted language which spells out the semantics of the language.¹ Let's call this relationship the translation relation. Theories differ as to how this relation is specified. In the Aspects theory of syntax, taken together with a Katz-Postal view of "semantic rules" (Chomsky 1965; Katz and Postal, 1964), it was assumed that the relation was defined on deep structures. Serious problems led to modifications of this view in several directions ('generative' and 'interpretive' semantics). Chomsky's latest papers (1976, 1975a: Ch. 3) assume that a modified 'intermediate structure' (surface structure with traces) is transformed by rules of interpretation to a level of representation called 'logical form' (which is input to further rules). Common to all of these approaches is the assumption that the translation rules are defined initially on syntactic structures of one sort or another (sometimes from several 'levels' at once). Let us call this assumption the configurational hypothesis: the translation rules all have the form: [183]

1. Given a structure of such and such a form, translate the structure into an expression in the interpreted language of such and such a form.

In sharp contrast to this view of the translation relation is another, in which it is assumed that the syntax and the translation go hand in hand. More precisely, the syntax builds up a syntactic structure by the application of syntactic rules which operate on syntactic structures, while for each syntactic rule there is a unique translation rule specifying the translation of the resultant expression as a function of the translation of the parts. Let us call this hypothesis the rule-to-rule assumption about the translation relation. It is characteristic of Montague grammar as well as a number of other approaches all falling within the limits of what T. Parsons (class lectures, U. Mass., 1976) has called 'recursive grammar.' Montague himself (in Montague 1973) described the translation relation as operating on structures, the so-called 'analysis trees'; but since the analysis trees are simply a record of the
syntactic derivation it is easy to conceive of his fragments as exemplars of a rule-to-rule procedure which simultaneously builds up a well-formed syntactic structure and its translation into an intensional logic. (See Montague, 1974, and Partee, 1976a for examples and extensions of Montague's work.)

The earliest work within the transformational-generative framework on the translation problem was of a mixed character. In Katz and Fodor (1972), the Type 1 projection rules were configurational, the Type 2 rules were rule-dependent. Although few explicit rules were given, it was clear from the examples that the procedure was thought of as one in which the translations of kernel structures were given "bottom to top". The contributions of the transformations to the meaning of sentences were associated with the rules and were functions of the application of the rule to a particular analysis of a syntactic structure and the already specified translation of that structure.

In Bach 1977 a preliminary study of the syntax of a version of classical transformational theory as modified by Fillmore (1963) was carried out. I drew the conclusion that the earlier theory had been too hastily abandoned, probably because of the appeal of the Kate-Postal hypothesis, and that it was an attractive alternative to various more abstract theories that have gained the attention of most transformational grammarians since Aspects. The present paper is a continuation of the earlier one, but a departure from it in that it attempts to deal with the translation relation within the framework of classical transformational grammar. I will try to show that a rule-to-rule conception of this relation fits much better with a revised version of the older theory than with the standard theory and its descendants, and I shall also try to show that the resulting system offers certain advantages over a strictly 'Montague grammar' framework. I draw heavily on the work of Cooper and Parsons (1976).

1. The translation relation. I have claimed that the two views of the translation relation are fundamentally different. In this section I will establish this claim by giving a couple of examples, one showing that the configurational view can get results that the rule-to-rule view cannot and one the converse.

Consider the problem of the correct structure for restrictive relative clauses in English (or other similar languages). Partee (1976b) has argued that a compositional view of the translation procedure is inconsistent with an analysis such as that given in A, but must choose the analysis B, since the translation of the definite NP in A will necessarily be incorrect for the translation of the top NP.

```
A. NP      B. NP
   /\       /\  \\
  / \  /   / \       / \
 /   \   /   \     /   \```

1. The translation relation. I have claimed that the two views of the translation relation are fundamentally different. In this section I will establish this claim by giving a couple of examples, one showing that the configurational view can get results that the rule-to-rule view cannot and one the converse.

Consider the problem of the correct structure for restrictive relative clauses in English (or other similar languages). Partee (1976b) has argued that a compositional view of the translation procedure is inconsistent with an analysis such as that given in A, but must choose the analysis B, since the translation of the definite NP in A will necessarily be incorrect for the translation of the top NP.

```
A. NP      B. NP
   /\       /\  \\
  / \  /   / \       / \
 /   \   /   \     /   \```
But this is true only within a rule-to-rule conception of the translation relation. Without additional assumptions, there is nothing in the nature of this hypothesis to prevent 'waiting' to give a translation for the whole NP in \( A \) until the top NP, that is, there is no inherent limitation in the hypothesis to prevent stating rules of the configurational sort for structures of arbitrary complexity. On the other hand, a rule-to-rule hypothesis must fix the translation of the constituents as they are put together. [186]

For the opposite kind of example, there is nothing in the rule-to-rule hypothesis as formulated here to prevent positing different syntactic rules that map the same syntactic input into the same syntactic output but are each associated with different translation rules, although this is excluded in Montague's general theory, according to Cooper (personal communication).

Thus the two approaches are essentially different and incomparable. Hence, we cannot give any a priori arguments for one or the other in terms of relative strictness.

