



March 1990

A Lexicalized Tree Adjoining Grammar for English

Anne Abeillé
University of Paris

Kathleen Bishop
University of Pennsylvania

Sharon Cote
University of Pennsylvania

Yves Schabes
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/cis_reports

Recommended Citation

Anne Abeillé, Kathleen Bishop, Sharon Cote, and Yves Schabes, "A Lexicalized Tree Adjoining Grammar for English", . March 1990.

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-90-24.

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/cis_reports/527
For more information, please contact repository@pobox.upenn.edu.

A Lexicalized Tree Adjoining Grammar for English

Abstract

This paper presents a sizable grammar for English written in the Tree Adjoining grammar (TAG) formalism. The grammar uses a TAG that is both lexicalized (Schabes, Abeillé, Joshi 1988) and feature-based (Vijay-Shankar, Joshi 1988). In this paper, we describe a wide range of phenomena that it covers.

A Lexicalized TAG (LTAG) is organized around a lexicon, which associates sets of elementary trees (instead of just simple categories) with the lexical items. A Lexicalized TAG consists of a finite set of trees associated with lexical items, and operations (adjunction and substitution) for composing the trees. A lexical item is called the *anchor* of its corresponding tree and directly determines both the tree's structure and its syntactic features. In particular, the trees define the domain of locality over which constraints are specified and these constraints are local with respect to their anchor. In this paper, the basic tree structures of the English LTAG are described, along with some relevant features. The interaction between the morphological and the syntactic components of the lexicon is also explained.

Next, the properties of the different tree structures are discussed. The use of S complements exclusively allows us to take full advantage of the treatment of unbounded dependencies originally presented in Joshi (1985) and Kroch and Joshi (1985). Structures for auxiliaries and raising-verbs which use adjunction trees are also discussed. We present a representation of prepositional complements that is based on extended elementary trees. This representation avoids the need for preposition incorporation in order to account for double wh-questions (preposition stranding and pied-piping) and the pseudo-passive.

A treatment of light verb constructions is also given, similar to what Abeillé (1988c) has presented. Again, neither noun nor adjective incorporation is needed to handle double passives and to account for CNPC violations in these constructions. TAG'S extended domain of locality allows us to handle, within a single level of syntactic description, phenomena that in other frameworks require either dual analyses or reanalysis.

In addition, following Abeillé and Schabes (1989), we describe how to deal with semantic non compositionality in verb-particle combinations, light verb constructions and idioms, without losing the internal syntactic composition of these structures.

The last sections discuss current work on PRO, case, anaphora and negation, and outline future work on copula constructions and small clauses, optional arguments, adverb movement and the nature of syntactic rules in a lexicalized framework.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-90-24.

**A Lexicalized Tree
Adjoining Grammar For
English**

**MS-CIS-90-24
LINC LAB 170**

**Anne Abeillé
Kathleen Bishop
Sharon Cote
Yves Schabes**

**Department of Computer and Information Science
School of Engineering and Applied Science
University of Pennsylvania
Philadelphia, PA 19104-6389**

April 1990

A Lexicalized Tree Adjoining Grammar for English

Anne Abeillé, Kathleen Bishop, Sharon Cote and Yves Schabes*

March 1990

ABSTRACT

This paper presents a sizable grammar for English written in the Tree Adjoining grammar (TAG) formalism. The grammar uses a TAG that is both lexicalized (Schabes, Abeillé, Joshi 1988) and feature-based (Vijay-Shankar, Joshi 1988). In this paper, we describe a wide range of phenomena that it covers.

A Lexicalized TAG (LTAG) is organized around a lexicon, which associates sets of elementary trees (instead of just simple categories) with the lexical items. A Lexicalized TAG consists of a finite set of trees associated with lexical items, and operations (adjunction and substitution) for composing the trees. A lexical item is called the *anchor* of its corresponding tree and directly determines both the tree's structure and its syntactic features. In particular, the trees define the domain of locality over which constraints are specified and these constraints are local with respect to their anchor. In this paper, the basic tree structures of the English LTAG are described, along with some relevant features. The interaction between the morphological and the syntactic components of the lexicon is also explained.

Next, the properties of the different tree structures are discussed. The use of S complements exclusively allows us to take full advantage of the treatment of unbounded dependencies originally presented in Joshi (1985) and Kroch and Joshi (1985). Structures for auxiliaries and raising-verbs which use adjunction trees are also discussed. We present a representation of prepositional complements that is based on extended elementary trees. This representation avoids the need for preposition incorporation in order to account for double wh-questions (preposition stranding and pied-piping) and the pseudo-passive.

A treatment of light verb constructions is also given, similar to what Abeillé (1988c) has presented. Again, neither noun nor adjective incorporation is needed to handle double passives and to account for CNPC violations in these constructions. TAG's extended domain of locality allows us to handle, within a single level of syntactic description, phenomena that in other frameworks require either dual analyses or reanalysis.

In addition, following Abeillé and Schabes (1989), we describe how to deal with semantic non compositionality in verb-particle combinations, light verb constructions and idioms, without losing the internal syntactic composition of these structures.

The last sections discuss current work on PRO, case, anaphora and negation, and outline future work on copula constructions and small clauses, optional arguments, adverb movement and the nature of syntactic rules in a lexicalized framework.

*We have benefited immensely from our numerous discussions with Aravind Joshi. Thanks are due also to Jack Hoeksema, and Anthony Kroch. This work was partially supported by ARO grant DAA29-84-9-007, DARPA grant N0014-85-K0018, NSF grants MCS-82-191169 and DCR-84-10413 to the University of Pennsylvania.

Contents

1	INTRODUCTION	3
2	Organization of the Grammar	10
2.1	Syntax	10
2.2	Interaction of the Morphological and Syntactic Lexicons	12
3	Sentential Complements	15
3.1	S or VP arguments?	15
3.2	Extraction properties	16
3.3	Selecting the Appropriate Type of Sentential Argument	21
4	Auxiliaries	23
4.1	Passives	25
4.2	Ordering Auxiliaries and Specifying Verb Form in Sentential Arguments	25
5	Raising constructions	26
5.1	Raising passives	27
5.2	Raising Adjectives and Nouns	29
6	Prepositional Complements	31
7	Verb-Particle Combinations	35
8	Light Verb Constructions	38
9	Idioms	43
10	Ongoing Work	45
10.1	PRO, Case, and Anaphora	45
10.2	Negation	45
11	Future Topics	46
11.1	Copula Constructions and Small Clauses	46
11.2	Adverb Movement	46
11.3	Optional Arguments	48
11.4	Syntactic and Lexical Rules	48
12	Conclusion	50
A	APPENDIX 1: Organization of the syntactic lexicon	51
A.1	Verbs	51
A.2	Auxiliaries	52
A.3	Complementizers	53
A.4	Nouns and Pronouns	53
A.5	Determiners	54
A.6	Adjectives	54
A.7	Prepositions	56
A.8	Adverbs	57
B	APPENDIX 2 : List of the Feature Structures used	59
B.1	The features <MODE> and <COMP>	59
B.2	Tense and Related Features	60
B.3	Agreement Features	60
B.4	Selectional Restrictions	61

C	APPENDIX 3 : Samples of structures handled	62
C.1	Notational Conventions	62
C.2	List of Tree Families	63
C.2.1	Verbs with nominal arguments	63
C.2.2	Verbs with sentential arguments	63
C.2.3	Light-verb constructions	64
C.2.4	Verb-particle combinations	65
C.2.5	Idioms	67

1 INTRODUCTION

Most current linguistic theories give lexical accounts of several phenomena that used to be considered purely syntactic. The information put in the lexicon is thereby increased in both amount and complexity: see, for example, lexical rules in LFG (Bresnan and Kaplan, 1983), GPSG (Gazdar, Klein, Pullum and Sag, 1985), HPSG (Pollard and Sag, 1987), Combinatory Categorical Grammars (Steedman 1985, 1988), Karttunen's version of Categorical Grammar (Karttunen 1986, 1988), some versions of GB theory (Chomsky 1981), and Lexicon-Grammars (Gross 1984).

Following Schabes, Abeillé and Joshi (1988) we say that a grammar is 'lexicalized' if it consists of:¹

- a finite set of structures each associated with a lexical item; each lexical item will be called the *anchor*² of the corresponding structure; the structures define the domain of locality over which constraints are specified; constraints are local with respect to their anchor;
- an operation or operations for composing the structures.

'Lexicalized' grammars (Schabes, Abeillé and Joshi, 1988), systematically associate each elementary structure with a lexical anchor. These elementary structures specify extended domains of locality (as compared to CFGs) over which constraints can be stated. The 'grammar', consists of a lexicon where each lexical item is associated with a finite number of structures for which that item is the anchor. There are no separate grammar rules at this level of description, although there are, of course, 'rules' which tell us how these structures are combined. In general, this is the level of description that we will be describing in this paper. At a higher level of description, the grammar rules and principles that are implicit in the form of the lexicon would be stated explicitly. For example, there are principles which govern which trees are grouped together into tree families, and rules which describe the relations between structure types across tree families (see subsection 2.1 for a discussion of tree families.) The information explicitly provided in this more abstract representation of the grammar may be thought of as an interpretation of the data in the lower-level representation.³

Not every grammar is in a 'lexicalized' form.⁴ In the process of lexicalizing a grammar, the 'lexicalized' grammar is required to produce not only the same language as the original grammar, but also the same structures (or tree set).

For example, a CFG, in general, will not be in a 'lexicalized' form. The domain of locality of CFGs can be easily extended by using a tree rewriting grammar (Schabes, Abeillé and Joshi, 1988) that uses only substitution as a combining operation. This tree rewriting grammar consists of a set of trees that are not restricted to be of depth one (as in CFGs). Substitution can take place only on non-terminal nodes of the frontier of each tree. Substitution replaces a node marked for substitution by a tree rooted by the same label as the node (see Figure 1; the substitution node is marked by a down arrow ↓).

However, in the general case, CFGs cannot be 'lexicalized', if only substitution is used (for further explanation why, the reader is referred to Schabes, Abeillé and Joshi, 1988). Furthermore, in general, there is not enough freedom to choose the anchor of each structure. This is important because we want the choice

¹By 'lexicalization' we mean that in each structure there is a lexical item that is realized. We do not mean simply adding feature structures (such as head) and unification equations to the rules of the formalism.

²In previous publications, the term 'head' was used instead of the term 'anchor'. From now on, we will use the term anchor instead; the term 'head' introduces some confusion because the lexical items which are the source of the elementary trees may not necessarily be the same as the traditional syntactic head of those structures. In fact, the notion of anchor is in some ways closer to the notion of function in Categorical Grammars.

³There may also be some linguistic generalizations which can not be stated as an explicit representation of implicit lexical information. These would likely be statements concerning the constraints on syntactic structures that are needed to evaluate the acceptability of a new lexical item or structure. We have not yet needed to account for any statements of this kind.

⁴Notice the similarity of the definition of 'lexicalized' grammar with the offline parsibility constraint (Kaplan and Bresnan 1983). As consequences of our definition, each structure has at least one lexical item (its anchor) attached to it and all sentences are finitely ambiguous.

of the anchor for a given structure to be determined on purely linguistic grounds.

