Nominal Structures and Structural Recursion*

Robert Frank
Department of Linguistics
University of Delaware
46 E. Delaware Avenue
Newark, DE 19716
(302) 831-6809
rfrank@cis.udel.edu

Anthony Kroch
Department of Linguistics
University of Pennsylvania
619 Williams Hall
Philadelphia, PA 19104-6305
(215) 898-3212
kroch@linc.cis.upenn.edu

*The authors wish to thank the following for their helpful comments and suggestions: Michael Hegarty, Aravind Joshi, Young-Suk Lee, Owen Rambow, Beatrice Santorini and three anonymous *Computational Intelligence* reviewers.

Abstract

It is possible with a Tree Adjoining Grammar to reproduce many of the syntactic analyses originally formulated by linguists in transformational terms. To the extent that these analyses are well-motivated empirically, this fact makes TAG interesting for use in developing computational learning and processing models, since the use of other non-transformational formalisms sometimes forces choices of linguistic description different from those ordinarily made by descriptive syntacticians. Thus, using TAG, one can take advantage in the construction of parsers and learners of the computational tractability of a mathematically restrictive formalism without having to reinvent empirical syntax in order to do so. At the same time, TAG analyses are not identical in every detail to their transformational counterparts; and it is interesting to compare them where they diverge. The differences arise because of a fundamental difference in the way that syntactic recursion is treated in the two frameworks. In TAG, recursive structures are generated by composing elementary syntactic objects, with the result that recursion is factored apart from the representation of local syntactic dependencies. By contrast, in transformational grammar, as in many other frameworks, recursive structure and local dependencies are represented together in a single, full representation of a complex sentence. Because of this difference in the treatment of recursion, it often turns out, when TAG is used to emulate a transformational analysis, that the TAG version has advantages, of both elegance and empirical coverage, over the original. This paper is a demonstration in a new empirical domain, that of nominal constructions, of the advantages of TAG-based syntax. By presenting a linguistically detailed account of these constructions and showing the advantages of using TAG to analyze them, we strengthen the case for the use of mathematically constrained and computationally tractable representational systems in competence-based as well as in computational linguistics.

Keywords: Syntax, Nominalizations, Gerunds, Recursion, TAG, Phrase Structure

1 Introduction

Most contemporary theories of natural language syntax have assumed that a phrase structure tree is the foundation of a sentence's representation, but the form of this tree must be heavily constrained, since tree structure *per se* is insufficient to characterize syntactic well-formedness in natural language. One of the major discoveries in the study of syntactic structure, stemming largely from the work of Ross (1967), has been that the restrictions on and deformations of tree structure which the grammars of existing languages exhibit are highly local; that is, specific grammatical relationships hold within small portions of a sentence's phrase structure tree, regardless of the size of the tree as a whole. Much of the comparison of competing syntactic frameworks has focussed, therefore, on how they capture this locality. Some theories use abstractly defined structural relations like government, while others tie locality to the argument structure of predicates.

Tree Adjoining Grammar (TAG) is a formalism in which sentences are generated out of a finite set of small pieces of phrase structure called *elementary trees*. The elementary trees are composed to build larger pieces of tree structure in a constrained way, and the compositional character of phrase structure directly yields an understanding of why grammatical locality exists: Grammatical constraints can only be expressed as restrictions on the structure and combining possibilities of the (boundedly large) elementary trees. Composition of the elementary trees in a TAG derivation uses the two operations of *substitution* and *adjoining*. Substitution combines pieces of phrase structure in a way analogous to the stacking of blocks onto a pile, with one piece of structure sitting atop another. Adjoining, on the other hand, inserts recursive subtrees, i.e. pieces of tree structure which have nodes of a single category label at the top and bottom of their spine, directly into the body of another tree. In this paper, we provide evidence for the linguistic reality of such a compositional derivation of phrase structure, demonstrating that whether a subtree in a phrase structure is or is not recursive determines an important aspect of its syntactic behavior. This result is expected under the TAG view of structure; but under a non-compositional view of phrase structure, it is mysterious.

Focussing for concreteness on the analysis of two types of nominal construction, so-called "derived" nominalizations and gerunds[1], we show that if the two constructions are taken to differ in whether they form recursive structures with an embedding verbal predicate, certain differences in their distributional properties can be explained. Although gerunds and nominalizations have properties in common that have generally led syntacticians to classify them both as nominal constituents, they often do not behave alike. We adduce new types of evidence concerning the interpretation of amount questions, the licensing of parasitic gaps, and the licensing of negative polarity items which distinguish the constructions; and, given this evidence, we show that certain gerunds should be classified as clauses rather than as noun phrases. We then show further how using TAG as the underlying formalism of the theory of grammar allows us to reduce the difference in the distribution of gerunds and nominalizations to the fact that, treated as clauses, gerunds will form a recursive structure with an embedding clause while nominalizations do not.

2 The Problem of Gerunds

The phrase structure and categorial status of gerunds like that in (1) has long been a puzzle.

(1) Ron couldn't recall [$_{\text{GERUND}}$ (his) meeting the sheik]

From the perspective of their internal structure, gerunds share many of the properties of non-finite clauses: they take empty (i.e. PRO[2]) subjects (cf. (2)), their direct objects

[1]Gerunds are *ing* forms, as in (i) below, while nominalizations are nouns related more indirectly to verbs, as in (ii):

(i) Eating pickles is fun.
(ii) Discussion of pickles is not fun.

[2]In transformational grammar, PRO is the phonetically null but structurally active subject of an infinitive, which functions as the external argument of the verb and ordinarily receives its interpretation from a "controller," an overt antecedent higher in the sentence which c-commands the PRO.

receive accusative case [3] (cf. (3)), and they take adverbial modifiers (cf. (4)).

(2) a. Leonard hates [PRO going home]

b. Leonard hates [PRO to go home]

(3) a. Sally hates [kissing him]

b. Sally hates [to kiss him]

(4) a. [totally messing up on the exam] was not our plan

b. [to totally mess up on the exam] was not our plan

In at least the latter two respects, moreover, gerunds differ from lexically derived nominalizations, their closest syntactic relatives.[4] In (5)a we see that derived nominalizations are unable to assign accusative case, requiring a preposition to govern their object and assign it case. The example in (5)b shows further that such nominals cannot be adverbially modified.