In the following I will adopt a strong form of the rule-by-rule hypothesis:

II.

a. For every syntactic rule, there is a unique translation rule specifying the translation of the output of the rule as a function of the translation(s) of the input(s).

b. All rules apply strictly locally in a derivation, that is, no rule has access to earlier or later stages of a derivation.

c. Syntactic rules apply to syntactic structures; translation rules to (already built-up) translations. Neither type of rule has access to the representations of the other type except at the point where a translation rule corresponding to a given syntactic rule is applied. \( \tilde{y} \)

d. The translations of the inputs to a syntactic rule must be 'intact' in the translation of the output (except possibly for changes in the variables, to
avoid accidental binding of variables).

(For these and other possible constraints on syntactic and semantic rules, see Partee, 1976c). [187]

(IIa) is simply a restatement of the rule–by–rule conception of the translation relation.
(IIb) is a consequence of this hypothesis; it amounts to saying that there is no level of
representation like the T–markers of the classical theory, or the analysis trees of
Montague.4 (IIc) requires a kind of 'autonomy' of a different sort from that discussed in
Chomsky (1975c). (IId) is a kind of 'recoverability' condition on the translation relation.
It disallows, for example, rules which map some part of the translation into a constant
(including null). Taken together with external requirements of descriptive adequacy it
leads to many of the consequences of the recoverability condition of transformational
grammar. For example, if there is a rule freely deleting NP's, the translation of the
result will necessarily preserve the translation of the NP, and the resultant sentence
had better be infinitely ambiguous or else the grammar will be wrong.

2. Classical transformational grammar. The earliest versions of transformational
grammar (CTG, Chomsky, 1957; 1975b (1955)) differed in a number of respects from
the theory of Aspects and its progeny. Most important for our purposes was the
assumption that complex sentences were built up by so-called generalized
transformations that embed the transform of a sentence into a matrix. This aspect of
the theory contributed a great deal to the relatively 'concrete' nature of the theory as
compared to later versions. The most straightforward way to incorporate this idea in
the present frame–work is to assume that the kernel rules generate some
structures that contain variables for various kinds of structures. Consider for example
simple embeddings of that–clauses. Assume that among the elements of the lexicon
are variables over sentences: say, that0, that1, that2, ... Then the embedding rule will
take a structure like that underlying (4) and embed into it a second sentence like (3):

3. John says that.
4. everyone is here

The rule is in effect a rule schema (as in Montague's quantification rules) looking like
this:

\[
\begin{align*}
T-\text{that} & \quad S^1: X, \text{that}_i, Y \quad \quad \text{(4)} \\
\text{1} & \quad \text{2} & \quad \text{3} \quad \text{===>} \quad \text{1, 2 + 4, 3} \\
S^2 & \quad Z \quad \quad \text{(3)} \\
\text{4} & \\
\end{align*}
\]
The variable is then a reinterpretation and slight modification of the 'dummy' symbols that were used in later versions of CTG (Fillmore, 1963). The associated translation rule is extremely simple; it applies the result of forming the lambda-abstract over (the translation of) that$_i$ of $S_1$ to the translation of the second sentence. A similar treatment can be given for verb-phrase embeddings, and relative clauses. (See Delacruz, 1976, for a treatment of propositional level constructions in Montague grammar.)

3. **Quantification.** One of the rocks on which the Katz–Postal hypothesis foundered was the problem presented by pronouns and the scope [189] of quantifiers like every, some, many (including the scope of negative elements). There have been Innumerable attempts to solve such problems, none completely successful. Linguists of the generative semantics persuasion tended to construct relatively abstract underlying structures in which relative scopes were structurally represented (Bach, 1968; McCawley, 1968, 1970). Interpretivists (Jackendoff, 1972) adopted rules that operated on surface, near-surface, or end-of-cycle structures assigning coreference indices (or non-coreference indices). The latter approach has led to the hypothesis currently being developed by Chomsky (1975a, 1976, forthcoming) that quantification and coreference relations are determined by rules that operate on structures which contain traces (indexed empty nodes) left behind by "moved" NP's. As I noted above, common to all of these approaches is the assumption (I) above: the translation relation is defined on some set of structures. I believe that a good deal of the argumentation about these alternatives is artifactual and that a rule-to-rule approach, implicit in the classical theory, explicit in Montague's work, provides an elegant alternative for treating the problems of quantification.

Before getting into the details of such an analysis, however, I would like to mention two errors that crept into a good deal of the early work on pronouns.

The first error is the assumption that all pronouns that are somehow 'coreferential' to antecedents are to be accounted for in the same way. The untenability of this assumption was pointed out [190] in Partee, 1972. Most linguists and philosophers of language would now agree that the assumption is incorrect. In the following I will be dealing mostly with a relatively narrow subset of such relations, namely, those corresponding pretty closely to the use of bound variables in logic. Because of the semantic difference between quantified noun-phrases and proper names, many conclusions about anaphoric relations have been wrong, since the two were confounded.

The second error (closely related to the first) is to be found in the work of many linguists who have advocated the use of variables in underlying structures (Bach, 1968;
McCawley, 1970). It is the assumption that, if there are variables in underlying structures, then every full noun-phrase that appears in a sentence has gotten there by substitution for a variable. I will show below that dropping this assumption makes it possible to give an explanation for certain facts about bound-variable pronouns.

Montague's paper 'The proper treatment of quantification in English' (1973, reprinted with some editorial corrections in Montague, 1974 — henceforth PTQ) is an elegant treatment of pronouns-as-bound-variables in English. It exploits the rule-to-rule conception of the translation relation. The idea, very briefly and roughly, is the following. A given English sentence like (5) can be derived in a number of different ways:

5. Every woman is seeking a unicorn.

(I'm replacing Montague's simple present by progressive forms through—[191]out out.)