If adjunction is used as an additional operation to combine these structures, CFGs can be lexicalized. Adjunction builds a new tree from an auxiliary tree β and a tree α . It inserts an auxiliary tree into another tree (see Figure 1). Adjunction is more powerful than substitution. It can weakly simulate substitution, but it also generates languages that could not be generated with substitution.⁵

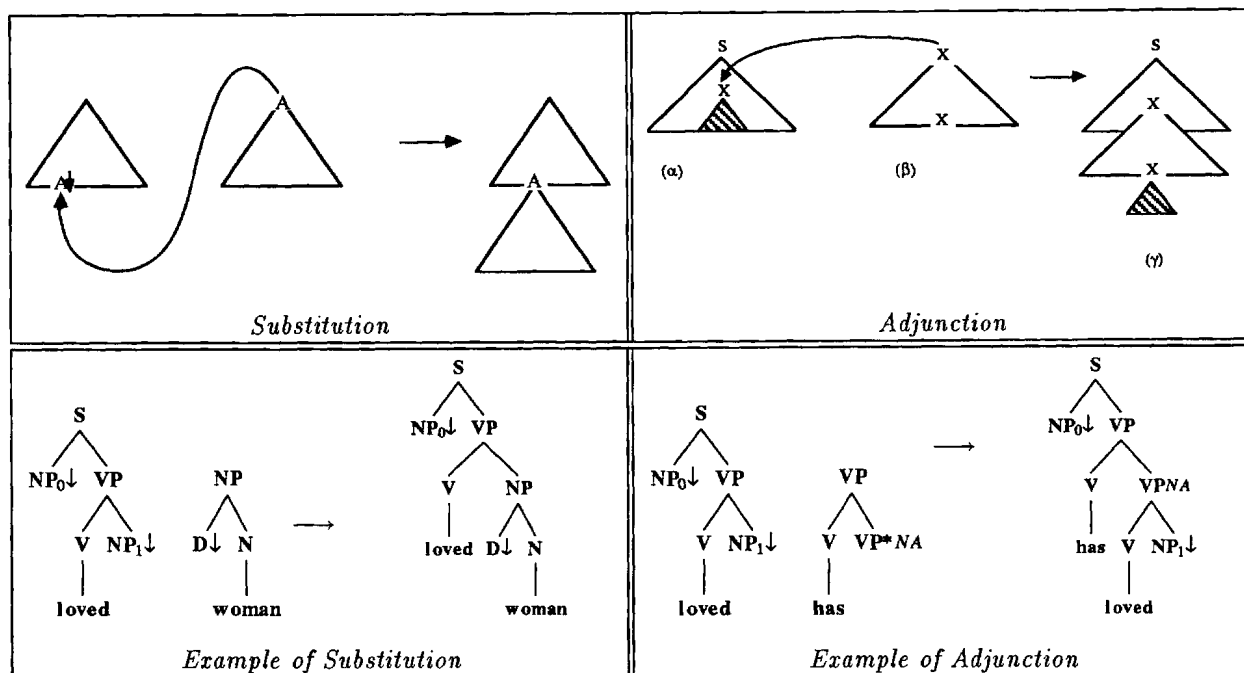


Figure 1: Combining Operations

Substitution and adjunction enable us to lexicalize CFGs. The anchors can be chosen on purely linguistic grounds (Schabes, Abeillé and Joshi, 1988). The resulting system now falls in the class of mildly context-sensitive languages (Joshi, 1985, and Joshi, Vijayshanker and Weir, 1990). Elementary structures of extended domain of locality combined with substitution and adjunction yield Lexicalized TAGs (LTAGs).

TAGs were first introduced by Joshi, Levy and Takahashi (1975) and Joshi (1985). For more details on the original definition of TAGs, we refer the reader to Joshi (1985), Kroch and Joshi (1985), or Vijay-Shanker (1987). It is known that Tree Adjoining Languages (TALs) are mildly context sensitive. TALs properly contain context-free languages.

TAGs with substitution and adjunction are naturally lexicalized because they use an extended domain of locality.⁶ A Lexicalized Tree Adjoining Grammar is a tree-based system that consists of two finite sets of trees: a set of initial trees, I and a set of auxiliary trees A (see Figure 2). The trees in $I \cup A$ are called **elementary trees**. Each elementary tree is constrained to have at least one terminal symbol which acts as its anchor.

The trees in I are called **initial trees**. Initial trees represent minimal linguistic structures which are defined to have at least one terminal at the frontier (the anchor) and to have all non-terminal nodes at the

⁵It is also possible to encode a context-free grammar with auxiliary trees using adjunction only. However, although the languages correspond, the set of trees do not correspond.

⁶In some earlier work of Joshi (1969, 1973), the use of the two operations 'adjoining' and 'replacement' (a restricted case of substitution) was investigated both mathematically and linguistically. However, these investigations dealt with string rewriting systems and not tree rewriting systems.

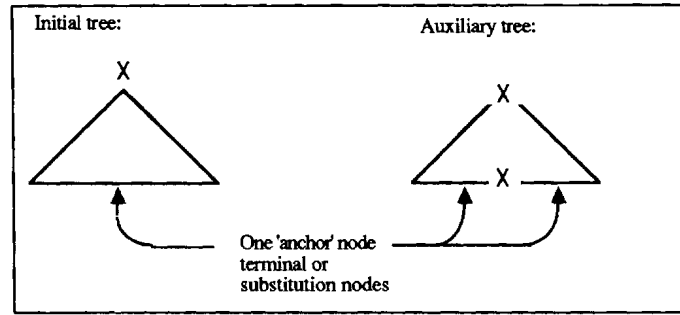


Figure 2: *Schematic initial and auxiliary trees*

frontier filled by substitution. An initial tree is called an X-type initial tree if its root is labeled with type X. All basic categories or constituents which serve as arguments to more complex initial or auxiliary trees are X-type initial trees. A particular case is the S-type initial trees (e.g. the left tree in Figure 2). They are rooted in *S*, and it is a requirement of the grammar that a valid input string has to be derived from at least one S-type initial tree.

The trees in *A* are called **auxiliary trees**. They can represent constituents that are adjuncts to basic structures (e.g. adverbials). They can also represent basic sentential structures corresponding to verbs or predicates taking sentential complements. Auxiliary trees (e.g. the right tree in Figure 3) are characterized as follows:

- internal nodes are labeled by non-terminals;
- leaf nodes are labeled by terminals or by non-terminal nodes filled by substitution except for exactly one node (called the **foot node**) labeled by a non-terminal on which only adjunction can apply; furthermore the label of the foot node is the same as the label of the root node.⁷

The **tree set** of a TAG *G*, $T(G)$ is defined to be the set of all derived trees starting from S-type initial trees in *I* whose frontier consists of terminal nodes (all substitution nodes having been filled). The **string language** generated by a TAG, $\mathcal{L}(G)$, is defined to be the set of all terminal strings on the frontier of the trees in $T(G)$.

We now define by an example the notion of derivation in a TAG. Take for example the derived tree in Figure 3.

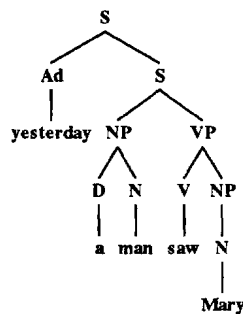


Figure 3: Derived tree for: *yesterday a man saw Mary*

⁷A null adjunction constraint (NA) is put systematically on the footnode of an auxiliary tree. This disallows adjunction of a tree on the footnode.

It has been built with the elementary trees in Figure 4.

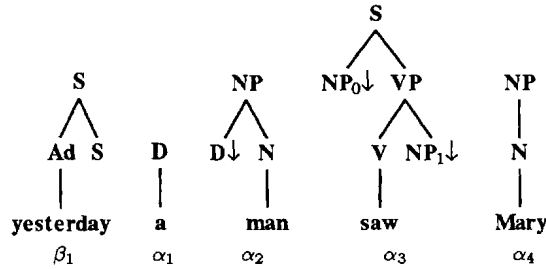


Figure 4: Some elementary trees

Unlike CFGs, from the tree obtained by derivation (called the *derived tree*) it is not always possible to know how it was constructed. The *derivation tree* is an object that specifies uniquely how a derived tree was constructed.

The root of the derivation tree is labeled by an *S*-type initial tree. All other nodes in the derivation tree are labeled by auxiliary trees in the case of adjunction or initial trees in the case of substitution. A tree address is associated with each node (except the root node) in the derivation tree. This tree address is the address of the node in the parent tree to which the adjunction or substitution has been performed. We use the following convention: trees that are adjoined to their parent tree are linked by an unbroken line to their parent, and trees that are substituted are linked by dashed lines.

The derivation tree in Figure 5 specifies how the derived tree was obtained:

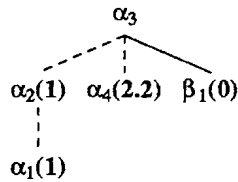


Figure 5: Derivation tree for *Yesterday a man saw Mary*

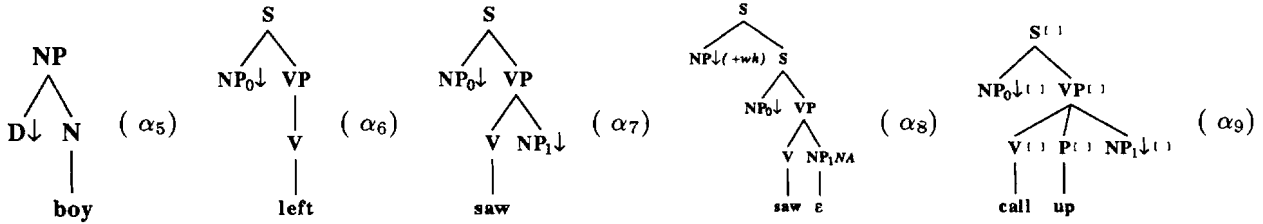
α_1 is substituted in the tree α_2 at node of address 1, α_2 substituted in the tree α_3 at address 1, α_4 is substituted in the tree α_3 at node 2 · 2 and the tree β_1 is adjoined in the tree α_3 at node 0.

By lexicalizing TAGs, lexical information is associated with the ‘production’ system encoded by the TAG trees. Therefore the computational advantages of ‘production-like’ formalisms (such as CFGs, TAGs) are kept while allowing the possibility of linking them to lexical information. Formal properties of TAGs hold for Lexicalized TAGs.

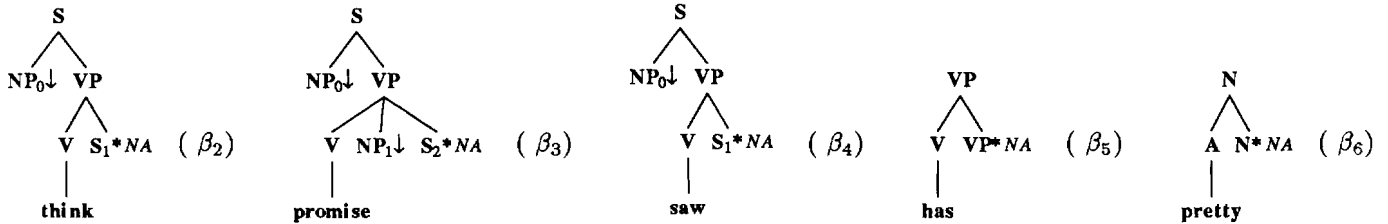
As first shown by Joshi (1985) and Kroch and Joshi (1985), the properties of TAGs permit us to encapsulate diverse syntactic phenomena in a very natural way. TAG’s extended domain of locality and its factoring of recursion from local dependencies lead, among other things, to a localization of so-called unbounded dependencies. An extended domain of locality and lexicalization of the grammar have already been shown, in Abeillé (1988c), to account for NP island constraint violations in light verb constructions. Under this analysis, extraction out of NP is correctly predicted, without the use of reanalysis. Abeillé (1988a) also uses the distinction between substitution and adjunction to capture the different extraction properties of sentential subjects and complements. The relevance of Lexicalized TAGs to idioms has been suggested by Abeillé and Schabes (1989).

Below are some examples of structures that appear in a Lexicalized TAG lexicon.

Some examples of initial trees are (for simplicity, we have omitted unification equations associated with the trees):⁸



Examples of auxiliary trees (they correspond either to predicates taking sentential complements or to modifiers):



In this approach, the category of a word is not a non-terminal symbol as in CFGs but the entire structure it selects. Furthermore, the notion of anchor in LTAG does not correspond exactly to traditional notions of head although in many cases it appears identical. In LTAG, the anchor of a phrase is intended to be the lexical source of both the elementary syntactic structure and the basic semantic structure, and may be represented by more than one word in the phrase (we think of these words as forming one multi-component anchor). Multi-component anchors are used to account naturally for phenomena like light verb constructions, verb-particle combinations (for example *call up someone*, tree α_9) and idioms.

An argument structure in a Lexicalized TAG is not just a list of arguments. It is the syntactic structure constructed with the lexical value of the predicate and with all the nodes for its arguments, eliminating the redundancy often noted between phrase structure rules and subcategorization frames.⁹

The Lexicalized TAG we are using is a feature structure-based formalism. As defined by Vijay-Shanker (1987) and Vijay-Shanker and Joshi (1988), to each adjunction node in an elementary tree two feature structures are attached: a top and a bottom.¹⁰ When the derivation is completed, the top and bottom features of all nodes are unified simultaneously. If the top and bottom features of a node do not unify, then an auxiliary tree must be adjoined at that node. This definition is easily extended to substitution nodes. To each substitution node we attach one feature structure which acts as a top feature. The updating of feature structures in the cases of adjunction and substitution is shown in Figure 6.

Following Schabes and Joshi (1989), dependencies between DAGs are expressed with unification equations¹¹ in an elementary tree. The extended domain of locality of TAGs allows us to state unification equations

⁸The trees are simplified and the feature structures on the trees are not displayed. \downarrow is the mark for substitution nodes, * is the mark for the foot node of an auxiliary tree and NA stands for null adjunction constraint. This is the only adjunction constraint not indirectly stated by feature structures. We put indices on some non-terminals to mark them as arguments (0 for subject, 1 for first object, etc.).