(5) a. Leonard hates [discussions *(of) pickles]

b. [*totally/total failure on the exam] was not our plan

On the other hand, in their external distribution, gerunds do seem to behave like nominal expressions and not like clauses. Unlike clauses, but like nominals, they do not undergo extraposition (cf. (6) and (7)), and they do undergo subject–auxiliary inversion when they serve as the subjects of questions (cf. (8)). Also, like nominals, they do not tolerate the movement of a wh- element to phrase-initial position (cf. (9)).

(6) a. * It bothered me [(Bush's) winning the election]

b. It bothered me [that Bush had won the election]

c. * It bothered me [(Bush's) victory in the election]

[3]Of course, in English this case is abstract and appears morphologically only on pronouns. We will be concerned here only with gerunds which assign accusative case to their objects and not those which require *of*-insertion such as (i).

(i) Leonard hates [the eating of pickles]

These *of*-gerunds pattern unambiguously like nominals with respect to all of the diagnostics which we discuss, and we will therefore take them to be structurally identical to nominals.

[4]We will put aside the question of whether nominals without genitive subjects obligatorily have PRO subjects.

(7) a. * I resented it [(Bush's) winning the election]

b. I resented it [that Bush had won the election]

c. * I resented it [(Bush's) victory in the election]

(8) a. Would [(Bush's) winning the election] really bother you?

b. * Would [that Bush won the election] really bother you?

c. Would [(Bush's) victory in the election] really bother you?

(9) a. * Ted knew about [who$_i$ (her) kissing t_i]

b. Ted knew about [who$_i$ she had kissed t_i]

c. * Ted knew about [who$_i$ (her) passion for t_i]

Abney (1987) captures these facts by analyzing the gerund phrase as a verb phrase embedded within a noun phrase[5], aiming thereby to explain why gerunds behave like verbs with respect to their internal structure, and like nouns in their relation to the context in which they are embedded. However, while Abney's treatment may be correct for some cases, gerunds cannot always be nominals externally. At least where they serve as complements to a verbal predicate, the evidence shows that they must be clauses, both internally and externally. As mentioned above, we will examine in detail the interaction of English gerunds with three types of structure-sensitive element: wh-amount quantifiers, negative polarity items, and parasitic gaps. In each of these interactions, gerunds pattern like clauses and contrast with nominalizations, just as they do with respect to the diagnostics in (3) through (5). Yet these additional contrasts do not seem explicable on the basis of an internal structural difference between the two constructions like that postulated by Abney.

We propose to account for the observed differences between gerunds and nominalizations as follows: While maintaining an analysis of nominalizations as noun phrases, we will show that gerunds should be assimilated to the defective clausal complements of verbs like *seem* or *believe*. As such, they lack the complementizer position of a

[5]Strictly speaking, Abney analyzes the gerund as a VP embedded within a Determiner Phrase (DP), based on his treatment, now in wide use, of noun phrases as headed by a functional category, namely, the determiner. This analysis is attractive, because it assigns parallel syntactic representations to determiners in noun phrases and to tense inflection in sentences.

full subordinate clause, and so, in current transformational terminology, are IPs rather than CPs. More specifically, they are headed by the functional category INFL(ection) rather than by COMP(lementizer). Since functional projections share the lexical features of their complement categories, sentences with gerund complements of their verbs will exhibit structural recursion, having one [+V,–N] phrase embedded below another [+V,–N] phrase. Nominalizations, on the other hand, will be [–V,+N], and therefore will not produce recursive structures when they appear as the complements of verbs. This difference has no obvious significance for a transformational analysis; but, as we will see, if we state our grammatical analysis of these constructions in the formalism of Tree Adjoining Grammar (Joshi 1985) and use, with some refinements, the analysis of wh-movement in Kroch (1989b), we can derive the differences between nominalizations and gerunds from differences in the formal operations by which they are incorporated into a syntactic derivation: gerunds through simple TAG adjoining, and nominalizations through substitution and multi-component adjoining.

3 Scope of Amount Quantification

Cinque (1990) and Longobardi (1990) discuss the interaction between the possible interpretations of quantificational expressions and transformational movement. Cinque points out, in particular, that the possible interpretations of amount questions (AQ) using the wh-quantifier *how many* differ according to whether extraction is from the CP complement of a bridge verb[6] or from an indirect question. Thus, in (10)a below, under the predominant reading, only a number is at issue and no pre-existing set of books is presupposed by the questioner. In (10)b, on the other hand, the only interpretation available takes the questioner as presupposing a pre-existing set of books. Cinque labels these readings the "quantificational" and "referential" interpretations of amount questions.

(10) a. How many books$_i$ did they decide [to publish t_i]? (ref: OK, qua nt: OK)

[6] A bridge verb is a verb like *think* or *say*, which freely allows the extraction of any questioned element, complement or adjunct, from its complement clause.

b. How many books$_i$ did they decide [whether to publish t_i]? (ref: OK, quant: *)

The same difference in interpretation appears when we contrast extractions out of noun complements, as in (11)a, with corresponding bridge verb cases, as in (11)b.

(11) a. How many cities$_i$ did they make [a plan to attack t_i]? (ref: OK, quant: *)
b. How many cities$_i$ did they plan [to attack t_i]? (ref: OK, quant: OK)

It appears that the absence of the quantificational reading of an AQ is associated with subjacency-violating (so-called "long") movement out of syntactic islands. This association is expected if we follow Longobardi (1990) in saying that the quantificational reading of an AQ occurs when it is interpreted[7] at a position in which it takes scope over only the lower clause. Note that associated with any AQ is a presupposition of the existence of some set which the speaker is asking the size of. The referential reading, then, corresponds to the case where an existential quantifier over this set takes wide scope, namely scope over the matrix predicate. Thus, the referential reading for (10)a corresponds to the following "logical paraphrase":

(12) What is the number y such that there is a set of books x with cardinality y such that they decided to publish x?

The quantificational interpretation, on the other hand, corresponds to a representation in which the existential quantifier takes narrow scope; that is, where it falls within the scope of the matrix predicate:

(13) What is the number y such that they decided that there should be a set of books x with cardinality y such that they publish x?

Under a transformational analysis, such narrow scope is possible because the derivation of a sentence like (10)a involves cyclic movement of the wh- phrase, first to the front of

[7]Such interpretation occurs at L(ogical) F(orm) in a transformational grammar.

the embedded clause (to the Spec(ifier) of CP) and then on to the matrix. The movements leave traces, so that later interpretation has available the lower operator position for the determination of scope. In indirect questions like (10)b, the lower Spec(CP) position is filled by a wh- phrase so that cyclic movement is not possible. Instead the matrix wh- moves in one "long" step to its surface position, yielding slightly degraded acceptability. Since there is no intermediate trace position to which scope reconstruction can take place, only matrix scope is possible here. In the nominal complement case in (11)a, it is plausible to suppose that the deverbal nominal head *plan* will not license a trace in the Spec(CP) position of its complement[8] and so only the long movement derivation, and hence wide scope, is possible.[9]

When we extend the AQ diagnostic to extraction from adjectival small clauses[10] or noun phrases, we see that they behave like the long movement cases.