I'll use the following Informal notation (borrowed from PTQ in part): let \( \alpha' \) represent the translation of \( \alpha \) into the intensional logic; let \( x ) (...x...) \) etc. represent the translation which results from quantifying in every woman etc. in the place of \( x \) in the (open) sentence (...x...). Syntactically the rule simply substitutes the noun-phrase for the first occurrence of the syntactic correspondent to \( a \) and changes all other occurrences of that variable (in PTQ, he, i) to pronouns. Notice that the different scope relations are not represented syntactically at all (in the configurational sense). Rather, the idea is that scope differences are the result of different orders of application of the same rules. If represented at all, the differences are represented in a record of the different derivations (in PTQ, different analysis trees). This kind of representation is exactly what was captured in the T-markers of the classical transformational grammar (as has been noted by Partee, 1975a) but never exploited for this kind of problem. Note that all NP's can be directly generated in place or 'quantified in.'

Taking the one derivation in which the two NP's of (5) are directly generated and the various possibilities for quantifying in, we get derivations corresponding to the following five representations (I'm ignoring two others that result from the rule which quantifies into verb-phrases in PTQ):

5. a. every' woman' is' seeking' a' unicorn'
   b. (every' woman' \( x \) ( \( x \) is' seeking' a' unicorn')
   c. (a' unicorn'y ) (every' woman' is' seeking' y) [192]
   d. (every' woman' \( x \))(a' unicorn' \( y \) (x is' seeking' \( y \))
   ) (a* unicorn' \( y \)) (x is' seeking' \( y \))
   e. (a' unicorn' \( y \)) (every' woman' \( x \) ( \( x \) is'
   seeking' \( y \))

3.1 Extending CTG. What is the difference between ECTG and Montague grammar? One
difference is this: in Montague grammar (even with transformational extensions, as in Partee (1976b), there is no separation between various kinds of rules, not only is there no extrinsic ordering, there is no ordering imposed by a separation of rules into various components. In CTG the rules are still segregated into types and the simple (singulary) transformations are extrinsically ordered. Fillmore (1963) assumes something like this (ignoring conjoining transformations, which pose a separate problem) (Fig. 1): [193]

```

```

"matrix"

```

"constituent"

```

Pre-sentence structures

```

Final simple transformations (FST)

```
[194] The ordering assumptions are represented by the arrows. Note in particular that every embedding rule embeds a (derived) structure into a kernel structure (in our adaptation, in place of a variable).

Notice that the cyclicity of derivations follows from this hypothesis about the organization of a grammar. But it is still necessary to make an additional assumption to prevent some bad derivations. Let's assume that the structures to be analysed for the operation of the PST must be the maximal ones available, that is, that no PST can apply to an embedded structure (this amounts to a kind of strict cyclicity in the sense of Chomsky, 1973).

If, as I assume, the PST are extrinsically ordered, then without some such restriction it would be possible to construct derivations in which a crucial ordering of two rules were violated. In an embedded sentence we could fail to apply a certain rule on the first pass, then apply it on the second pass when the matrix structure was undergoing the PST. I will use this restriction crucially below.

Where do our new NP-embedding rules fit into this schema? Since NP's embedded into embedded sentences can have wide-scope interpretations, it seems that we cannot assume that the matrix for an NP-embedding must be a kernel structure. Moreover, if we wish to incorporate L. Karttunen's analysis of questions (Karttunen, 1977) as arising by embedding of wh-phrases we must allow this operation to apply into embedded sentences. The following sentences illustrate these facts: [195]

6. Every man thinks that some boy on our block is harassing him. (possible wide scope on some boy) 7. Who did you think would be here?

So let's assume that the NP rules embed NP's into pre-sentence (derived) structures. The system then looks like this (Fig.2):
3.2 The proper treatment of quantification in ECTG. In this section I will sketch a treatment of quantification within our revised version of the classical theory. I'll show how we can take over Montague's treatment in PTQ more or less intact. I'll then (Section 4.1) show how it is possible to incorporate an explanation for some further facts about quantification and its relation to negation in the framework of the classical theory in a way that is not easily reproducible in MG. In Section 4.2, I'll take up some interactions between this theory and the classical treatment of embedded structures. Finally (Section 4.3), I'll present an explanation for some facts that is possible only within a theory which allows NP's to appear in sentences either by direct generation or by quantification. Recall that I am dealing primarily with pronoun–antecedent relations of the sort that most closely resemble bound variable relations in the predicate calculus.

Our assumptions about the translation relation (II) and the separation of syntactic and translation rules require that such pronouns be represented as variables in kernel structures. The alternative, to start with pronouns as such, requires the addition of rules of indexing and is not available within the restrictions of (II) (see Cooper and Parsons, 1976, for examples of such rules). I'll assume that gender is syntactic and that the variables have the form of indexed pronouns, indices disjoint for the three genders: she\(_{0}\), she\(_{3}\),..., she\(_{n}\); he\(_{1}\), he\(_{4}\),..., he\(_{n+1}\),..., it\(_{2}\),..., it\(_{5}\),..., it\(_{n+2}\),.... The pronouns are translated into sets of properties: pro\(_{n}\) = \(\overline{P}\{x_{n}\}\), as in PTQ. [For typographical convenience I use a macro rather than a cap over P, EB 2007]