⁹Optional arguments are stated as such in the structure.

¹⁰The top feature structure corresponds to a view to the top of the tree from the node. The bottom feature corresponds to the view to the bottom.

¹¹Directed Acyclic Graphs which represent the feature structures.

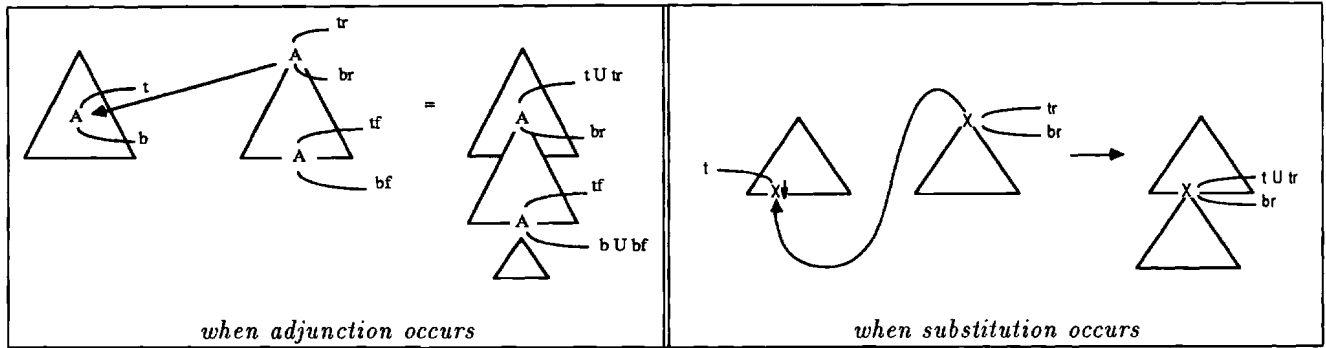


Figure 6: Updating of feature structures

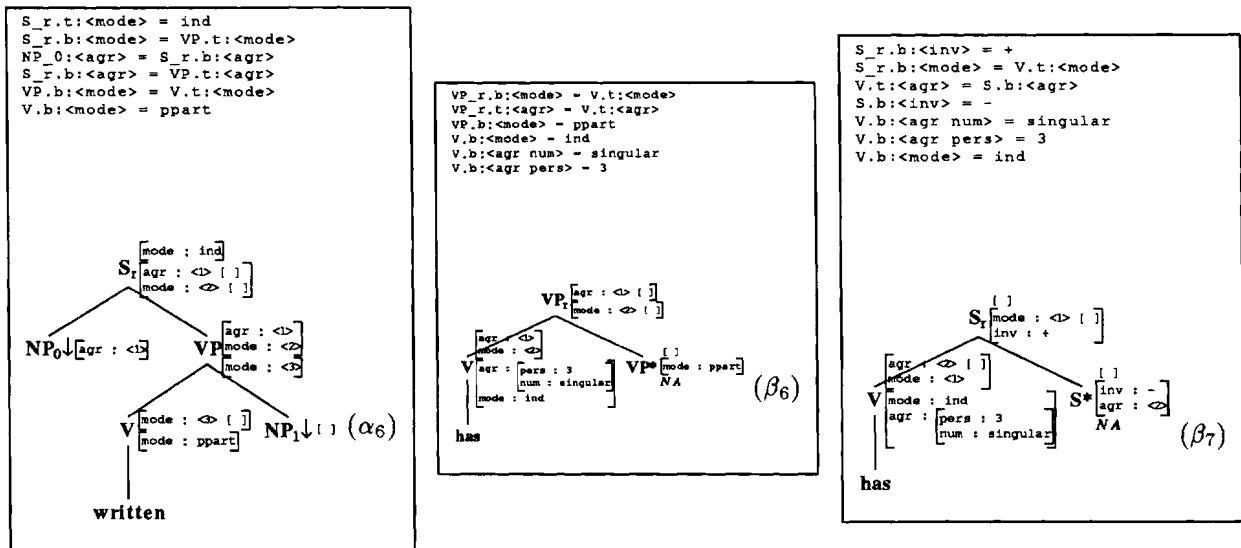


Figure 7: Examples of unification equations

between features of nodes that may not necessarily be at the same level.

The system consists of a TAG and a set of unification equations on the DAGs associated with nodes in elementary trees.

An example of the use of unification equations in TAGs is given in Figure 7.¹²

Coindexing may occur between feature structures associated with different nodes in the tree. Top or bottom features of a node are referred to by a node name (e.g. S_r)¹³ followed by $.t$ (for top) or $.b$ (for bottom). The semicolon states that the following path specified in angle brackets is relative to the specified feature structure. The feature structure of a substitution node is referred to without $.t$ or $.b$. For example, $V_b.<agr num>$ refers to the path $<agr num>$ in the bottom feature structure associated with the adjunction node labeled by V and $NP_0.<agr>$ refers to the path $<agr>$ of the substitution node labeled by NP_0 . The top and bottom feature structures of all nodes in the tree α_6 (Figure 7) cannot be simultaneously unified: if the top and bottom feature structures of S_r are unified, the $mode$ will be ind which cannot unify with $ppart$ (VP

¹²In these examples we have merged the information stated on the trees and in the lexicon. We write unification equations above the tree to which they apply. We have also printed to the right of each node the matrix representation of the top and bottom feature structures. The example is only intended to illustrate the use of unification equations and of the feature structures.

¹³We implicitly require that each node have a unique name in an elementary tree. If necessary, subscripts differentiate nodes of the same category.

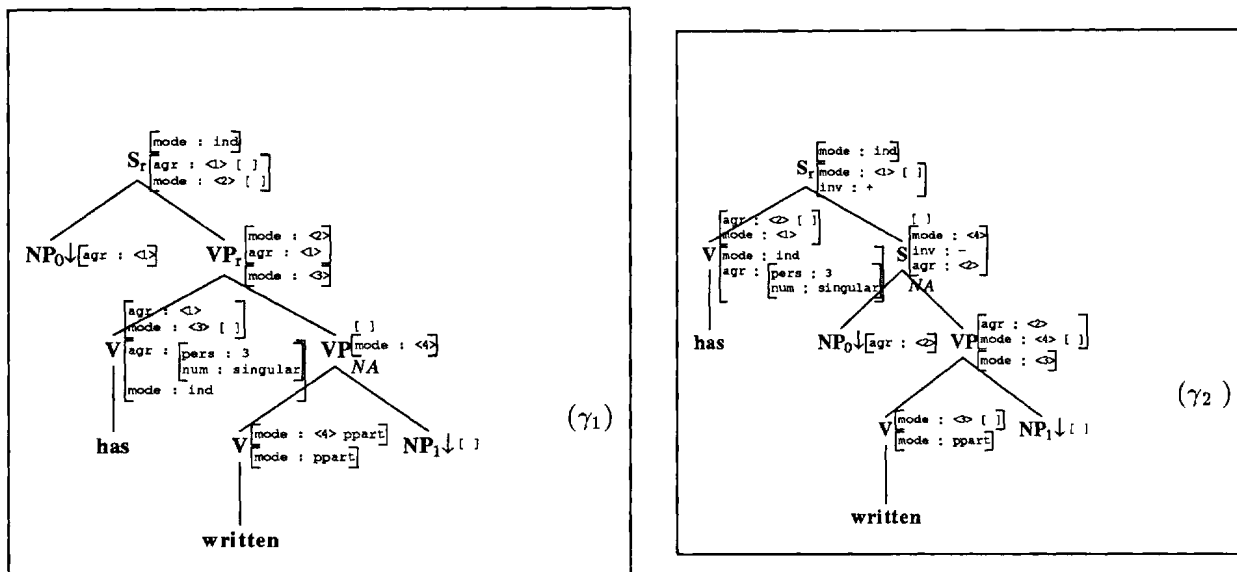


Figure 8: NP_0 has written NP_1 and Has NP_0 written NP_1 ?

and V node). This forces an adjunction to be performed on VP (e.g. adjunction of β_6 to derive a sentence like *John has written a book*) or on S (e.g. adjunction of β_7 to derive a sentence like *Has John written a book?*). The sentence *John written a book* is thus not accepted.

Notice that in the tree α_6 agreement is checked across the nodes NP_0 , S and VP . These equations handle the two cases of auxiliary: NP_0 has written NP_1 and has NP_0 written NP_1 ?. The corresponding derived trees are shown in Figure 8. γ_1 derives sentences like *John has written a book*. It is obtained by adjoining β_7 on the VP node in α_6 . γ_2 derives sentences like *Has John written a book?*. It is obtained by adjoining β_6 on the S node in α_6 . The obligatory adjunction imposed by the mode feature structure has disappeared in the derived trees γ_1 and γ_2 . However, to be completed, γ_1 and γ_2 need NP -trees to be substituted in the nodes labeled by NP (e.g. *John* and *a book*).

TAGs were proved to be parsable in polynomial time (worst case) by Vijay Shanker and Joshi 1985 and an Earley-type parser was presented by Schabes and Joshi (1988). The Earley-type algorithm for TAGs has been extended to parse Lexicalized TAGs with unification equations on elementary trees (Schabes and Joshi, 1989).

A general two-pass parsing strategy for 'lexicalized' grammars follows naturally (Schabes, Abeille and Joshi, 1988). In the first stage, the parser selects a set of elementary structures associated with the lexical items in the input sentence, and in the second stage the sentence is parsed with respect to this set. Schabes and Joshi (1989) discuss in more detail the relevance of lexicalization to parsing.

Several software packages running on Symbolics machines have been developed to aid in the implementation of grammars in the LTAG framework. The Lexical Interface (Bishop, 1990) helps the user to add entries to the lexicons. The TAG System (Schabes, 1989) includes a graphical tree editor and a TAG parser (Schabes and Joshi, 1988) itself, and gives access to the grammar files. This system provides an environment for building and editing trees, editing the lexicons, and testing possible derivations both by hand and through the parser.

In this paper, we present the structure and current linguistic coverage of a feature-based English LTAG, noting the natural distinction made in a TAG between arguments and adjuncts. We rely heavily on the linguistic work that has been done in this framework on English by Joshi (1985) and Kroch and Joshi (1985), Kroch (1987), and for French by Abeillé (1988a, 1988b and 1988c). We have also benefited from the work

of Maurice Gross on Lexicon-Grammars for Romance languages.

The English LTAG currently contains about 3000 lexical entries. There are about 80 tree families for predicate terms, each containing at most 12 different trees. The total number of elementary trees is approximately 600. A more detailed breakdown of the current size of the grammar is available in the appendices to this report.

2 Organization of the Grammar

The grammar consists of a morphological lexicon, which lists the possible morphological variations for a word, and a syntactic lexicon, which is the domain of structural choice, subcategorization and selectional information. Currently, we are including only syntactic information but much semantic information will also ultimately be contained here. Lexical items are defined by the tree structure or the set of tree structures they anchor. One of the choices made in the organization of the lexicon was to mark predicative nouns and adjectives for their ability to act as part of a multi-component anchor for a sentential clause, as was done by Abeillé (1988c) for ‘light-verb’ constructions. Nouns, adjectives and prepositions can also state subcategorization constraints about their arguments. Nouns and adjectives that do not act as Predicate Nominals or Predicate Adjectives in any way have simple entries that are NP-type or AP-type trees. Adjectives used as modifiers of N define N-type auxiliary trees. Other lexical categories such as determiners and adverbs also have simple entries.

2.1 Syntax

The entries for anchors of all types belong to the syntactic lexicon and are marked with features to constrain the form of their arguments. For example, a verb which takes a sentential argument uses features to constrain the form of the verb acceptable in the complement clause. An interesting consequence of TAG’s extended domain of locality is that features imposed by a clausal anchor can be stated directly on the subject node as well as on the object node. These features need not be percolated through the VP node as in context-free formalisms.

When a word can have several structures, corresponding to different meanings, it is treated as several lexical items with different entries in the syntactic lexicon. Morphologically, such items can have the same category and the same entry in the morphological lexicon¹⁴:

```
\DEMONSTRATE\,V: Tnx0V(pnx1)[p=against/for] (verb with prep complement: to protest in the streets)
\DEMONSTRATE\,V: Tnx0Vnx1, Tnx0Vs1 (verb with np or s complement: to show)
```

```
\DECLARE\, V: Tnx0Vs1 (verb with S complement: to state)
\DECLARE\, V: Tnx0Vnx1 (verb with NP complement: to admit possession)
```

```
\RESOLUTION\,N: αNXdn(s1) (noun with S complement: decision)
\RESOLUTION\,N: αNXdn (intransitive noun: precision of an image)
```

```
\WILL\, N: αNXdn(s1) (noun with S complement: mental strength/desire)
\WILL\,N: αNXdn (intransitive noun: legal document)
```

The lexical entries can be differentiated by the basic type of their trees or by the feature structures they impose upon their arguments.