(14) a. How many cities$_i$ did Hitler insist on [the destruction of t_i]? (ref: OK, quant: *)

b. How many people$_i$ do you consider [John friendly with t_i]? (ref: OK, quant: *)

These sentences allow only the referential reading of the amount question. Note that the lack of a quantificational reading in (14)a is incompatible with the analysis of extraction from noun phrases proposed in Stowell (1989). Stowell adopts Abney's analysis of noun phrases as DPs (i.e. as headed by the determiner, see note *** above) and that such extractions involve using the spec(DP) position as an "escape hatch" in just the way

[8]In more technical terminology one might say that a noun, unlike a verb, is not a proper governor for the trace. Therefore, the presence of an intermediate trace in the Spec(CP) of the nominal's sentential complement would violate the ECP.

[9]Cinque (1990) derives the association between long-movement and the impossibility of quantificational interpretation from the requirement that any operator–variable relationship which obtains without obeying antecedent government must be the result of binding, a relationship which is possible only with referential elements. Thus, the possibility of an intermediate trace and scope reconstruction, in Longobardi's terms, will generally coincide with the possibility of chains of antecedent government relations between the moved AQ and its base position. Note that this view of the referential/quantification distinction depends upon an inherent notion of referentiality, and is therefore quite different from our scope-based view presented here. See also Kroch (1989a) for a proposal that all of these asymmetries are the result of the pragmatic (im)plausibility of the existence presupposition in the various cases.

[10]A small clause is a verb-less predication structure, as in (14)b.

that spec(CP) is used in clauses. Under such an analysis, we would, contrary to fact, expect scope reconstruction to apply and the quantificational reading to be available in (14)a just as it is in (10)a. Thus, we conclude that the intermediate positions through which wh-expressions may move, and therefore to which scope reconstruction may take place, must be governed by a verbal, and not by a nominal or adjectival, head.

If we now apply the amount question diagnostic to the gerund case, we see that it admits both the referential and quantificational interpretations:

(15) How many cities$_i$ did Hitler insist on [(our) destroying t_i]? (ref: OK, quant: OK)

If gerunds are assigned the structure given by Abney, this fact is mysterious. As we saw in the example in (14)a, the DP projection does not make its specifier available as a possible site for scope reconstruction.[11] Therefore, the site to which reconstruction is taking place in (15) must be within the VP complement to the D^0 head. We might, in response, adopt an analysis of wh-movement like the one in Chomsky (1986) where movement proceeds first by adjunction to the VP within the DP.

(16) How many cities$_i$ did Hitler insist on [$_{DP}$ (our) [$_{VP}$ t_i^2 [$_{VP}$ destroying t_i^1]]]?

It would then be the t_i^2 position which licensed scope reconstruction. However, if we take this line, we lose our explanation for the lack of ambiguity in extraction from wh-islands, as in (10)b: the extracted element could first adjoin to VP and then move directly to the matrix spec(CP):

(17) How many books$_i$ did they decide [whether to [$_{VP}$ t_i^2 [$_{VP}$ publish t_i^1]]]?

If the VP adjoined position is a possible scope position, then this sentence should also admit a quantificational reading, which it does not. Therefore, we can conclude that there must be some other functional head present in gerunds whose specifier position

[11] Or alternatively, in Cinque's proposal, the DP blocks antecedent government.

serves as a possible scope position and which is absent in true nominals.[12] We will see that this functional head, under a TAG analysis, is actually the complementizer position.

4 Licensing of Parasitic Gaps

Our second piece of evidence that gerunds are clausal appears in the distribution of adjuncts containing parasitic gaps (PGs). PGs are secondary empty positions, usually found in adjunct clauses, in sentences which already contain one gap left by wh-movement, like the second trace (e) in the following example.

(18) a. Which book$_i$ did John lose t$_i$ before reading e?

We assume that PGs result from the movement of an empty operator to the front of the adjunct clause[13] and that this empty operator is itself licensed and given its interpretation by a local binder in a wh- operator position[14]. Frank (1991) observes that the attachment sites of an adjunct with a PRO subject are detectable by what noun phrase can function as controller of the PRO, because the controller must be higher in the structure than the position it controls (see note *** above). Using this diagnostic, we see that adjuncts can ordinarily attach both to nominalizations and to clauses.

(19) a. The newspaper reported on Trump's$_i$ renovation of the building after PRO$_i$ abandoning it years ago

b. The newspaper reported that Trump$_i$ renovated the building after PRO$_i$ abandoning it years ago.

[12]Once again, the alternative proposal in terms of antecedent government is straightforward to state: there must be some projection present in the case of true nominals, but absent in gerunds, which blocks antecedent government.

[13]Here we follow Contreras 1984, Chomsky 1986 and others.

[14]We follow Aoun and Clark 1984 and others.

When an adjunct contains a PG, however, it can attach to a clause but not to a nominal, as indicated by the difference in the possible controllers of PRO in (20)a and (20)b.

(20) a. Which building did the mayor$_i$ report on [$_{\text{DP}}$ Trump's$_j$ renovation of t] [O_k after PRO$_{i/*j}$ previously abandoning e_k]?

b. Which building did the mayor$_i$ report [$_{\text{CP}}$ that Trump$_j$ renovated t] [O_k after PRO$_{i/j}$ previously abandoning e_k]?

The possibility of attachment of an adjunct clause to a nominalization returns immediately when the parastic gap in the adjunct clause is removed, as evidenced by the return of the ambiguity in the controller of PRO:

(21) Which building did the mayor$_i$ report on [$_{\text{DP}}$ Trump's$_j$ renovation of t] [after PRO$_{i/j}$ previously abandoning it]?

This asymmetry is explained if nominals, but not clauses, prevent the empty operator within the PG-containing adjunct from being bound. If we accept, as argued in section ***, that there are no intermediate traces within DP, there will be no local binder for the empty operator in a sentence like (20)a. Hence, attachment of the adjunct within the DP is impossible.

In the behavior of parasitic gaps, gerunds again pattern with clauses.