Montague's quantification rules include rules which seek out the first of a series of (zero or more) indexed pronoun*, substitute the NP for that pronoun and remove the
subscripts from all other pronouns bearing the same subscript. Rather than building the desubscripting function into the quantification and other NP-embedding rules, I am including among the PST a rule which does this desubscripting operation before the embedding rules can have their effect:

T-pro:

\[
\begin{array}{cccc}
X, \text{ pron}_n, & Y, \text{ pron}_n, & Z \\
1 & 2 & 3 & 4 & 5 \\
1 & 2 & 3 & \text{ pro} & 5
\end{array}
\]

Conditions: obligatory

2 must command 4

(Strictly speaking, this rule is not a transformation since its elementary operation changes a terminal element by removing part of it. It is easy to set up a representation which does not violate the restrictions on transformations, but I think rather premature.) I assume that this rule applies iteratively (alternatively it could be a rule schema which applies 'across-the-board'). Like its interpretive analogue (Jackendoff, 1972) the rule must follow all PST that involve movement. The quantification rule, relative clause formation, the question rule, and various embedding rules [198] like EQUI are formulated for syntactic structures containing subscripted pronouns. For example, the quantification rule looks like this:

Because T-pro is in the PST, the quantification rule will always embed into the left-most position of a structure containing a string of originally like-subscripted pronouns. Thus quantifying phrases will always stand to the left of their bound-variable pronouns. The command condition on T-pro ensures that binding can occur down into but not up out of embedded clauses:

8. Every man said that he was happy.
9. *That every man is happy disturbs him.

(I use underlining to indicate the binding relation between an antecedent and its bound-variable pronouns.) Note that proper names, which I assume to be 'rigid designators' as in PTQ (same for all possible worlds), can occur in contexts linked to pronouns where quantified phrases can't. Thus sentences like (10) receive their interpretation by some other -- possibly discourse-level, or pragmatic -- rules

10. That John is happy disturbs him.

[199]

The desubscripting rule is a syntactic rule. Its corresponding translation rule is the identity function so that the rule has no semantic effect. The translation rule
associated with quantification can be taken over directly from PTQ (assuming the same translations for noun phrases (term-phrases). It applies the translation of the NP (the set of properties that every man has in example (8)) to the intension of the appropriate lambda abstract of the matrix (the set of $x_i$ such that $x_i$ said that $x_i$ was happy).

Sentence (8) would have this derivation (for the bound-variable reading). The Kernel Rules would derive three structures:

\[
\begin{align*}
S_1: & \text{ he}_4 \text{ said that}_1 \\
S_2: & \text{ he}_4 \text{ was happy} \\
\text{NP:} & \text{ every man}
\end{align*}
\]

$T$-that embeds $S_2$ into $S_1$, giving

\text{he}_4 \text{ said that he}_4 \text{ was happy}

$T$-pro removes the subscript from the second $\text{he}_4$:

\text{he}_4 \text{ said that he was happy}

Finally, $T$-quantification substitutes every man man for $\text{he}_4$, yielding (8) with its correct translation. Similarly, the five different derivations of example (5) arise as follows:

\[
\begin{align*}
5a. & \text{ every woman is seeking a unicorn (kernel sentence)} \\
5b. & \text{ every woman + she}_0 \text{ is seeking a unicorn} \\
5c. & \text{ a unicorn + every woman is seeking } \text{it}_2 \\
5d. & \text{ every woman = (a unicorn + she}_0 \text{ is seeking } \text{it}_2 )[200] \\
5e. & \text{ a unicorn + (every woman + she}_0 \text{ is seeking } \text{it}_2)
\end{align*}
\]

(The plus sign signifies the application of the quantification rule.) These derivations give exactly the same translations as do the corresponding derivations in PTQ.

4.1 Negation and quantification. In general although Montague's fragment is quite strict in generating only well-formed English sentences (modulo the details of English syntax such as reflexivization that Montague was not particularly interested in), it overgenerates possible pairings of sentences and translations quite a bit and is hence descriptively inadequate. It is quite difficult to tell in individual cases whether these inadequacies are matters of principle or of descriptive detail.\textsuperscript{8} Given the power of the general theory it is probably impossible to show that any particular analysis cannot be reproduced within his framework. In this section, I will exhibit some cases where it can
be argued that a mode of explanation not available in Montague grammar can be used to account for certain facts of English grammar in a general and elegant way.

PTQ provides only one way of negating a sentence. Thus, corresponding to Sentence (5) above there is only one negative version:

11. Every woman isn’t seeking a unicorn.

(Again, I’m using progressive forms for Montague’s simple present; also, I’m ignoring forms like is not.) The translation rules (again ignoring VP quantification) give these interpretations: [201]

11. a. (NOT (every’ woman’ is’ seeking’ a’ unicorn’))
   b. (every’ woman’ x) NOT (x is’ seeking’ a’ unicorn’)
   c. (a’ unicorn’ y) NOT (every’ woman’ is’ seeking’ y)
   d. (every’ woman’ x) (a’ unicorn’ y) NOT (x is’ seeking’ y)
   e. (a’ unicorn’ y) (every’ woman’ x) NOT (x is’ seeking’ y)

These readings can be paraphrased as follows:

11. a’. It’s not the case that every woman is seeking a unicorn.
    b’. Every woman is such that she’s not seeking a unicorn.
    c’. There’s a unicorn such that it’s not the case that every woman is seeking it.
    d’. Every woman is such that there’s a unicorn that she’s not seeking.
    e’. There’s a unicorn that is such that every woman is such that she isn’t seeking it.