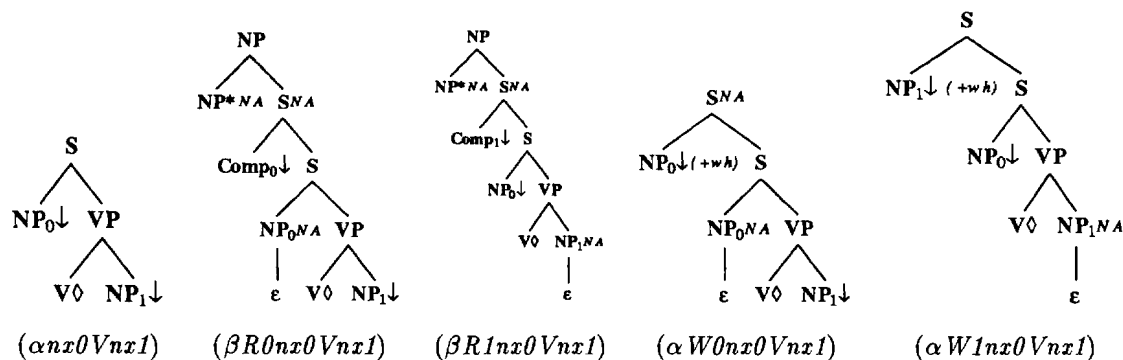
¹⁴The lexical entries below are simplified for the purpose of exposition; for example, feature structures are ignored and light verb constructions for nouns too.

An elementary tree is either a complete argument structure representing a predicate or an adjunct, or the maximal projection of a category that represents an argument. The predicate structures will be explained in more detail below. Adjuncts (such as adjectives, adverbs and relative clauses) are represented as auxiliary trees: they thus have in their structures not only their argument structure (if any) but also a root node and foot node that match the category of the word they modify. Arguments are usually either nominal or sentential. Nominal arguments are always initial trees that are substituted in the elementary tree of their predicate. Sentential arguments can either be substituted in or receive the adjunction of their predicate.

We consider sentential clauses to be elementary trees, usually anchored by their main verb. In cases of light verb constructions or idioms, a multicomponent anchor comprised not only of the verb but of nouns, adjectives or adverbs as well is defined (see below). A sentential tree is the proper (full) syntactic category of the corresponding predicate in the lexicon¹⁵. Notice that we consider the subject to be selected for by a predicate in the same way as the other arguments. The distinction between verbs that select sentential subjects and those that don't is thus made in the same way as the distinction between verbs that select sentential objects and those that don't.

Lexical items which are clausal anchors usually select a set of trees, called a **tree-family**. In essence, a tree-family is a set of elementary trees corresponding to different syntactic structures that share the same subcategorization type. Because all trees of a type are grouped together, LTAG does not need to specify subcategorization frames separately. Argument structure is represented in the groupings of such syntactic structures. As a result, predicate terms, which are each defined by one or more tree families, have associated with them both complete syntactic structure type(s), and complete predicate argument structure(s). This arrangement takes advantage of TAG's extended domain of locality and allows subcategorization to be satisfied in some locality domain.

For example, a verb like *eat*, which takes a nominal subject and a nominal object, selects the transitive tree family $Tnx_0 Vnx_1$:¹⁶ The following trees are some of the members of this tree family:



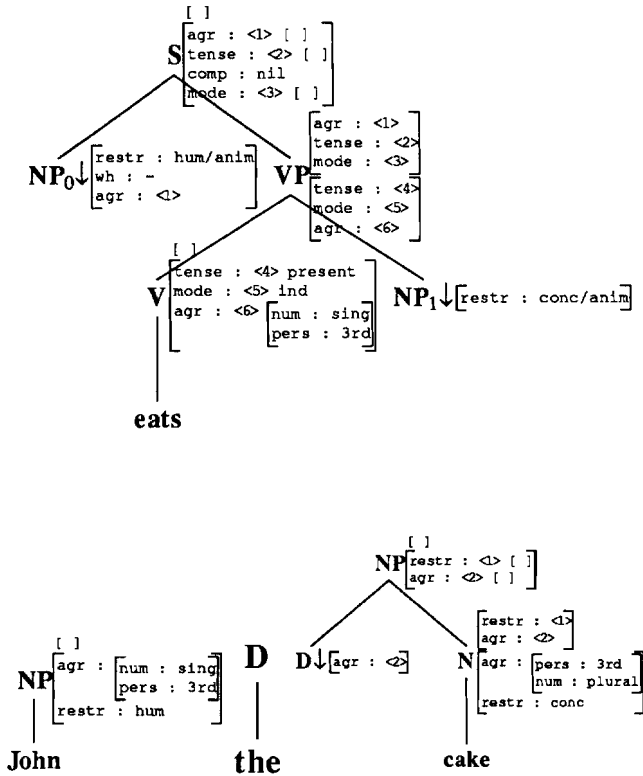
$\alpha nx_0 Vnx_1$ is an initial tree corresponding to the declarative sentence, $\beta R_0 nx_0 Vnx_1$ is an auxiliary tree corresponding to a relative clause where the subject has been relativized, $\beta R_1 nx_0 Vnx_1$ corresponds to the relative clause where the object has been relativized, $\alpha W_0 nx_0 Vnx_1$ is an initial tree corresponding to a wh-question on the subject, $\alpha W_1 nx_0 Vnx_1$ corresponds to a wh-question on the object.

At the level of description of the tree families, the grammar is not lexicalized since the trees in a tree family are not associated with a lexical item. This higher (abstract) level of description does not violate the lexicalized condition since when a tree in a tree family is selected, its anchor will be lexically inserted. At the level of description of the tree families, the rules used to build the tree families need not be lexicalized.

¹⁵This is similar to the approach adopted by Pollard and Sag (1987) and Abeillé 1988a.

¹⁶The trees shown here are simplified (the feature structures are not shown). \diamond is the mark for the node under which the anchor word of the tree is attached. See Appendix 3 for the list of tree families.

When selected by the verb *eat*, $\alpha n x_0 v n x_1$ is a concrete tree which combines with argument trees to generate simple sentences like, ‘John eats the cake’, as shown below.



For a complete listing of the tree families currently used see Appendix 3.

Finally, note that lexical variations may also be specified for a tree family. For instance, a particular verb may select certain trees in a given tree family and not others. It has been observed, for example, that the passive, dative movement, and ergative alternation apply with many lexical idiosyncrasies. Therefore, lexical items may have features associated with them to limit the trees selected within a tree family.

2.2 Interaction of the Morphological and Syntactic Lexicons

The English LTAG is organized as two interacting lexicons of different types: a morphological lexicon and a syntactic lexicon.

Words are marked with appropriate morphological features in the morphological lexicon. The **morphological lexicon** associates a word with a abstract class of words (class of the inflected forms of the same morphological root), a preterminal symbol (the category of the word) and a set of morphological features. For instance, each morphological form for a verb is marked with the relevant value for the <Mode> and <Agr> features, as in the example below:¹⁷

¹⁷Note that these examples represent only a subset of the morphological entries.

went : \GO\,V {(mode)=ind, (tense)=past}.
 goes: \GO\,V {(mode)=ind, (tense)=pres, (agr num)=sing, (agr pers)=3}.
 go: \GO\,V {(mode)=base}.
 gone: \GO\,V {(mode)=ppart}.
 going: \GO\, V {(mode)=ger}.

The **syntactic lexicon** describes the structures for the abstract \GO\, thus predicting that the following entries will have the same morphological properties, despite their syntactic and semantic differences:

Syntactic Lexicon

\GO\,V: TnxOV(pnx1). (John went to the store)
 \GO\,V : TnxOVa. (John went crazy)
 \GO\,V : TnxOVpdn1 [p=to, d1=the, n1=dogs]. (idiom: The project went to the dogs)

Occasionally, differences in syntactico-semantic properties reflect upon the morphology. These cases are handled by distinguishing the different syntactic entries by number, and choosing only the appropriate references in the morphological lexicon:

Morphological Lexicon

can : \CAN-1\,V {(mode)=base}.
 cans : \CAN-1\,V {(mode)=ind, (tense)=pres, (agr num)=sing, (agr pers)=3}.
 canned : \CAN-1\,V{(mode)=ind, (tense)=past}.
 can : \CAN-2\,V{(mode)=ind, (tense)=pres}.
 could : \CAN-2\,V{(mode)=ind, (tense)=past}.

Syntactic Lexicon

\CAN-1\,V: TnxOVnx1. (to make preserves)
 \CAN-2\,V: β vVP. (the auxiliary verb)

When a word has no morphological variations, it is entered directly into the syntactic lexicon. This is the case for adjectives, prepositions, determiners and complementizers in English. Nouns usually vary in number, and the two alternative forms are listed in the morphological lexicon. The few nouns that do not exhibit such variation, are entered directly in the syntactic lexicon:

Morphological Lexicon

man : \MAN\,N {(agr num)=sing}.
 men : \MAN\,N {(agr num)=plur}.

Syntactic Lexicon

\MAN\,N: α NXd_n, α NXn{N.b : (agr num)=plur}.
 luggage,N {(agr num)=sing}: α NXd_n, α NXn.

A distinction is also made between homonymous nouns on the basis of the morphological variations they do or do not exhibit:

Morphological Lexicon

hair : \HAIR\, N{(agr num)=sing}.
 hairs: \HAIR\, N{(agr num)=plur}.

quarter : \QUARTER\, N {(agr num)=sing}.
 quarters: \QUARTER\,N {(agr num)=plur}.

Syntactic Lexicon

hair, N{(agr num)=sing} : α NXd_n, α NXn. (collective human hair)

\HAIR\, N: $\alpha NXdn$, $\alpha NXn\{N.b:\langle agr\ num\rangle=plur\}$. (individual hair(s))

\QUARTER\,N: $\alpha NXdn$, $\alpha NXn\{N.b\langle agr\ num\rangle=plur\}$. (25 c coin)

quarters,N{\langle agr num\rangle =pl } : $\alpha NXdn$. (living quarters)

A given word is thus considered an autonomous entry in either the morphological lexicon (for most inflected forms), or in the syntactic one (for most lemmas and words without inflectional variations). When an inflected form is part of a regular morphological family but has a unique syntactic construction or unique semantic meaning (eg 'quarters'), it will have autonomous entries in both lexicons. Words that do not have inflectional variations and occur only as parts of idiomatic constructions (such as 'umbrage') will not have autonomous entries.

3 Sentential Complements

3.1 S or VP arguments?

We have chosen in our grammar not to use VP arguments. All arguments anchored in a verb (or multi-component anchor including a verb) are treated as sentences. Other grammars, such as LFG and GPSG, posit the existence of VP arguments for cases where there is no overt subject. There is a long history and a large literature on the right representation of these cases. We have adopted the GB approach making use of PRO subjects both because of the theoretical generalizations that it allows and for practical reasons.

Some of the theoretical reasons are discussed below. Although many earlier works have presented strong arguments for considering so-called VP-complements as S-complements, we will review here just a few types of examples. A full discussion of the theoretical benefits of S-complements is outside the scope of this report.

It has been observed (eg., by Borsley (1984)) that infinitival complements are subcategorized for by the same verbs as sentential ones and can be coordinated with them:

- (1) *John expects that he will see Mary today.*
- (2) *John expects to see Mary today.*
- (3) *John expects to be hired and that Mary will be his boss.*

- (4) *John wonders whether to go to Macy's (or not).*
- (5) *John wonders whether he should go to Macy's (or not).*
- (6) *John wonders whether to go to Macy's and whether Mary will notice him.*

Although an imperfect test, coordination is often an indication of similarity of syntactic categories. Notice also that in the second example, *Whether*, which is considered to be either a complementizer (as in LTAG) or a Wh-term of some type, dominates both the tensed and the untensed clauses.

A similar phenomenon can be seen to exist between infinitival clauses and 'for' clauses, as shown below:

- (7) *John prefers to go.*
- (8) *John prefers for Mary to go.*
- (9) *John prefers that Mary leaves early.*

- (10) *John condescends to come.*
- (11) *John condescends for Mary to come.*
- (12) * *John condescends that Mary leaves early.*

If we consider *for* a complementizer (which will account for its being in sentential subjects as well), then the NP following it is the subject of the infinitival. There are of course cases in which such alternations do not hold:

- (13) *John happened/ceased/began to talk.*
- (14) * *John happened/ceased/began for Mary to talk.*
- (15) * *John happened/ceased/began that Mary talked.*

All such verbs can be shown to be raising verbs in the sense that they do not provide an independent thematic role for any subject and therefore either take *there* as subject or 'share' the subject of the embedded sentence (even with idioms).