(22) Which building did the mayor$_i$ report on [$_{\text{GERUND}}$ Trump's$_j$ renovating t] [O_k after PRO$_{i/j}$ previously abandoning e_k]?

In (22), both attachments of the adjunct are possible, indicating that the gerund does not block binding of the empty operator. Once again we conclude that gerunds are clausal.

5 Licensing of Negative Polarity Items

Our final piece of evidence on the categorial status of gerunds concerns the distribution of negative polarity items (NPIs).[15] Klima (1964) observes that while inherently negative verbs are able to license NPIs within their clausal complements, as in (23)a, they are unable to license them within their own clause, as in (23)b.

(23) a. Jones denied that a single allegation of the prosecution had merit.
b. * Jones denied a single allegation of the prosecution.

Progovac (1988) and Laka (1990) analyze this contrast as resulting from an indirect licensing of the NPIs in (23)a. Both authors suggest that the negative verb licenses an abstract element in its complement CP projection which in turn licenses the NPI: for Progovac this element is a negative operator in spec(CP) position, while for Laka it is a C^0 head with negative features.

Note that NPIs are illicit inside the NP complements of negative verbs, as seen in (24)a, just as they are as the bare complement itself. In contrast, example (24)b shows that gerund complements do permit NPIs within them.

(24) a. * The owner denied [the proposal of a single improvement].
b. I denied [GERUND (?his) eating a single cookie].

Pursuing the line of Laka and Progovac, we must assume that this contrast lies in a difference in the ability of the negative verb to select/license a negative head/operator in the outermost projection of a nominalization and a gerund, which can then serve to license the NPI. The data in (24) tell us that gerunds permit such a negative element, while nominalizations do not. Yet, if gerunds and nominalizations are identical in their outer structure (both being DPs), it will be difficult to explain this contrast. Under this view, one would expect that either both or neither of these DPs could license such a negative head or operator.

[15] Negative polarity items are lexical items that must occur within the scope of a negative. Examples are: *give a damn*, and some uses of *any* and *a single*.

Let us generalize Laka's proposal to allow inherently negative verbs to select for negative features on any [+V] head whose projection it takes as a complement. A negative C^0 is only one instance of this more general possibility. Now, if the gerund is externally clausal (i.e. if its highest functional projection is [+V]), then we explain the judgment in (24)b. On the other hand, since the maximal projection of the nominalization is a DP (hence [-V]), an inherently negative verb will not be able to select for negative features on the head of the complement; and NPIs will correctly be predicted to be ungrammatical.

We can derive further support for our proposal that negative predicates select for a [+V] negative functional head from clear cases of IP complementation. In example (25), it is standardly assumed that the negative predicate *unlikely* selects an IP out of which subject raising takes place. As our extension of Laka's proposal predicts, this IP can contain an NPI.

(25) John is unlikely to see a single person at the rally.

If negative features could be borne only by C^0 heads, this example would be problematic since raising is usually taken to proceed only out of IP, and not out of CP complements. This example also supports Laka's proposal, in which licensing of the NPI is via a negative head, over Progovac's proposal, where an operator in specifier position is the licensing element. Spec(IP) is filled by the trace of *John*, and therefore cannot host a negative operator.

6 Gerunds as IPs

The evidence we have considered suggests that complement gerunds are clausal. Yet the dilemma posed by examples (6) through (9) remains: The distribution of gerunds is not entirely parallel to that of clauses. We propose that these problems can be solved by assimilating governed gerunds to bare IPs, much like the exceptional bare IP arguments of raising and ECM predicates. This hypothesis allows us reduce the contrasts in (6) and (7) to the fact that IPs cannot be extraposed (Baltin 1981). We can also account

for the possibility of subject–aux inversion over gerunds seen in (8)a as deriving from the stipulation that IP sentential subjects may appear in the spec(IP) position of a clause, while CP sentential subjects may not (cf. Koster 1978). Finally, the lack of wh-movement to the initial position of gerunds shown in (9)a is explained by the absence of a spec(CP) position necessary to license the appearance of a wh-element. Here, we follow the general approach of Rizzi (1991) in assuming that only C^0 heads can be endowed with the [+wh] features capable of licensing a wh-operator in their specifiers. Thus, IP complements can never license wh-elements in their specifier positions so as to form indirect questions. Compare (26)a with (26)b. [16]

(26) a. * Lenore couldn't believe [$_{IP}$ which pasta to have finished cooking first]
b. Lenore couldn't believe [$_{CP}$ which pasta had finished cooking first]

Note finally that viewing complement gerunds as IPs accounts immediately for the grammaticality, for many speakers, of examples like (27) (Postal 1974).

(27) Through its actions, the EU stopped/prevented there from being an all-out war.

Under theoretical assumptions in which the effects of a raising to object transformation are disallowed by the projection principle (but see Lasnik and Saito 1992), this sentence poses an apparent problem since *there* is licensed not as an argument of the verb *prevent* but rather as a part of the copular construction expressed by the gerund. Given our analysis of gerunds as IPs, we can construe *from* as being categorially an I^0. In fact, *from* will be an overt negative counterpart of the abstract I^0 element in (24)b. As the following example shows, this negative I^0 freely licenses an NPI in its clause:

[16]It might be interesting to compare this behavior with that in a language such as Polish which has been argued to allow wh-movement to an IP-adjoined position (Lasnik and Saito 1984). Also, note that if Lasnik and Saito (1992) are correct that topicalization in English involves adjunction to IP, we predict that gerunds should also allow such topicalization. This prediction, however, is not in accord with the facts:

(i) * Fred likes these beans$_i$ eating t_i, and those beans$_j$ cooking t_j.

We have nothing to say about this at the moment except that it raises doubts about the Lasnik and Saito analysis.

(28) The alarm prevented the burglar from taking a single jewel.

Thus, if we assume that the element *there* originates within the VP, then (27) will be assigned the structure in (29).[17]

(29) The EU prevented [$_{\text{IP}}$ there$_i$ [$_{\text{I}'}$ from [$_{\text{VP}}$ t_i being an all-out war]]]

7 Structural Recursion and Tree Adjoining Grammar

We have just argued on the basis of data concerning amount quantifiers, parasitic gaps and negative polarity items that the structure of certain gerunds is clausal. In particular, we have proposed that they should be treated as IPs, in contrast to lexically derived nominalizations, which should be treated as DPs. In the last section, we saw how this proposal accounts, not only for the differences between gerunds and nominalizations but also for certain similarities between them in contexts of occurrence. However, we have not explained why the structural distinction between gerunds and nominalizations should have the particular effects on these two constructions that it does. That is, why should the difference between a [+V,–N] and [–V,+N] argument have the effects that it does upon amount quantifier scope, parasitic gap licensing, and negative polarity item licensing? In each of these cases, we have simply stipulated that clausal or [+V] elements should behave in certain way, while nominal elements should behave in another. We will now show that by looking at these phenomena within the Tree Adjoining Grammar (TAG) formalism, the observed differences between clausal and nominal elements can be explained.