In my judgment, only readings (b), (d), and (e) are natural for (11).

Let’s now recall Klima’s analysis of negation (1964), in its essentials this: Negative sentences in underlying form arise by the optional generation of a sentence initial NEG element. The negative element is positioned within the Aux by a rule of NEG-placement. There is a rule of NEG-incorporation which attaches the negative to certain noun phrases which are marked +Indefinite (some/any, many, a, every etc.). Klima assumed that the latter two rules applied in the order just given and that the incorporation rule applied optionally to the right but obligatorily to the left (to subjects). This last assumption was intended to explain the badness of examples like this [202]

12. *Any people didn’t come to the party.

But note that even in the optional case it is necessary to prevent the rule from skipping
over an indefinite element to prevent sentences like (13):

13. *John gave any candy to no friends.

I'll assume that the rules apply in the opposite order and that the same restriction that prevents the derivation of (13) prevents the NEG-placement rule from applying in a case like (12). Thus the negative will remain in initial position if the subject is 'indefinite.' Now note that in the present framework the position of the negative will differ depending on whether the subject NP is (when indefinite) directly generated in the kernel rules or quantified in. In the latter case, since the subject position is occupied by a variable pronoun at the point of application of the PST, NEG-placement (an obligatory rule) can apply. These assumptions lead to the derivation of the following sentences:

11. a''. Not every woman is seeking a unicorn.
   b''. Every woman isn't seeking a unicorn.
   c''. Not every woman is seeking a unicorn.
   d''. Every woman isn't seeking a unicorn.
   e''. Every woman isn't seeking a unicorn.

It seems to me that these sentences come somewhat closer to matching the readings (a)-(e) above than does the single sentence (11).

The treatment suggested here is consistent with the view of transformational grammar that there are abstract underlying forms that get turned into English by obligatory rules and the view [203] of CTG that the transformational rules are ordered. It is inconsistent with the strong well-formedness condition of Montague grammar (Partee, 1976c) and with the view of Montague grammar and extensions of it that reject any ordering of rules, either by types or by extrinsic ordering conditions. I stress again that Montague grammar does not exclude in principle the means for providing a descriptively equivalent account of the facts assumed above. Hence, alternatives must be judged on the basis of the ability to capture generalizations, simplicity, etc. The above account does not depend crucially on the assumptions we have made that exclude reapplication of PST to simple sentences or already embedded ones. But other facts follow from this assumption. We can derive the following sentences (with binding of pronouns as indicated):

14. Every man is seeking a woman who loves him.
15. Every man believes that he is seeking a woman that loves him.

On the assumption that there is a Passive transformation in the PST, we can also derive (16).
16. Every man believes that a woman that loves him will be found by him.

We cannot derive the following sentence:

17. *A woman who loves him is being sought by every man.

In order for every man to bind him it must be quantified in. The NP a woman who loves him could have gotten into the sentence in two 204 ways (1) by direct generation of a woman such that \( r, (r^a \text{ property variable}) \) and the application of the relative clause rule, or (2) (if relative clauses can also be embedded into NP's by themselves, an open question) by quantification of the whole phrase. On either assumption because of the command condition, we would have at the point of the application of quantification for every man, the structure

\[
18. \text{a woman who loves } h_0 \text{ is being sought by } h_0
\]

T–pro cannot apply because of the command condition. The rules so far would yield either (19) or (20):

19. A woman who loves every man is being sought by \( h_0 \).
20. A woman who loves \( h_0 \), is being sought by every man.

I assume that sentences with subscripted variables are ill–formed and that all pronouns other than bound variables arise from directly generated 'real' pronouns, which have a different translation. Although semantically well–formed, the sentences will be syntactically deviant. Notice that the strict–cycle property will exclude the derivation of (17) from (14).

Consider next the problem of restricting the scope of quantified phrases in embedded sentences. The facts are hard to determine here. I indicated above (Example 6) that wide scope is possible out of embedded clauses. Cooper (1975) has provided an example that seems to show that this can happen even out of relative clauses.

21. John wants to date every girl who goes out with a professor who flunked him out of Linguistics 101.

[205]

Our rules co far do not disallow assignment of wide scope to quantifiers in embedded sentences, even complex ones (in this they are like PTQ, see Rodman, 1976, for a method for building in island constraints). But the rule of desubscripting, with its command condition, does disallow the binding of pronouns by such quantified phrases. Here the facts seem to be relatively clear. Compare the following pairs. The (a) sentences are to show that wide–scope interpretations are possible, the (b)
sentences to show that binding of pronouns cannot occur.

22. a. ?Some man makes the promise that he will love every woman.
       b. *Some man makes the promise that he will love every woman to her.
23. a. ?A professor who had dated every student in the class was at the party.
       b. *A professor who had dated every student in the class spoke to him.
(See next section for discussion of some apparent counterexamples, which I will argue do not derive from quantification.)

If I am correct about these facts, then it would seem that the desubscripting rule is not just a weird variant of Montague's treatment (which includes the desubscripting operation in the quantification rules themselves). Montague's rules incorporate the claim that binding of variables and width of scope go hand in hand. The rules given here do not, but they make a different claim, that facts about binding will be the same for all NP-embedding rules. For example, Karttunen's rules for questions, taken together with [206] my rules predict the following facts;

24. *Who did you tell him Mary loves?
25. *Who did you tell that Mary loves him?
26. *Which man that saw which girl will tell about it?