- (16) *There happened/ceased/began to be a problem.*

(17) *Close tabs happened/ceased/began to be kept on Mary.*

For gerunds, the same parallels with tensed S-complements hold. An additional significant parallel holds between prepositional gerunds and that-clauses. This was first mentioned by Rosenbaum (1967) and was more recently studied by Freckelton (1984):

- (18) *John insisted on going to the beach.*
- (19) *John insisted that we go to the beach.*
- (20) *Going to the beach was insisted on (by John).*
- (21) * *That we go to the beach was insisted (by John).*
- (22) *That we go to the beach was insisted on (by John).*

Notice that, although two different subcategorization frames seem to be involved in the active sentences, passivization shows that the that-clause is in fact to be analysed as a prepositional clause with (Prep + that) being reduced to *that*. The tensed clause therefore does alternate with the gerund clause which is thus considered a sentential clause as well.

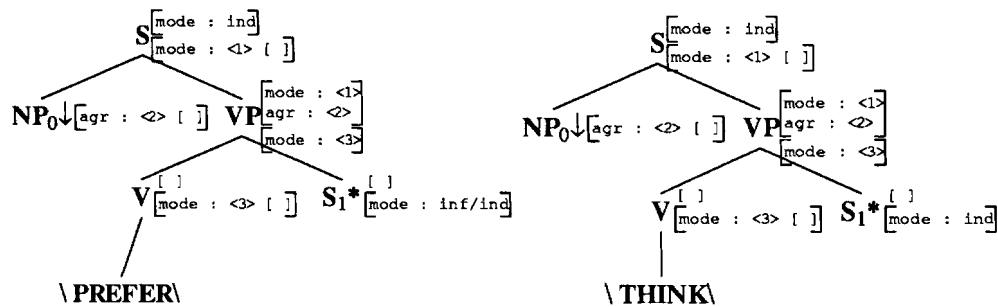
Again, the few predicates that take only gerunds can be shown to have raising properties:

- (23) *John stopped/quit lying to Mary.*
- (24) * *John stopped/quit that he lies to Mary / for Mary to be angry.*
- (25) *It stopped/quit raining in BC.*
- (26) *There stopped/quit being troubles around here.*

As mentioned above, there are also some practical benefits to adopting the S-complement approach for infinitivals and gerunds. First, the same basic elementary trees used to represent tensed clauses can be used to represent infinitivals as well, making the grammar more efficient. Second, this approach is the only one consistent with earlier work on English TAGs, if one wants to account for extraction out of infinitivals and gerunds (see next subsection).

3.2 Extraction properties

Treating gerunds, infinitival complements and that-clauses as sentential trees allows us to define sentential auxiliary trees for the tree families of verbs that take these forms as complements. For example, the tree family for the verbs *think* and *prefer* would include the following trees:

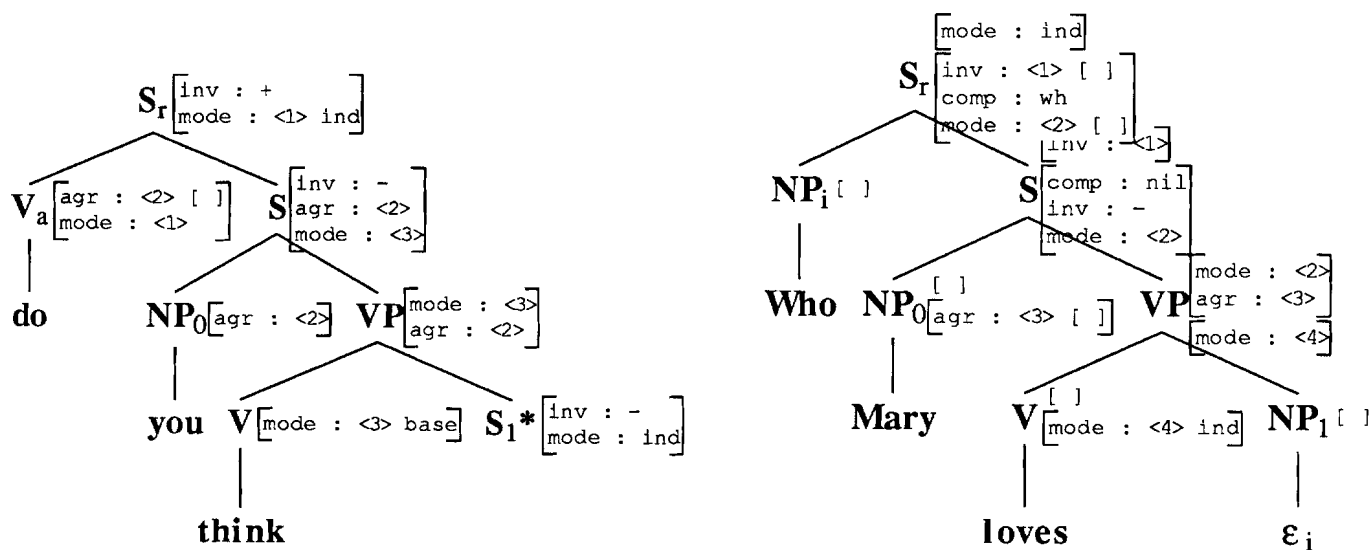


Such a representation has been shown by Kroch and Joshi (1985) to be exactly what one needs for a ‘natural’ account of unbounded dependencies.

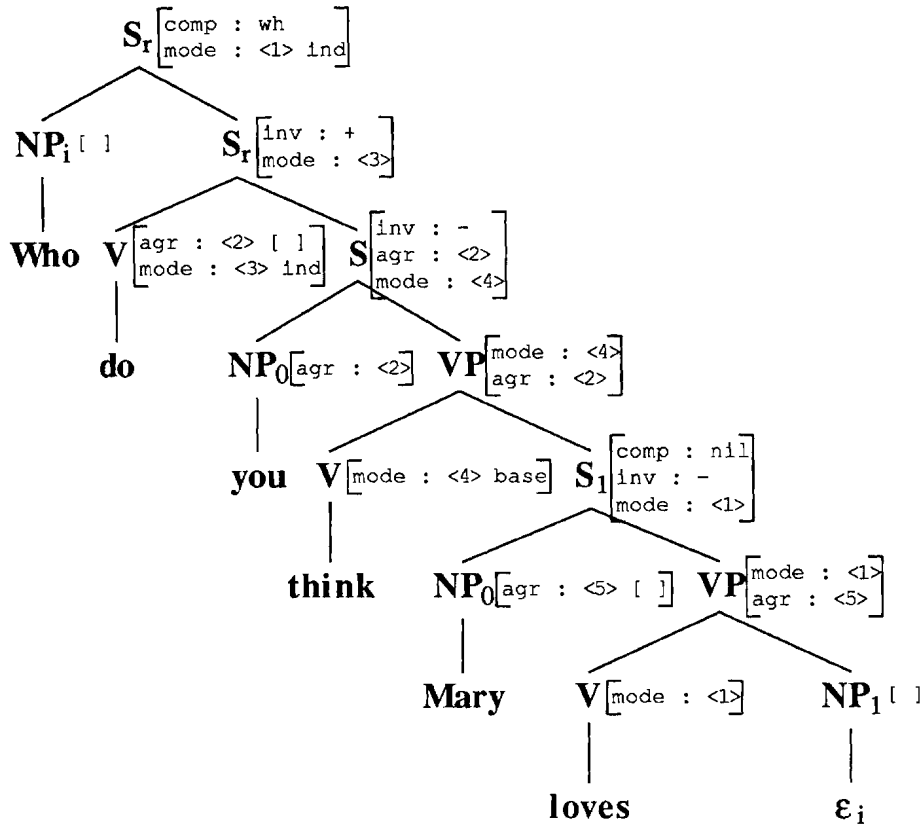
Following Kroch and Joshi (1985), extraction out of sentential complements is accounted for in terms of elementary structures. Complement clauses are represented as initial sentential trees, and matrix clause auxiliary trees may adjoin to them. Since adjunction can happen at the internal S-node of a wh-tree, extraction is predicted with the matrix clause getting inserted between the wh-element and the rest of the complement clause. Adjunction allows this insertion of matrix clauses to be recursive.

This analysis has numerous advantages. First, filler-gap relations are localized because the wh-element belongs to the same tree that its trace is an argument of. There is no need for ad hoc procedures to compute where a wh-element comes from or where it has to be passed to in the case of unbounded dependencies. For example, devices such as functional uncertainty used in LFG become a mere corollary in a TAG (Joshi and Vijay Shanker, 1989).

The derivation of the sentence, "Who_i do you think Mary loved ϵ_i ?" starts with structures shown below:



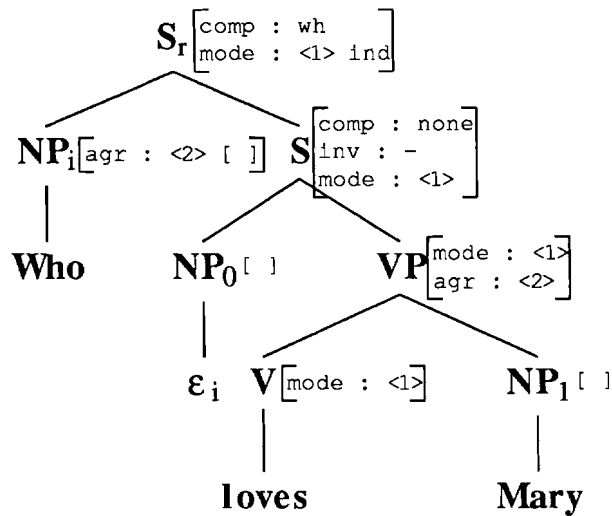
Note that the top and bottom values of the **inv** feature on node S_r in the second tree do not unify, forming an obligatory adjunction constraint. The resulting structure for that sentence is below:



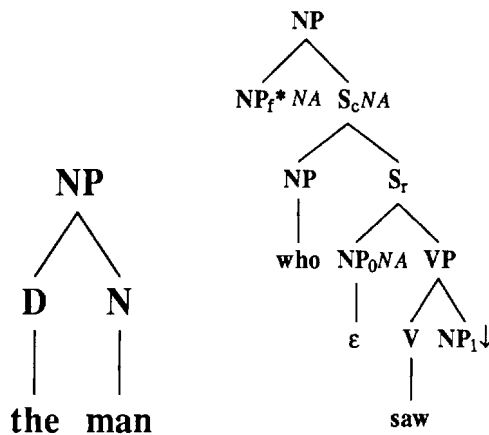
Who do you think Mary loves?

Recursive adjunction provides derivations for the sentences “Who do you think Bob said Mary loves?”, “Who do you think Anne believes Bob said Mary loves”, and so on.

ECP may be implemented either as a constraint on the form of initial trees, or as a feature constraint on the types of auxiliary trees that can adjoin to wh-trees. Our current approach is to specify $\langle \text{COMP} \rangle = \text{none}$ (described further below) on the root node of tree-structures containing subject gaps (see below), so that a sentence such as **Who_i do you think that ε_i loves Mary?* can not be generated.



Extraction properties are also accounted for as constraints on the structure of the elementary trees, as was first shown by Kroch 1987. In the case of relative clauses, they follow directly from the structure of the elementary trees themselves :

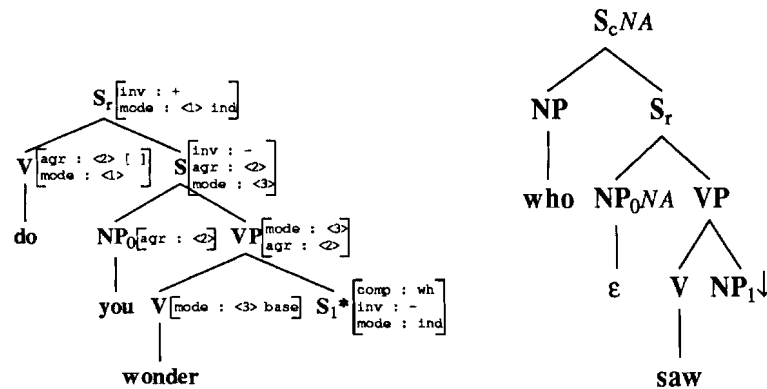


Extraction out of relative clauses is thus ruled out because there is no way a sentence like:

(27) * *Who_i did you meet the man who loves e_i ?*

could be derived, with such elementary trees, without either loosing the filler-gap relation or the desired word order.