7.1 TAG Basics

In TAG, complex structures are built out of the atomic objects of the formalism, which are called the elementary trees. Frank (1992) argues that the domain of each of these

[17] For discussion of the deep structure position of *there*, see Moro (1990). Strictly speaking, the structure in (29) is incompatible with Moro's proposal in that he argues that *there* is generated as the predicate of a small clause.

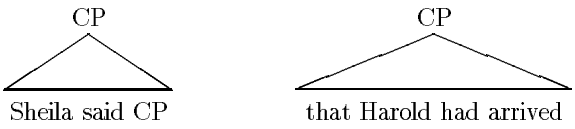
elementary trees is limited by the following principle:

(30) **Condition on Elementary Tree Minimality (CETM):** Every elementary tree consists of the extended projection of a single lexical head.

This condition utilizes Grimshaw's (1991) notion of extended projection which, roughly stated, is the domain up to and including the functional projections which are associated with a given lexical head. Thus, the extended projection of a V^0 head can include the IP and CP projections which dominate it, but not the projections of any phrases which the V^0 takes as arguments, and not the projection of a head which takes this CP–IP–VP complex as its complement. In the nominal domain, the extended projection of an N^0 head includes the DP projection which takes the NP as its complement. As a result of the CETM, then, each nominalization, gerund, or clause must originate in a separate elementary tree since each is headed by a distinct lexical head.

Elementary trees can be combined using one of the operations defined by the formalism: substitution and adjoining. The application of the adjoining operation requires the presence of a recursive piece of phrase structure called an auxiliary tree. An auxiliary tree is an elementary tree which has a non-terminal along its frontier, called the foot node, which is categorially identical to the root node. When complex constructions involve recursion, such as cases of sentential embedding, the phrase structure may be constructed using either adjoining or substitution. Thus, the elementary trees in (31) can be combined either by substituting the tree on the right into the CP node on the frontier of the tree on the left, or by adjoining the tree on the left to the root of the tree on right.[18]

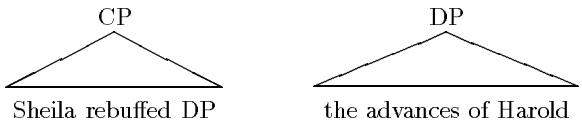
(31)



[18]For the sake of readability, we will instantiate the DP arguments in the trees that follow except where this instantiation is a crucial part of the derivation. We will assume that these DPs are filled through substitution in the course of the derivation.

In complex constructions which do not involve recursion, however, only the substitution derivation is available. To insert the nominalization elementary tree on the right in (32) into the complement position of the tree on the left, we must use substitution since the clausal tree is not a recursive structure as it was in (31), and hence cannot function as an auxiliary tree.

(32)



Under our assumption that gerunds are IPs, they will admit both the substitution and adjoining derivations. If we allow a main clause elementary tree to be rooted in IP as well as in CP, then the trees needed for the derivation of a sentence with a gerund complement will be as in (33).

(33)

These trees are essentially parallel to those shown in (31) and correspondingly allow both the substitution and adjoining derivations.

On first consideration, it might appear that the possibility or impossibility of the adjoining derivation should no effect on the structure which is derived. In fact, however, the difference between the adjoining and substitution derivations is strongly reflected in the way that features are percolated through the derived structure, assuming the system of feature percolation of Vijay-Shanker (1987). In this system, each node in an elementary tree is associated with a pair of feature structures. When an adjoining takes place at a certain node, the top feature structure of the root node of the auxiliary unifies with the top feature structure at the site of the adjoining, and bottom feature structure of the foot node unifies with the bottom feature structure at the site of adjoining. During substitution, on the other hand, the feature structures at the site of substitution

all unify with the top feature structure of the tree which is being substituted. In all TAG derivations, the pairs of feature structures at each node must unify with one another as a last step in the derivation.

The linguistic intuition behind the use of a pair of feature structures at each node has been this: The top feature structure represents a description of the requirements imposed on a node from the structural context above. Thus, any restrictions which a head imposes on its complement that are instantiated via features, case for example, must be placed in the top feature structure of the complement node. The bottom feature structure, on the other hand, is taken to represent the description of the lower structural context and the requirements that this context imposes on the node. Thus, if we have a DP which exhibits a particular lexically expressed case, the lower feature structure of this DP will express the fact. Ultimately, the two descriptions of a node which are imposed from its upper and lower context must be consistent with one another; otherwise, we will have instances of conflicting requirements. This fact lies at the root of the requirement that all pairs of feature structures be unified at the conclusion of the derivation.[19]

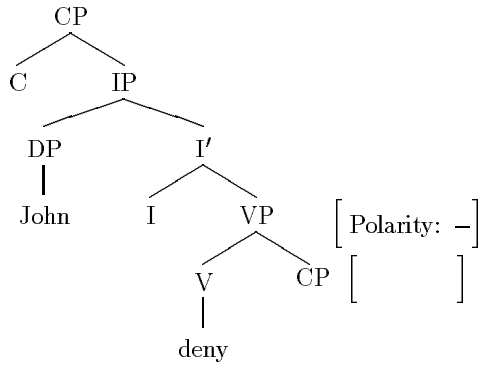
7.2 Explaining the Gerund/Nominal Contrasts

We now are in a position to understand in formal terms why gerunds and nominalizations exhibit the contrasting behavior described in earlier sections of this paper. We begin by investigating how Vijayshankar's TAG approach to feature expression and feature percolation can explain the facts concerning the licensing of negative polarity items. Suppose that NPI licensing is mediated through the existence of a feature we will call *Polarity*. This feature can take one of two values, + or −. We will interpret the − value of this feature as the one which is relevant to the licensing of NPIs. In particular, let us suppose that an NPI must be licensed by the presence of a locally commanding head, i.e. within its own elementary tree, which is specified as having a − value for Polarity. Now, we must ask what gives rise to the existence of a [Polarity:−] specification in a

[19]See Joshi et al. (1991) and Frank (1992) for further discussion.