Examples like (24) and (25) are usually taken to be evidence for some kind of cross-over condition (or, in Chomsky 1976) a restriction that variables -- from trace elements -- cannot be antecedents of pronouns to their left). If my judgment of (26) is correct, it provides a different kind of consequence in support of our hypotheses, since there is no movement of the phrase which girl.

Similarly, all of the following will be excluded by our rules:

27. *He loves every man.
28. *She is loved by every woman.
29. *Herself is loved by every woman.
30. *Who did the woman he loved betray.
32. *The man who the woman he loved betrayed is despondent.

(27) and (28) are taken by Lasnik (1976) and others as evidence for the necessity of a rule of 'disjoint reference.' On the assumption that reflexivization is defined on variables and follows Passive (or if there is no rule of Passive), the ungrammaticality of (27) – (29) follow (but without some further stipulation, our rules will generate the counterpart to (29) with herself
and every woman interchanged). (30) – (32) are from Chomsky (1976), where (31) is especially interesting since Chomsky's explanation depends on the assumption that the logical form of (31) (to which the principle applies) is this: 33. (for some x, x a person) (the woman he loved betrayed x ) But this seems to require that the sentence be given wide scope interpretation of someone and a different explanation must be provided for parallel sentences with narrow scope readings (Chomsky does not discuss problems of opacity and narrow scope). On the assumption that ordinary restrictive relatives arise by embedding sentences with (subscripted) variables in them, (32) will be excluded.

4.2 Verb–phrase embedding. As mentioned above, PTQ provides not only rules for quantification into sentences, but also rules for quantifying into verb–phrases (and common–noun phrases). The main reason for this addition, as I understand it, is to provide sources for sentences like (34):

34. John wants to catch a fish and eat it.

(with narrow scope on a fish). The classical theory of transformational grammar accounted for complement sentences of all sorts by embedding of transforms of full sentences. One such rule (for EQUI type sentences) might be the following: [208]

\[
\begin{align*}
\text{T–to:} \\
\text{Matrix S:} & \quad X, \text{NP}, \text{Aux}, V, \text{to}, q_n, Y \\
& \quad 1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \\
\text{Const. S:} & \quad \text{pro}_n, \text{Pres}, \text{PredPhrase} \\
& \quad 8 \ 9 \ 10
\end{align*}
\]

where 2 and 8 agree in gender

Structural change: substitute 10 for 6

I interpret $q_n$ as a variable over [properties of EB 2007] properties. The associated translation rule applies the lambda abstract over $q_n$ of the translation of $S_1$ to the abstract (over $x_m$) of $S_2$. There are a number of interesting possibilities opened up by this rule.

First, notice that an extension of the use of $q$ variables makes it possible to treat VP ellipsis as a result neither of deletion or surface Interpretation, but simply the result of the fact that variables for VP's are not pronounced in English (just as, in Japanese, for example, bound–variable pronouns are not pronounced). For example, we can derive a sentence like (35) as follows
(assuming we have solved the problem of conjunction):

35. John wants to marry a Swede and Bill wants to too.
   $S_1$ John wants to $q_0$ and Bill wants to $q_0$
   $S_2$ he$_1$ Pres marry a Swede
   or $S_2$ he$_1$ Pres marry she$_0$

By T–to we can derive (35) with two readings, which seems correct. On the second reading (wide scope for a Swede, specific reading), quantification takes place into the structure:

36. John wants to marry she$_0$ and Bill wants to($q_0$)

[209] Note that although there is no occurrence of she$_0$ in the syntactic structure of the right conjunct, there is an occurrence of the corresponding variable in the translation of (36), whether or not we have eliminated the second $q_0$. The reduction rule that distributes the lambda abstract of $S_2$ over the translation of $S_1$ will ensure that the individual concept variable appears in just the right places. This treatment predicts the fact noted by many writers that reduced phrases interpreted like full phrases always share the relevant readings of the full phrases. Facts about 'sloppy identity' also follow from this treatment (as they do from other adaptations of lambda abstraction for deriving verb–phrases).

Consider sentences like (37): [2007: this is really baaaad stuff??]

37. Mary kissed her husband and Alice did too.

The different readings arise from the two sources for the embedded verb–phrase (assuming either a more general version of T–to or a different VP–embedding rule):

i. she$_0$ Past kiss she$_0$'s husband
ii. she$_0$ Past kiss she$_3$’s husband

The latter — non–sloppy — reading arises by the following derivation

$S_1$: she$_3$ Past $q_0$ and Alice Past ($q_0$)
$S_2$: she$_3$ Past she$_3$’s and Alice Past ($q_0$)
T–pro: she$_3$ Past kiss she's husband and Alice Past ($q_0$)

Notice that Mary must come into (37) by quantification, otherwise the
subscript could never be removed from she in (ii). From our [210] analysis we can now predict that when the first subject of a sentence like (37) must have been directly generated we will only be able to derive such sentences with a sloppy reading, since the only way in which a directly generated subject can bind a pro-noun is by lambda abstraction (VP-embedding). According to our hypotheses of the last section, sentences with negatives left at the head must come from direct generation of 'indefinite' subjects. Hence we should predict a difference in the interpretation of sentences like (38) and (39):

38. Every woman from New York didn't kiss her husband, but Alice did.
39. Not every woman from New York kissed her husband, but Alice did.