In the case of indirect questions, subjacency follows from the principle that a given tree can not contain more than one wh-element:

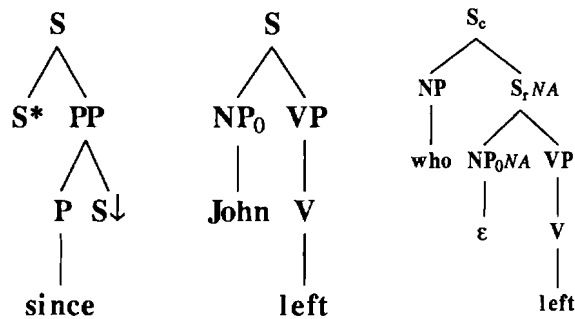


Extraction out of an indirect question is ruled out because a sentence like:

(28) * *Who_i do you wonder who loves e_i ?*

would have to be derived from the adjunction of 'do you wonder' into 'who_i who loves e_i' that is an ill-formed elementary tree.¹⁸

Extraction can also be ruled out by using substitution, which is forced to happen at leaf nodes only, instead of adjunction for combining sentential structures (Abeille, 1988). Extraction out of adjunct clauses, for example, is thus ruled out :



Thus the string 'who_i since e_i left' cannot be generated, although the echo-question, '... since who_i e_i left?' would be fine. Notice that here using substitution instead of adjunction is not an extra stipulation, it is imposed by the formalism, since otherwise the tree for 'since' would have two footnodes and would be thus ill-formed.

A similar device is also used for sentential subjects. It has long been observed that sentential subjects resist extraction (Ross, 1967). But it has less often been noted that extraposed subjects may allow it :

(29) *Going to the beach pleases John.*

¹⁸This does not mean that elementary trees with more than one gap should necessarily be ruled out. Such an option might actually be considered for dealing with parasitic gaps or gaps in coordinated structures.

- (30) * *Where does going (to) please John ?*
 (31) *It pleases John to go to the beach.*
 (32) ? *Where does it please John to go (to) ?*

In the family of the verb *please* with a sentential subject, the tree for the non-extraposed case will be an initial tree (ruling out extraction) whereas the tree for extraposed subject will be an auxiliary one (allowing for it).

A further distinction could be made between verbs that do allow extraction out of their sentential complements and those which don't :

- (33) *John said that he hit Mary.*
 (34) *Which woman did John say that he hit ?*
 (35) *John stammered that he hit Mary.*
 (36) * *Which woman did John stammer that he hit ?* (Erteschik, 1973)
 (37) *John answered that he hit Mary.*
 (38) * *Which woman did John answer that he hit ?* (Culicover and Wilkins, 1984)

Such phenomena require further study; but if the non-extractability is regular for all contexts of a given verb (and such seems to be the case for *stammer*), the corresponding tree family will probably be a different one with the complement clause being a substitution node rather than an adjunction node.

3.3 Selecting the Appropriate Type of Sentential Argument

Verbs that take sentential arguments may have basic constraints on the verb form and choice of complementizer in these arguments.¹⁹ For example, the verb *likes*, which takes an infinitive or a gerundive complement, will require that the highest VP node in the complement be anchored by either a verb in *-ing* form or *to*.²⁰ *Likes* will, of course, also need to require in these cases that the subordinate clause not have a complementizer. The feature $\langle \text{MODE} \rangle$ is used to constrain the verb form at the top of the embedded VP. This feature actually conflates a couple of different types of information (mainly, verb form and sentence mood), and will eventually need to be re-analyzed. The $\langle \text{COMP} \rangle$ feature constrains the presence and choice of complementizers. The exact use of these features is described in Appendix B.

For verbs taking prepositional sentential complements, there are no lexical variations regarding $\langle \text{Comp} \rangle$ and $\langle \text{Mode} \rangle$. Their values (respectively *none* and *gerund*) are thus stated directly at the level of tree families without appearing in the lexical entry of the matrix verb.

However, verbs that take direct sentential complements may vary widely (though within constraints) in the values they assign for these features. *Think*, for example, requires either *none* and an infinitival complement, or *that* or *none* and an indicative complement. *Wonder*, on the other hand, though it has same argument structure and thus selects the same tree family, takes only indirect wh-questions or *whether* clauses. Such constraints are stated by the verbs in the syntactic lexicon.

$\langle \text{MODE} \rangle$ and $\langle \text{COMP} \rangle$ of sentential arguments are also selected by nouns and adjectives taking sentential complements. The noun *fact* takes only that-clauses, the noun *question* only wh-clauses, and a noun like *urge* infinitival complements. These features can also be imposed by prepositions heading subordinate clauses. *Because*, for example, requires that the mode of the clause be indicative, while *after* allows indicative or gerundive complements:

¹⁹Other considerations, such as the relationship between the tense/aspect of the matrix clause and the tense/aspect of a complement clause are also important but are not covered at this time

²⁰Note that we do not make use of an INFL node and therefore treat *to* as an auxiliary verb.

- (39) *John is happy because he got a job.*
- (40) * *John is happy because getting a job.*
- (41) * *John is happy because to get a job.*

- (42) *After he killed Mary, John was unhappy.*
- (43) *After killing Mary, John was unhappy.*

As shown below, there are further variations, at least for verbs, depending on whether the context is interrogative, or negative, or neutral:

- (44) *John said that Mary was coming.*
- (45) ??*John said whether Mary was coming.*
- (46) *Did John say whether Mary was coming ?*
- (47) *John did not say whether Mary was coming.*

Other feature structures will be needed to capture these constraints on the tree for *say*. But notice that the possibility of such variation is by itself lexically determined:

- (48) *John prefers that Mary leaves early.*
- (49) * *John prefers whether Mary leaves early.*
- (50) * *Who prefers whether Mary leaves early ?*
- (51) * *John did not prefer whether Mary leaves early.*

4 Auxiliaries

We use the term ‘auxiliaries’ to refer to all of the following types of verbs:

the modals (*will, would, can, could, may, and might*),
the forms of *have* followed by a past participle,
the forms of *be* followed by a gerund or a passive participle,
do followed by the base form of a verb, and
to.²¹

The differences between these types is expressed by the specific tense or type of aspectual information they supply to a sentence.

Following McCawley (1975) and others, auxiliaries are not considered to be morphologically different from main verbs. But their syntax, i.e. the elementary trees selected by them, is quite different. We want to account for the following properties:

- Auxiliaries allow for questions and inverted structures
- Auxiliaries allow for sentential negation²²
- Recursive insertion of auxiliaries is very restricted
- Auxiliaries may agree with the subject of the sentence
- Auxiliaries do not subcategorize for a sentence

The first three properties are proper to auxiliaries. They share the fourth with all tensed verbs and, in at least one analysis of raising-verbs,²³ they share the last property with them.

Auxiliaries in the English LTAG are added into a sentence by adjunction, allowing all of the necessary properties to be achieved. A main verb anchors an entire sentence structure.²⁴ If the main verb is tensed, tense and agreement are represented directly in the elementary structure, and feature constraints prevent auxiliaries from adjoining into these structures. If, however, the main verb is untensed, a compatible auxiliary may adjoin. If that auxiliary is tensed or *to*, no further adjunctions are possible. If it is any other form, additional auxiliary adjunctions are possible.

The adjunction trees for auxiliaries are rooted in VP or S (the S tree is defined only for tensed auxiliaries), depending on whether or not they are being used to form an inverted sentence. Under this analysis, inversion is never present in an elementary S-tree. We can therefore use the same elementary trees for direct and indirect questions. Some example trees for auxiliaries, along with two transitive S-trees, are shown below:²⁵

²¹Because we are so far concerned only with ‘standard’ American English we have not added *dare, ought* and *need*. However, we would do so to account for all of the following sentences:

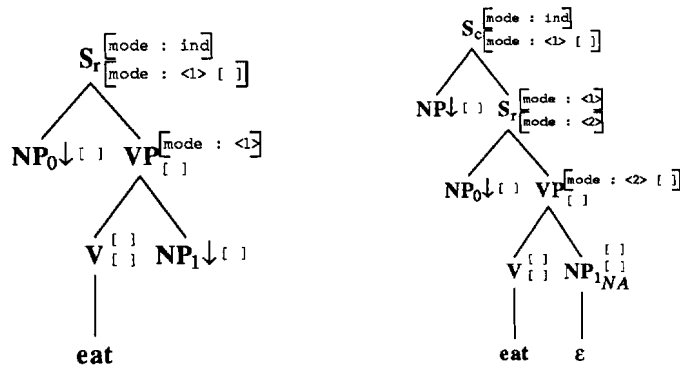
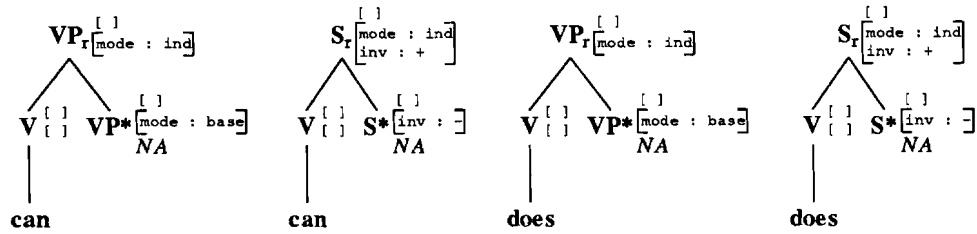
Need I tell Tom about that ? How dare you borrow this without my permission ! I need not tell you this. ? I dare not say this, but... Ought we to do this at once?

²²This property will not be developed here (see ongoing work section)

²³See the section on Ongoing Work for a discussion of these verbs.

²⁴This is a simplification. Actually, an S may be anchored by any predicate term and, in the case of light-verb constructions (as discussed in another section of this report), S is anchored by a multi-component anchor. This distinction does not bear on our discussion of auxiliaries.

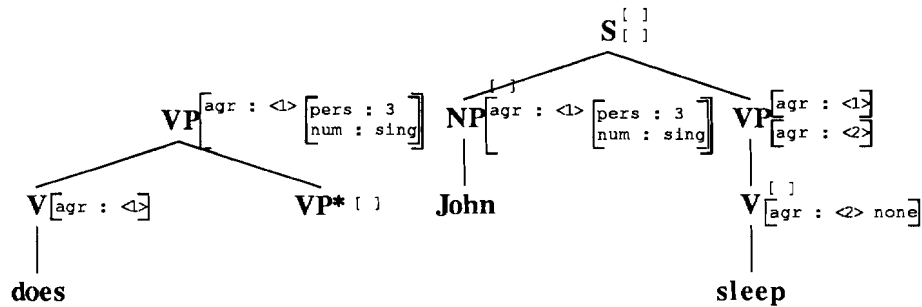
²⁵Another possible motivation for having auxiliaries adjoin to VP is to account for cases of VP ellipsis, but the corresponding tree structures are not implemented in our grammar yet.



We thus account straightforwardly for the difference in acceptability between the two pairs of sentences below because, in the second set, there is no S-structure available for ‘what’ and ‘John’:

- (52) *What does John eat?*
- (53) *What can John eat?*
- (54) * *What does John?*
- (55) * *What can John ?*

Agreement is co-indexed between the subject, the top VP-node, and S. All untensed verbs are unspecified for agreement and therefore place no agreement constraints on the subject. Adjunction of a tensed auxiliary verb changes features on either the top VP-node or S, triggering the required agreement, as shown below:



4.1 Passives

The analysis of the passive developed by Megan Moser requires additional feature constraints to be built into the auxiliary structures. The elementary tree for the passive is a sentential tree rooted in a $\langle \text{MODE} \rangle = \text{PPART}$ form of a main verb. Passive trees are added within the same tree family that contains the active structures for a particular subcategorization frame. The subject of a passive is labelled NP1 instead of NP0, so any constraints that a particular verb places on NP1 for that tree family, will apply to the object position for active sentences and to the subject position for passive sentences.

With the addition of a passive feature, auxiliary structures can adjoin into passive structures just as they do into active structures.