phrase marker. Following the views of Laka and Progovac discussed above, we propose that negative predicates like *deny* have certain selectional properties which license NPIs. In our terms, this translates as the licensing of the presence of a [Polarity:−] specification on the root node of their complement. By the CETM, this complement node must either be the foot node of an auxiliary tree or a substitution site; hence, it cannot dominate any other structure in the elementary tree. Consequently, the bottom feature structure will be empty and the Polarity specification will be expressed only on the top feature structure. An elementary tree headed by *deny*, then, will look like this:

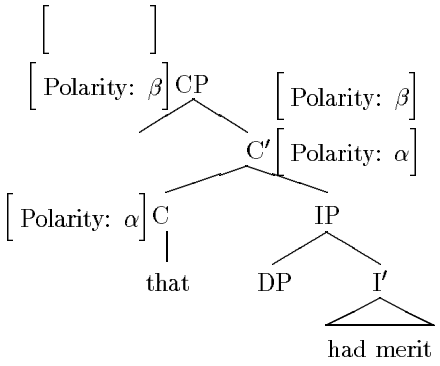
(34)



Now, let us suppose that all verbal projections, i.e. those with [+V,-N] categorial features (including CP, IP and VP), have the ability to host *Polarity* features, as we suggested above (cf. example (25)). In general, the value of this feature is lexically underspecified, as with the English complementizer *that* which can appear either in negative or positive polarity contexts, although the polarity value of a head can sometimes be restricted to −, as in the case of the inflectional element *from* (cf. example (27)) and in cases from Basque and Serbo-Croatian that Laka and Progovac discuss. A CP elementary tree for

a sentential complement will thus have the following form:[20]

(35)



If such a tree is substituted into the CP complement position of (34), or alternatively the tree in (34) is adjoined into the root of (35), the negative value for Polarity specified by the embedding predicate *deny* will collapse with the variable β to produce a negative Polarity feature on the complement CP. If nothing further is adjoined at the C′ node, this value will percolate down to the C^0 head as α and β are collapsed. Thus, if an NPI such as *a single argument* is substituted into the spec(IP) subject position, it will be licensed by the locally commanding C^0 head, now specified appropriately as to Polarity.

Note that this analysis works perfectly well for gerund complements. As IPs, they are categorially [+V,-N] and hence may bear Polarity features. Thus, in an exactly analogous fashion to the CP case, an I^0 head can inherit the negative Polarity features and thereby license the insertion of an NPI into a position filled by one of the gerund's arguments. Let us contrast these cases with the case a nominal complement to *deny*. We are assuming that nominal projections are unable to bear Polarity features. Consequently, there is no way for a negative predicate to express the selectional

[20]We assume, in fact, that Polarity features are percolated down not only through a single projection as shown here for CP, but also throughout the extended projection. Cases such as (i) provide strong support for such a conclusion.

(i) Alice denied that I seemed to anyone to be drunk.

Here, the NPI anyone will need to be licensed locally within the *seem* auxiliary tree which gives rise to the raising structure following the analysis of raising in Frank (1992) and Kroch and Joshi (1985). In order for the Polarity feature to percolate correctly into this structure, it must be present on the IP projection, in Frank's analysis on the top feature structure of the I′ node of the *be drunk* tree. A further discussion of these cases would take us quite far afield, however, and we leave them for another occasion.

relationship that would license an NPI within its complement. Consequently, an NPI within the nominal complement to a negative predicate can never be sufficiently local to the requisite [Polarity: −] features and can never be well-formed.[21,22]

We turn now to the differences between gerunds and nominals in AQ interpretation. Kroch (1989b) shows how cases of successive cyclic wh-extraction can be handled in a TAG using the adjoining operation. The case of wh-movement in (10)a (repeated as (38) below) can be derived by adjoining the auxiliary tree representing the matrix clause in (39) into the elementary tree in (40) at the C′ node.[23]

(38) How many books$_i$ did they decide [to publish t_i]?

[21]It is interesting to compare the situation of complements to verbal predicates that are nominals, which we have just discussed, to that of complements to nominals which are themselves clauses. As an anonymous reviewer points out, clausal complements to inherently negative derived nominals behave identically to those of their verbal counterparts with respect to NPI licensing.

(36) a. John denies that her case has any merits.
b. John's denial that her case has any merits is outrageous.
(37) a. Mary doubts that they will see a single person.
b. Mary's doubts that they will see a single person.

These facts suggest that the crucial property in such indirect NPI licensing is the categorial status of the complement, which impacts on its ability to bear Polarity features, as opposed to the categorial status of the negative predicate itself or its compatibility with the category of the complement, as we suggested in an earlier version of this paper.

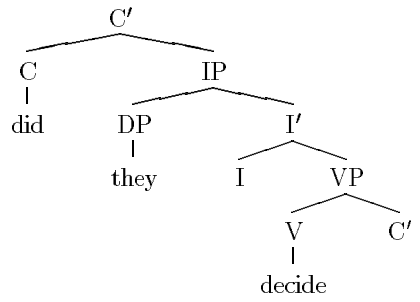
[22]We have put aside the question of how NPIs are able to appear in argument positions where the selecting predicate is not the relevant NPI licenser, such as occurs in the presence of sentential negation, interrogation, conditionals and the like (cf. Ladusaw 1980 and Linebarger 1980, 1987):

(i) Alice didn't like a single gift that we gave her.
(ii) Did you see a single person at the mall?
(iii) If I hear a single peep, I'll scream!

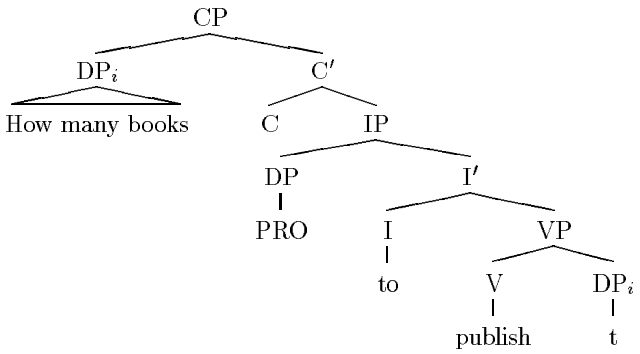
The natural proposal here would be to allow negation to license the presence of the − Polarity feature in a fashion similar to the negative predicates. Since this Polarity feature would command the complement position in a clausal structure, we would correctly predict that NPIs may appear in complement position as in (i). Certain details regarding the manner in which this can be percolated remain to be worked out, but the general direction is clear.