I think that (39) requires 'sloppy identity' that is, it cannot be asserting that Alice kissed some of the husbands of women from New York. A somewhat clearer case is the following:

40. No woman over thirty kissed her husband, but several under thirty did.

Examples like those we have been discussing point up the necessity within -- either our framework or the framework of PTQ -- of including in the system something like Partee's derived verb phrase rule. Without some such rule it is impossible to get a directly generated noun-phrase to bind a pronoun in the predicate. That is, neither our previous rules nor the rules of PTQ account for sentences like (41), or interpretations of sentences like (42) in which the subject has narrow scope from direct generation:\[12\]

41. No man loves himself.
42. Not every person loves his children.

Finally, notice that the assumption of VP embedding makes it unnecessary to have a separate rule of quantification into VP's. Montague's VP-quantification allows two further derivations for example (11) 11. Every woman isn't seeking a unicorn.

11. f. NOT (every' woman' x) (a' unicorn' y) (x is' seeking' y
   g. (every' woman' x) NOT (a' unicorn'y) (x is' seeking' y)
These readings can be derived in our system as follows:

f. NEG every woman Pres q0 + (a unicorn + sheo be Ing seek it2)
(Not every woman...
In the preceding sections I have tried to suggest some ways in which a modified classical transformational framework can be used to get different results from those obtainable in a purely Montague framework. Obviously, much more work remains to be done before any firm conclusions can be drawn about the relative merits of the two systems.

4.3 In support of two sources for quantified noun phrases. In this section I will give evidence in favor of the hypothesis of the preceding sections (shared by ECTG and Montague grammar) that surface noun-phrases can be derived either by direct generation or by quantification. It counts then as evidence against theories in which no noun phrases or all noun phrases arise by embedding rules.

If we assume that bound variable pronouns arise only by quantification (or lambda abstraction, which will be excluded by the nature of the constructions considered here) then if there are constructions which depend on the presence of directly generated NP's we should not be able to find any bound–variable pronouns construed with then. Moreover, the HP's in question should always have narrowest scope. Two such constructions are sentences with there–Insertion and sentences with have plus an indefinite object and a locative (possibly a wider class of have–sentences).

43. There’s a unicorn in the garden.
44. I have a unicorn in my garden.

Carlson (1973) noted that the indefinite NP’s in there sentences do in fact always have narrowest scope:

45. Every dreamer believes that there is a unicorn in his garden.

I believe the same fact holds for the class of have–sentences illustrated.

46. Many people want to have a car in their garage.

Now consider sentences like these:

47. There was a man in his garden.
48. Those people have a baby in its crib.

The most natural interpretation for these sentences is one in which the possessive pronoun is not bound by the respective indefinites. If we force a bound–variable interpretation then the NP–PP [213] must be interpreted as a unit. That is, (47) cannot be interpreted as parallel to (49):

\[ \text{g. (every woman) + NEG she\text{\textsuperscript{0} Pres} + (a unicorn + she\text{\textsuperscript{3} be Ing seek lt\text{\textsubscript{2}}})} \]
49. A man was in his garden.

Similarly, (48) must be interpreted in a way parallel to (50), not (51).

50. What those people have is a baby in its crib. 51. What those people have in its crib is a baby.

If we assume that such sentences are either directly generated as such (there in the kernel) or arise by transformations that require indefinites in their structural conditions, these facts will follow, since to bind the relevant pronouns we would have to have variables in the relevant positions (again, if there–insertion is a transformation this explanation depends crucially on the ordering of rules and the strict cycle principle).

Wasow (1975) has given a series of arguments against theories in which all full NP's arise by embedding into variable positions. Among his examples are the following (his numberings are given after the examples):

52. A man who discovered that there were some burglars in his house was shot by them. [12]
53. *Some burglars shot a man who discovered that there were they in his house. [11]
54. Some burglars shot a man who discovered that they were in his house.

(53) and (54) are no problem for our theory. (53) cannot be generated for reasons just given. (54) can be generated straightforwardly. But (52) is a problem if the undoubtedly possible anaphoric link between some burglars and them arises by quantification, that is, if them is a genuine example of a bound variable pronoun. In order to account for (52) we would have to give up the hypothesis that there requires directly generated Indefinites. Moreover, that example, as well as a further one, (55), would be counterexamples to the command condition on our desubscripting rule, T-pro.

55. A man who discovered that some burglars were in his house was shot by them. [6b]

Thus, to save our theory, we must show that these examples do not arise by quantification but by some other rule. In fact, I think this is correct.

Many linguists and philosophers agree that there are cases of anaphora that arise by some sort of (pragmatically conditioned?) rule or rules that must be kept separate from true cases of bound-variable interpretations.13 (Probably,
even here a number of cases must be distinguished.) Without a relatively explicit account of such rules, it is impossible to be even remotely sure about our analysis, but I think a number of suggestive arguments can be given to show that (52) and (55) involve some other process.

First, epithetic anaphora like the bastards, seem incompatible with the clearest cases of bound-variables:

55[ sic]. *Everyone believes that the bastard is going to win.
56. * thinks that they will elect the woman.

[215]

57. *Captain Smith was on the old seaman's last legs.