4.2 Ordering Auxiliaries and Specifying Verb Form in Sentential Arguments

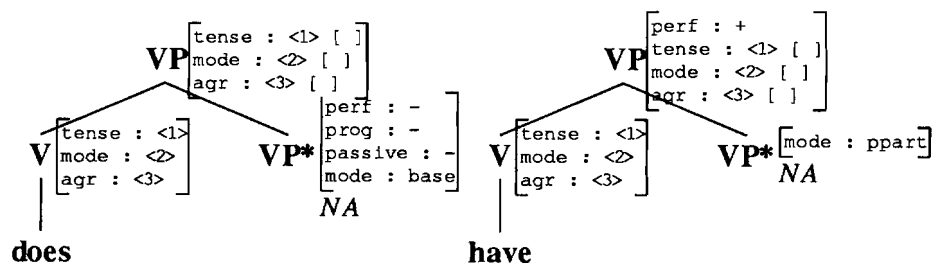
The canonical order for English auxiliaries is :

Modal base perfective progressive passive

Unlike the French modals, for example, English modals cannot be inserted recursively, and the longest acceptable string is of the type:

could have been being written

For a complement clause the ultimate form of the verb phrase (i.e., the form of the leftmost verb) is determined by the matrix verb. In a matrix clause, the leftmost (highest) verb is stipulated to have the feature $\langle \text{MODE} \rangle = \text{indicative}$. An indicative verb form carries the tense, and also agreement features which must match those of the subject. A to-infinitive (*to* + base) has no tense. Each auxiliary form determines which verb can follow immediately after it. This constraint is represented by having each auxiliary form impose the relevant mode, tense and aspect features on the foot node of its auxiliary tree. The root node of the auxiliary tree takes its features from the auxiliary verb itself:



5 Raising constructions

We represent raising verbs with the same VP-trees as auxiliaries²⁶, both for active and passive raising verbs.

The major differences between raising verbs and auxiliaries are captured by the different sets of trees associated with them in the grammar. They both select VP-auxiliary trees, but only auxiliaries also select S-auxiliary trees for inverted constructions. The VP-trees also bear different features accounting for limited recursion of the auxiliaries and ‘infinite’ recursion for the raising verbs:

(56) *John seems to tend to lie.*

(57) *John happened to seem to know about that.*

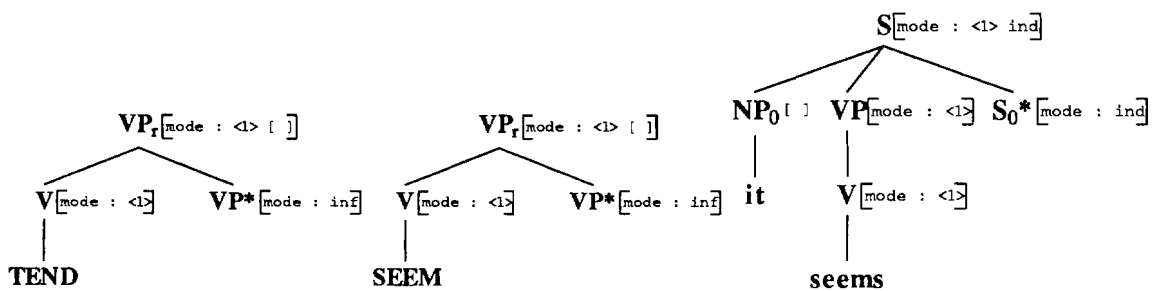
As discussed in the previous section, auxiliaries are reduced to their auxiliary structures, rooted in VP or S, depending on inversion. Raising verbs on the other hand can usually also be used as main verbs and will therefore select at least one subcategorization frame. They thus usually select a tree family that contains at least one sentential tree²⁷. Most raising verbs allow for sentential arguments, sentential subjects (usually extraposed) for verbs that are raising in the active, and sentential objects for verbs that are raising in the passive.

The possible existence of an extraposed sentential subject of the raising verb, such as :

(58) *It seems that John has left.*

(59) *It turned out that John was gone.*

is accounted for by having the corresponding S-tree as part of the tree family of the raising verb. The S trees for raising verbs will have a slot for a subject, namely *it*, or a *that*-clause (depending on whether the raising verb yields obligatory extraposition or not), whereas the S tree for the auxiliary will be reduced to the auxiliary verb itself²⁸:



Some raising verbs also occur in other non-raising contexts. For example *stop*, *begin* and *happen* take NP subjects and NP complements and, less naturally, sentential subjects:

²⁶The same representation was originally discussed by Kroch and Joshi (1985).

²⁷As noted by J. Bresnan 1982, this needs not always be the case:

John tends to lie.

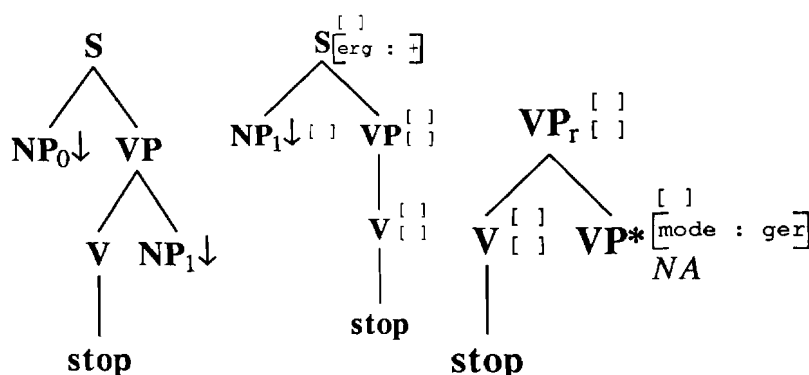
John tends that Mary lies to him.

²⁸For sentences such as 'John seems happy' see the subsection on small clauses in the future work section.

An accident happened.
The car stopped.

If we want to say that the raising and non-raising contexts are related and that the verb has roughly the same semantics in both cases, the raising tree has to be part of a larger tree family.²⁹

- (60) (non-raising) *John stopped the car.*
- (61) *The car stopped.*
- (62) ? *Killing the dolphins must stop.*
- (63) (raising) *There stopped being troubles around here.*
- (64) *Close tabs stopped being kept on Mary.*
- (65) (non-raising) *An accident happened.*
- (66) ?? *Killing seals used to happen quite often around here.*
- (67) (raising) *There happened to be troubles around here.*
- (68) *Close tabs happened to be kept on Mary.*



Few raising verbs will select only VP-trees (for their raising use); the only examples we found were *tend* and *end up*.

5.1 Raising passives

- (69) *Peter was believed to be fat (by John)*
- (70) *Peter was considered to be a fool (by Mary).*

Passive raising verbs such as *consider* or *believe* have been widely discussed in the literature, either as raising verbs or as part of a wider Exceptional Case Marking class of verbs (Chomsky 1981). Bresnan (1982) treats them as verbs that take NP VP complements, making the distinction with non-raising verbs such as *force* or *persuade* at the f-structure level. Her analysis is mainly on the basis of right-node raising and heavy NP-shift:

- (71) * *Mary believes, but Catherine doesn't believe, Peter to be fat.*
- (72) *I will consider to be fools in the weeks ahead those who drop this course.*

We think that the syntactic constraints on such phenomena are not altogether clear, and cannot compensate for the major drawbacks of a two-complements analysis. Such a representation cannot account for the fact that a regular alternation between such complements and true sentential complements occurs without

²⁹For a discussion of this issue, see Perlmutter 1970

a change of meaning:

- (73) *John believes Bill to be a fool.*
- (74) *John believes that Bill is a fool.*

- (75) *John expects Bill to come tomorrow.*
- (76) *John expects that Bill will come tomorrow.*

A two-complements analysis also fails to provide any prediction for similar raising verbs that only take a sentential complement :

- (77) *Everybody says that Bob eats ice cream with a fork*
- (78) * *Everybody says Bob to eat ice cream with a fork*³⁰
- (79) *Bob is said to eat ice cream with a fork (by everybody)*

Another property is also difficult to handle with a two-complements analysis. Omission of the infinitival seems never to be possible, as predicted by an S-complement analysis:

- (80) * *John believes Bill.* (without complete change of meaning)
- (81) * *John expects Bill.*

It is also consistent with an S-complement analysis of such verbs that their passive can only subcategorize for an S-subject. If an NP subject occurs in the passive, the sentence has to be a raising construction.

Finally, at least in a formalism like TAG, which relies on only one level of syntactic representation, it is difficult to distinguish raising verbs from verbs like *persuade* or *force*. For example, the two complements analysis does not account for the acceptability of *there* or frozen subjects in idiom chunks with raising verbs but not with equi-verbs:

- (82) *There was expected to be a flower vase on each table.*
- (83) * *There was forced to be a flower vase on each table.*

- (84) *Close tabs were believed to be kept on her.*
- (85) * *Close tabs were asked to be kept on her.*

In order to show the difference between these verbs and equi-verbs such as *force* or *ask*, we consider them to take a sentential complement in active sentences. We consider the passive of these verbs, on the other hand, to be raising verbs just like *seem* or *happen*. We thus only have one tree-structure that accounts for all raising phenomena in our grammar.

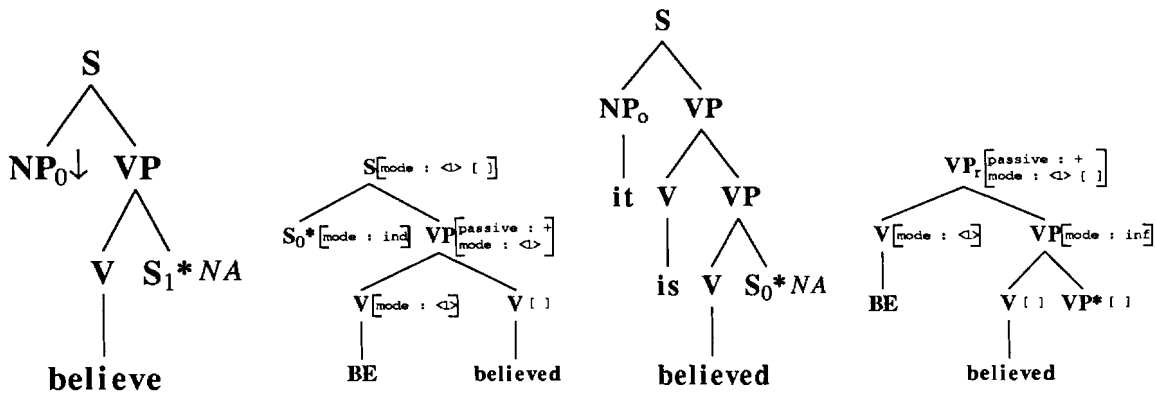
Since these verbs have sentential complements in the active, they usually have sentential subjects in the passive (extraposed or not), and the relationship between 'It seems that Mary left' and 'Mary seems to have left' is the same as that between 'It is considered that Bob is a fool' and 'Bob is considered to be a fool':

- (86) *That Mary was gone was widely believed.*
- (87) *It is (widely) believed that Mary is a fool.*
- (88) *Mary is believed to be a fool.*

- (89) *It was proved that Mary was a sinner.*

³⁰ *Everybody says for Bob to eat ice cream with a fork* would have a different meaning.

(90) *Mary was proved to be a sinner.*



The difference in our grammar between verbs such as *say* or *think*, and *believe* or *consider*, which all have raising passives, will be the $\langle mode \rangle$ and $\langle comp \rangle$ features of the S-complement they select in the active – respectively that-clauses and to-infinitives. The passive of all these verbs, on the other hand, always selects for a to-infinitive or a gerund, and this will be stated directly at the level of the VP auxiliary tree.

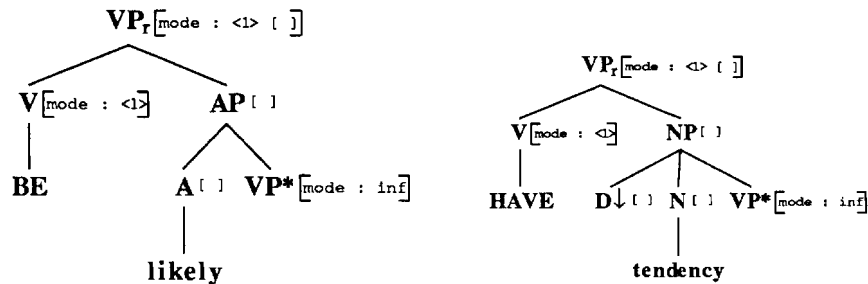
5.2 Raising Adjectives and Nouns

Raising adjectives in copular constructions have long been noticed. (See (91).) But there are also raising nouns contained in light verb constructions like (92):

(91) *It is likely to rain tomorrow.*

(92) *It has a tendency to rain a lot in this country.*

We represent both cases with extended VP-trees:



The motivation for such expanded trees, with multicomponent anchors, is given in the section on light verb constructions.