[23]The trees that we use in this derivation are those exploited in a revision of Kroch's analysis proposed in Frank (1992) which accounts for the contrasts between extraction from bridge and non-bridge verb complements.

(39)



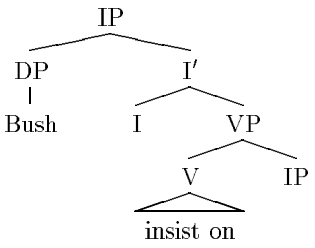
(40)



Extraction from a gerund complement can be accomplished in a similar fashion. To derive the sentence in (15) (repeated below as (41)), we need only adjoin the main clause auxiliary tree in (42), this time recursive on IP, into the embedded clause elementary tree in (43) at the node labeled IP.[24,25]

(41) How many tanks$_i$ did Bush insist on [destroying t_i]?

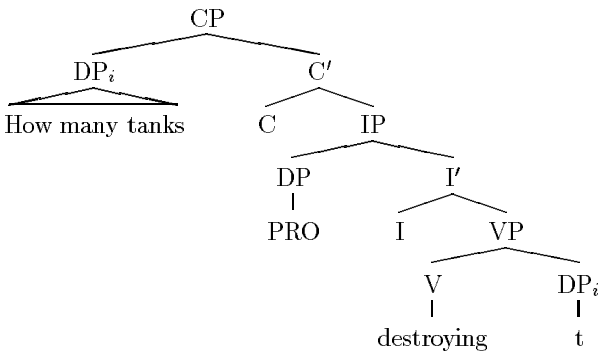
(42)



[24]We ignore the complicating issues relating to the placement of the auxiliary *did* in this example. See Frank (1992) for details.

[25]As an anonymous reviewer points out, it is necessary to block the use of the CP structure in (43) as an indirect question complement for reasons mentioned above in relation to the examples (9). This can be done in any number of ways by positing an incompatible pair of features at the IP level. We leave open the characterization of the features which give rise to this conflict.

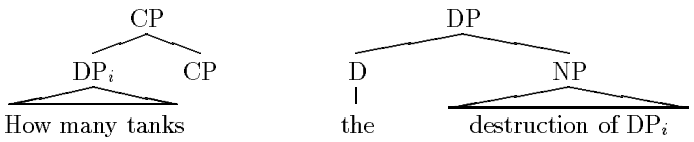
(43)



If we now try to derive the case of extraction from the nominalization in example (14)a in a similar fashion, we see that it is impossible. The reason for this is simple: There is no way for the wh-expression *how many tanks* to occupy a spec(CP) position in the same elementary tree as the nominal predicate in which the trace occurs since they will constitute two distinct extended projections, and hence two elementary trees, according to the CETM.[26]

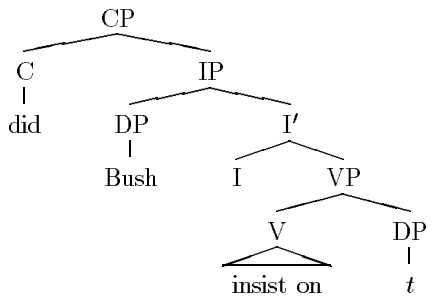
Kroch (1989b) suggests using multi-component TAG to derive these grammatical cases of extraction from object nominals. Multi-component TAG extends the concept of auxiliary tree to allow for auxiliary tree sets containing multiple auxiliary trees which together enter a derivation. The elements of a tree set are subject to the formal restriction that they must all adjoin into the same elementary tree at a single point in the derivation. Using this extended system, sentence (14)a can now be derived by adjoining the auxiliary tree set in (44) into the elementary tree in (45).

(44)



[26]We should point out that though the structure in (43) is rooted in CP, it does not violate our analysis of gerunds as IP projections. The interpretation of the "gerunds as IPs" proposal which we intend requires only that predicates which select for gerunds must select for IPs. Emphatically, it does not require all gerundival forms to be part of an IP-rooted elementary tree. This latter option would have the undesirable consequence of complicating the principles of projection, and the CETM.

(45)



Such multi-component adjoining is also necessarily involved in the derivation of cases of subjacency violating long movement like that in (10)b. In fact, Kroch (1989b) shows that the cases which induce subjacency violations are exactly those which require using multi-component derivations. Of course, even if a sentence has a derivation which exploits only simple non-multi-component adjoining, it also has a multi-component derivation.

We can now link the quantificational reading of the amount quantifier to a simple adjoining derivation, and the referential reading to the multi-component derivation. Since we have interpreted the notion of "quantificational reading" as narrow scope of the existential quantifier associated with the presupposition of the AQ, this link will be established if we adopt an interpretive procedure under which a quantifier's scope is limited to the elementary tree in which it is generated; and, of course, this procedure is essentially equivalent to scope reconstruction in a transformational analysis. For cases such as extraction from wh-islands and nominalizations, which necessarily involve multi-component adjoining, the AQ is not generated in the same clause as its base position and therefore is prevented from taking scope within this domain. Hence, it cannot be interpreted as taking narrow scope; i.e., quantificationally. Instead, the AQ may only take wide scope which, as we saw above, corresponds to the referential reading. Instances of extraction which can use simple adjoining, on the other hand, allow the AQ to be generated in its base clause and so to be interpreted with the quantificational reading. Since nothing blocks the multi-component derivation from occurring in these cases, the referential reading is also available. This view of AQ interpretation explains immediately why the gerund case patterns with the ordinary clausal case, and not

with wh-islands and nominalizations. Extraction from gerunds, as IP complements, can proceed by simple adjoining; and since the multi-component derivation is not forced, both the quantificational and the referential reading are licensed.

Our account of the AQ scope ambiguities is supported by the following example involving extraction from an IP complement.

(46) How many books$_i$ are they likely [$_{\text{IP}}$ to publish t_i]? (r ef: OK, quant: OK)

This sentence allows both referential and quantificational interpretations. Under the version of Longobardi's account we discussed in section 3, the ambiguity is surprising since there is no intermediate spec(CP) trace position at which the wh-expression *how many books* can be interpreted in order to take the embedded, i.e. quantificational, reading.[27] However, as we mentioned above, we do not want to allow a position adjoined to VP to serve a site for scope interpretation because then this case will collapse with that of wh-islands, an undesirable result. Under the TAG-based account, however, the sentence behaves exactly as predicted. It can be derived without the use of multi-component adjoining, and hence should allow the quantificational reading.