In none of these cases can we interpret the epithet as anaphoric to the subject. If this hypothesis is correct, we should not be able to use an epithet in examples (52) and (53), if they arise by quantification. But we can:

53. A man who discovered that there were some burglars in his house was shot by the villains.
50 Some burglars shot a man who discovered that the villains wore in his house.

Second, bound-variable pronouns cannot be controlled by quantifiers in separate sentences. But compare (60) and (61):

60. A man discovered that there were some burglars in his house. He was shot by them.
61. ?Every member of the committee was there. He had on white suit.

Third, the kind of anaphora we are considering here are governed by a condition that requires that they occur in modally congruent contexts (cf. Karttunen, 1970). If we vary the verb in the relative clause of (52) we find the sentence is deviant unless the second verb is correspondingly varied:

62. ?A man who pretended that there were burglars in his house was shot by them.
63. A man who pretended that there were burglars in his house pretended that he was shot by them.

True bound variables are not subject to such restrictions:

64. Everyone pretended that he would win.
65. No member of the committee pretended that they would elect her.

[216] 66. Captain Smith pretended to be on his last legs.

I conclude that (52) and (54) are not counterexamples to our theory. (Some of Wasow's other arguments are more difficult, for example, the case of clitic pronouns, but at least there-cases show that it can make a difference if only some NP's are introduced by embedding rules.)

5.1. Conclusions. I have tried to show that a rule-by-rule conception of the translation relation can be taken over quite nicely in the earlier theories of transformational grammar. I have also suggested that there are certain advantages to be gained from taking a basic transformational framework and adding to it certain features of Montague grammar as against adding transformational features to a basically MG framework.

In Bach (1968) it was argued that to give an adequate account of scope ambiguities for sentences like many of those discussed here it was necessary to posit a relatively abstract type of deep structure in which elements like seek were represented as try-to-find. The fallacy of this argument, which was modelled on Halle's argument against the taxonomic phoneme, lay in the assumption that for any general account there would have to be a single 'level of representation' on which to define the scope relations. Similar arguments are now being offered to justify a level of 'logical form' (Chomsky, 1976). But the underlying assumption is not necessary. I have tried to show that a different conception of the translation relation offers another dimension of analysis in which it is possible to give a quite general account of such facts, without positing artifactual levels of representation. Which of the ways turns out to be better must, of course, await a great deal of further work.\textsuperscript{14} [218]

Footnotes

\textsuperscript{1}In fact, it is not necessary to assume an intermediate disambiguated language (Cooper, 1975) and it is an open question whether positing such an intermediate language (= 'logical form') is empirically justified*. All transformational theories that I am aware of make this assumption. What is necessary for all adequate theories is that the interpretations of expressions be accounted for in an explicit way.

\textsuperscript{2}I assume that in this theory and its derivatives, the language of semantic markers would receive an interpretation, say by a model-theoretic semantics.
This is strictly true only if we fix a single translation for the phrase the man. In Cooper, 1975, it is shown that a technique developed for the analysis of adjoined relative constructions (as in Hittite) can be used to formulate a consistent compositional semantics for the NP–3 analysis (see Bach and Cooper, 1978).

Even without extrinsic ordering, it is possible to construct examples of bad derivations that result from 'going back down': [219]

i. John bought a popsicle and Mary a cone.
ii. *A popsicle was bought by John and Mary a cone.

Without much conviction; see Cooper, 1975, for an alternative. I'm restricting discussion here to singular pronouns and ignoring the problem of reflexives (as in PTQ).

The treatment here assumes that there is no 'backwards' pronominalization for bound variables, a matter of detail, not principle. If this turns out to be wrong T-pro will have to be correspondingly modified. Cf. Jacobson, in preparation, for some discussion, also Karttunen, 1970, who argues that many cases of apparent backwards pronominalization result from there being sequences of anaphora, both bound as 'discourse referents.

One such case is that of vacuous quantification, freely allowed in PTQ. Thus, the syntactic structure for John saw Mary has an infinite number of translations associated with it by vacuously quantifying it by phrases like a fish, the unicorn that Mary sees, etc. Such examples would seem to be the result of demanding that the syntactic functions be total functions, but it is possible that an analysis could be given which would exclude such pairings and still meet the requirement of total functions.

Some of the readings are more natural if we substitute some for a. The suggested analysis is obviously only a first approximation to a solution (if there is one in the syntax), which will have to take into account intonational facts as well. Partee (personal communication) has suggested that stress and use of some [220] (especially for unmodified 'light' phrases with a) might be required for those readings in which the order of quantification is not parallel to surface order.

For an example of an alternative within a more strictly Montague theory, see Lee, 1974, which includes an extensive analysis of negation, quantification, and their interactions.
Partee (1976a) proposed a derived verb phrase rule to solve certain problems about conjunction and quantification. The idea has been used by Sag (1976) and Williams (forthcoming) within a transformational framework.

This fact emerged from a discussion with Barbara H. Partee.


I wish to thank Barbara Partee for helpful discussion at all stages of the preparation of this paper. Naturally, all errors and confusions are my own.

References


Jacobson, Pauline. Ms. The pronominal pilot meets the variable Mig. [see next]


Note 2007: The terminology "rule-to-rule" has been cited quite often in the literature. The paper re-edited here has been very hard to come by, so it is
hoped that making the paper available will be of some service. It bears, of course, the marks of its time. The substantive points need a lot of rethinking, particularly in the light of developments in generative theories of all sorts since the time when the paper was first published. I hope to take up the issues raised in another place. EB