We turn finally to our parasitic gap evidence. Similar considerations to those relevant to amount question interpretation apply here. As suggested in Frank (1991), it is only when the derivation proceeds by multi-component adjoining in which the complement is non-recursive that the empty operator of a PG cannot be licensed outside of its local domain. Otherwise, the domain in which the operator may be bound can be extended by feature percolation which takes place under the appropriate instances of adjoining. We refer to Frank (1991) for details. This analysis leads to precisely the correct predictions. Since extraction from nominalizations requires multi-component adjoining because the complement is non-recursive, a parasitic gap is illicit within the nominalization domain, as shown in (20)a. However, in cases such as (22), where no multi-component adjoining is required, attachment to the gerund of the PG containing adjunct clause is possible. In addition, this analysis correctly predicts that cases requiring multi-component adjoining,

[27]Observe that Cinque's (1990) proposal makes the correct prediction in this case since antecedent government will be able to obtain across the IP boundary.

but which involve recursive complements, such as wh-island extractions, do allow low attachment of parasitic gaps. This is shown in (47) (once again see Frank (1991) for further discussion).

(47) Which car did Bill$_i$ figure out [$_{\mathrm{CP}}$ how the mechanic$_j$ h ad fixed t] [$_{\mathrm{PP}}$ without O_k PRO$_{i/j}$ dismantling e_k]

8 Conclusions

We have now shown that the way in which the TAG formalism composes complex structures out of elementary trees explains the syntactic differences between gerunds and nominalizations. In three empirical domains, negative polarity item licensing, amount question interpretation, and parasitic gap licensing, we have seen that gerund complements behave like clausal complements to bridge verbs, allowing long distance relations between positions within them and within their matrix clause, while nominalizations behave like syntactic islands. A representational system, like transformational grammar, which has available the entire structure of a sentence for the statement of relations among its elements, can capture the facts we have presented only by stipulating differences between the licensing and interpretive properties of the two constructions. It is certainly not necessary in such a system that the two behave differently, nor that they behave consistently across the environments that we have investigated. In contrast, under a TAG-based analysis, once we have determined that gerunds are [+V] while nominalizations are [+N], all of the differences we have described follow directly. Only extended projections with [+V] heads can allow the insertion, via adjoining, of clausal (hence also [+V]) auxiliary trees to stretch local dependences across the boundaries of predicates. The composition of phrase structure subtrees which disagree in categorial features cannot take place by simple adjoining and must rather invoke either substitution or multicomponent adjoining. These operations, as we have seen, have quite different locality properties than simple adjoining which account straightforwardly for the stricter locality properties of nominalizations when compared to gerunds.

References

Abney, Steven. 1987. *The English Noun Phrase in its Sentential Aspects*. PhD thesis, MIT.

Aoun, Joseph and Robin Clark. 1984. On non-overt operators. Manuscript, USC/UCLA.

Baltin, Mark. 1981. Strict bounding. In C. L. Baker and J. J. McCarthy (Eds.), *The Logical Problem of Language Acquisition*. MIT Press.

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Cinque, Guglielmo. 1990. *Types of $\overline{A}$-Dependencies*. Cambridge, MA: MIT Press.

Contreras, Heles. 1984. A note on parasitic gaps. *Linguistic Inquiry* 15.

Frank, Robert. 1991. Parasitic gaps and locality conditions. In *Proceedings of the 27th Regional Meeting of the Chicago Linguistics Society*. University of Chicago.

Frank, Robert. 1992. *Syntactic Locality and Tree Adjoining Grammar: Grammatical, Acquisition and Processing Perspectives*. PhD thesis, University of Pennsylvania.

Grimshaw, Jane. 1991. Extended projection. Manuscript, Brandeis University.

Joshi, Aravind K. 1985. How much context-sensitivity is required to provide reasonable structural descriptions: tree adjoining grammars. In D. Dowty, L. Kartunnen, and A. Zwicky (Eds.), *Natural Language Processing: Psycholinguistic, Computational and Theoretical Perspectives*. Cambridge University Press.

Joshi, Aravind K., K. Vijay-Shanker, and David Weir. 1991. The convergence of mildly context-sensitive grammatical formalisms. In P. Sells, S. Shieber, and T. Wasow (Eds.), *Foundational Issues in Natural Language Processing*. Cambridge, MA: MIT Press.

Klima, Edward S. 1964. Negation in English. In J. A. Fodor and J. J. Katz (Eds.), *The Structure of Language*. Prentice Hall.

Koster, Jan. 1978. Why subject sentences don't exist. In S. J. Keyser (Ed.), *Recent Transformational Studies in European Languages*. MIT Press.

Kroch, Anthony. 1989a. Amount quantification, referentiality, and long Wh-movement. Manuscript, University of Pennsylvania.

Kroch, Anthony. 1989b. Asymmetries in long distance extraction in a tree adjoining grammar. In Mark Baltin and Anthony Kroch (Eds.), *Alternative Conceptions of Phrase Structure*. University of Chicago Press.

Kroch, Anthony and Aravind K. Joshi. 1985. The linguistic relevance of tree adjoining grammar. Technical Report MS-CS-85-16, Department of Computer and Information Sciences, University of Pennsylvania.

Laka, Itziar. 1990. *Negation in Syntax: on the Nature of Functional Categories and Projections*. PhD thesis, MIT.

Lasnik, Howard and Mamoru Saito. 1984. On the nature of proper government. *Linguistic Inquiry* 15.

Lasnik, Howard and Mamoru Saito. 1992. *Move-α*. Cambridge, MA: MIT Press.

Longobardi, Giuseppe. 1990. Extraction from NP and the proper notion of head government. In A. Giorgi and G. Longobardi (Eds.), *The Syntax of Noun Phrases*. Cambridge University Press.

Moro, Andrea. 1990. The raising of predicates: Copula, expletives and existence. Manuscript, Università di Venezia and MIT.

Postal, Paul. 1974. *On Raising*. Cambridge, MA: MIT Press.

Progovac, Ljiljana. 1988. *A Binding Approach to Polarity Sensitivity*. PhD thesis, University of Southern California.

Rizzi, Luigi. 1991. Residual verb second and the WH criterion. Manuscript, Université de Genève.

Ross, John Robert. 1967. *Constraints on Variables in Syntax*. PhD thesis, MIT. published as *Infinite Syntax*, Ablex, 1986.

Stowell, Tim. 1989. Subjects, specifiers and X-bar theory. In M. Baltin and A. Kroch (Eds.), *Alternative Conceptions of Phrase Structure*. University of Chicago Press.

Vijay-Shanker, K. 1987. *A Study of Tree Adjoining Grammars*. PhD thesis, University of Pennsylvania.