It is worthy of note that such nouns and adjectives lose their raising properties when they are expanded to NP (or AP) only and not in light-verb constructions:

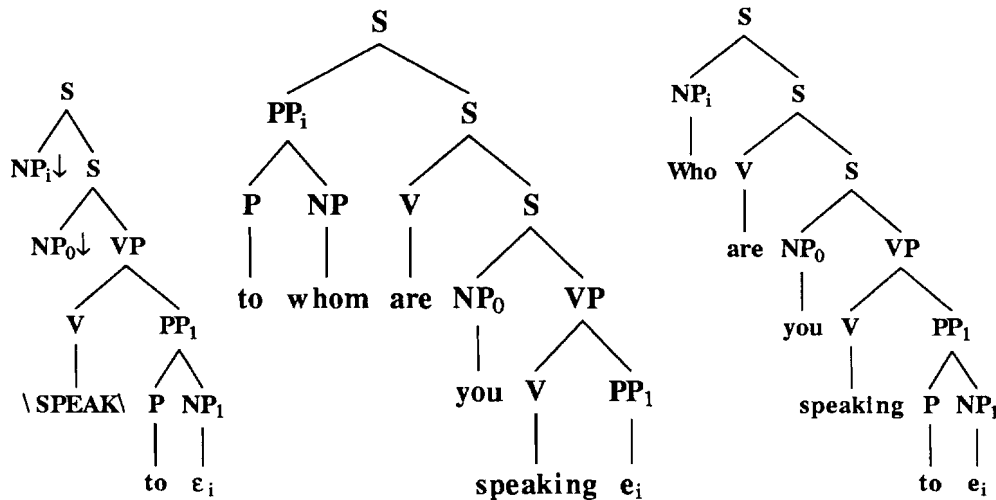
- (93) * *The clouds make it likely to rain.*
- (94) * *It's tendency to rain.*

As has been suggested by Kroch and Joshi 1985, the corresponding elementary NP or AP trees would not be well-formed and would not yield the right result (with the subject and the 's under the same node).

6 Prepositional Complements

In the English LTAG, verbs that take prepositional complements are defined by tree families containing expanded trees. Specifically, these verbs each select a tree family whose trees have a PP complement node expanded to specify an NP or an S argument. By using these expanded trees, we are taking advantage of TAG's extended domain of locality to achieve certain syntactic and semantic effects.

The basic syntactic benefit of the expanded tree structures is the ability to allow for both PP and NP extraction without resorting to preposition incorporation or re-analysis. Below are some examples of the trees available for the verb *speak* under this approach.



The grammar will thus consider both PP_1 and NP_1 to be eligible syntactic arguments, and will generate the following sentences:

- (95) *To whom are you speaking?*
- (96) *Whom are you speaking to ?*
- (97) *John was spoken to (by Mary)*

Preposition incorporation, originally proposed in Bresnan (1982) to account for prepositional passives, faces the following problems. First, in order to account for sentences that allow both PP and NP movement, a double analysis has to be postulated – a (VP) (NP) structure for preposition stranding, and a (V) (P NP) structure for pied-piping:

- (98) *They look at John with pride.*
- (99) *It is at John that they look with pride.*
- (100) *It is John that they look at with pride.*

However, as in the example above, the original sentence is not ambiguous and the double analysis misses a reasonable match between the syntactic and the semantic component of the grammar. In the TAG analysis, pied-piping is simply defined as extraction at the PP_1 node, and preposition stranding, as well as the pseudo-passive if needed, are defined as extraction at the NP_1 node.

A second problem with the incorporation analysis comes from the fact that verbs and prepositional arguments are not always contiguous. This is straightforward in the case of verbs with more than one complement:

- (101) *What did you speak with Mary about ?*
 (102) *? About what did you speak with Mary ?*

As Bresnan (1982) notes herself, insertion of an adverbial is also usually allowed between the verb and preposition in preposition stranding cases, and it is also acceptable in some prepositional passive constructions:

- (103) *That's something that I would pay ((easily) twice) for.*
 (104) *These are the books that we have gone ((most) thoroughly) over.*
 (105) *John is spoken ((very) highly) of.* (from Bresnan 1982)

Incorporation seems difficult to define over these unbounded strings. In a TAG, on the other hand, the insertion of an adverbial does not cause a problem³¹. The TAG analysis also naturally rules out sentences like the one below, where an adverbial occurs in between the preposition and its argument:

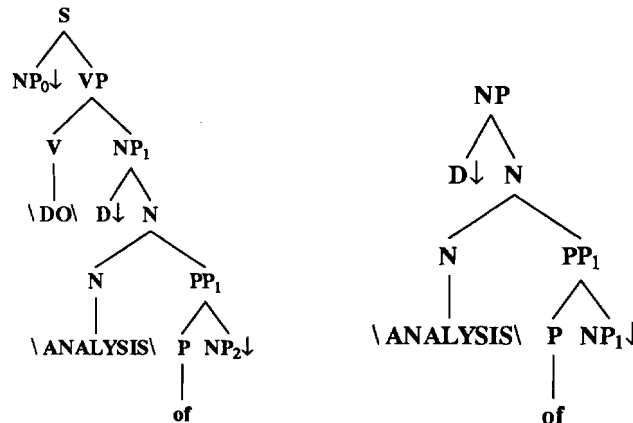
- (106) **John went over most thoroughly the books.*

The extended TAG structure also provides a direct expression of dependencies between a subject and an NP argument inside PP, such as those that hold in the case of 'symmetric' predicates:

- (107) *John alternates with Mary/ * the rain.*
 (108) *The sunshine alternates with * Mary/ the rain.*

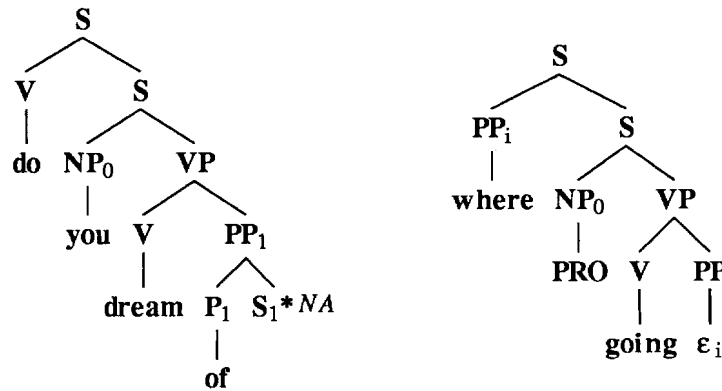
Our use of local constraints contrasts with the need to pass constraints in context-free based grammars like GPSG and HPSG.

Nouns and adjectives that take prepositional complements are also represented with trees extended to P and NP (or S). Most of these predicate terms also enter light-verb constructions as described in the next section, and their arguments in these cases must also be available for extraction. The two figures below are light-verb and non-light-verb trees for *analysis*:



Verbs taking sentential prepositional complements also make use of an extended domain of locality. But they select auxiliary S-trees just as other verbs taking sentential complements do (when allowing for extraction).

³¹This is perhaps an overstatement. If the scope of the adverbial is V then there is certainly no problem. If, however, the scope is VP, a normal adjunction structure will not work. We need to use a more sophisticated version of TAGs, which factors out dominance and precedence relations. Such a system has been described in Joshi (1987).



One interesting observation for prepositional S complements is that, in their tree families, we could also add a tree for ‘direct’ sentential complements, when there are good reasons to consider this form to be an ‘underlying’ prepositional complement. The examples below suggest that extraction phenomena may provide one reason:

- (109) *You agreed/insisted on going to the beach.*
- (110) *He agreed/insisted that we go to the beach.*
- (111) *Going to the beach was agreed/insisted on (by you).*
- (112) *That we go to the beach was agreed/insisted on (by you)*

The phenomena can be found in the complements of adjectives and nouns in copular constructions. This predicts that in these cases, wh-questions and clefts are possible with the preposition:

- (113) *John is happy to go.*
- (114) *John is happy about going/this job.*
- (115) *What is John happy about ?*
- (116) * *What is John happy ?*
- (117) ? *It is that he is be able to graduate that John is happy about.*
- (118) * *It is that he is able to graduate that John is happy.*
- (119) *What he is happy about is that now he can graduate.*
- (120) *John had the/an idea of going abroad.*
- (121) *John had the idea that Mary is lying to him.*
- (122) *What does John have the idea of ?*
- (123) * *What does John have the idea ?*

Clearly, the TAG analysis gives a more elegant treatment of the above syntactic facts than was previously available. The semantic benefits of this analysis, on the other hand, may not be as straightforward. It may be useful to consider the meaning of verb + Prep to be non-compositional, like the analysis of verb-particle constructions described in the previous section, but with different structures. Syntactically, the preposition would be the head of PP, but it would also form a single multi-component anchor with the verb.

Examples are constructions using prepositions like *to* and *of*, which have been argued to lack a precise and unified meaning when they are used with verb complements. As noted by Pollard and Sag (1987) (among others), the same semantic relation can be expressed by different prepositions depending on which verb is involved :

- (124) *John was charged with the murder.*

- (125) *John was accused of the murder.*
(126) *John was blamed for the murder.*

Using the multi-component anchor analysis prevents us from attaching any independent semantics to the preposition. More meaningful prepositions, particularly the locatives, will always be treated as independent anchors, and will have to be substituted into the S-tree.

This does not prevent the preposition from having several values in some cases as long as the resulting predicates are synonymous as for : *rely on/upon* . The preposition can also be optional in such cases as :

John entered the room = John entered into the room
John protested the order = John protested against the order.

As a final comment on prepositional complements, we should note that we will probably need to require that for some verbs taking locative complements their tree families contain an additional tree where PP is either unexpanded or expanded to a P alone. This seems to be necessary to account for certain locative adverbials that can fill the PP position, as shown below:

- (127) *John put the book somewhere/elsewhere/far away*

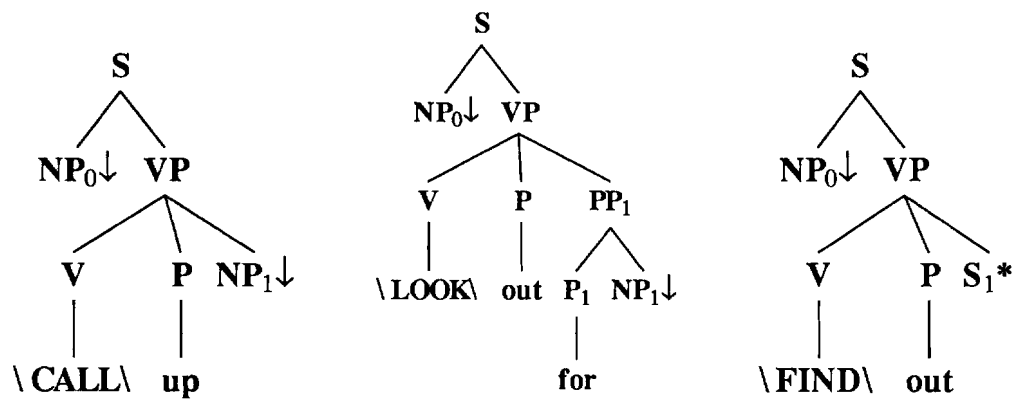
7 Verb-Particle Combinations

Argument structures containing verb-particle combinations are separated from those structures with ordinary prepositions. The reason for having separate tree-families for verb-particle combinations is that they do not undergo the syntactic transformations that similar constructions with ordinary prepositions do.

A preposition is considered to be a particle if pied-piping of it is ruled out, as well as coordination with another PP:

- (128) *Who are you calling up?*
 (129) * *Up who(m) are you calling ?*
 (130) *Mary, who I was calling up, was out.*
 (131) * *Mary, up whom I was calling, was out.*
 (132) *I called up Mary and John.*
 (133) * *I called up Mary and up John.*

Verb-particle combinations allow for basically the same variety of structures as simple verbs³². In other words, there will be close to the same number of tree-families for verb-particle combinations as for simple verb constructions. For example, as shown below, verb-particle constructions can include NP, PP, or S complements.



The particle is thus considered to be a preposition but it is treated as a sister to the verb. In contrast with prepositional complements, there is no PP node and the particle cannot move together with the NP object. The structures of the elementary trees themselves thereby exclude pied piping and coordination with a PP. Since the particle and the NP object are two separate sisters of V, 'pied-piping' would amount to double extraction, which the grammar rules out, as mentioned in the extraction subsection above.

The NP (or S) arguments in a verb-particle combination are expected to have the same properties as the NP (or S) argument of simple verbs. Extraction is allowed out of S complements, and the NP argument can passivize.

- (134) *John found out that Mary went to New York.*
 (135) *Where did John find out that Mary went to ?*
 (136) *A neighbour called up the authorities.*
 (137) *The authorities were called up by a neighbour.*

³²Appendix 3 gives a listing of the tree-families for verb-particle combinations.

Usually, verb particle combinations taking NP-arguments can undergo the following movement pattern:

- (138) *John called up Mary.*
 (139) *John called Mary up.*

A parallel might be drawn between particle hopping and the more general permutation of sisters of verbs, as shown below:

- (140) *John spoke about the party with Mary.*
 (141) *John spoke with Mary about the party.*

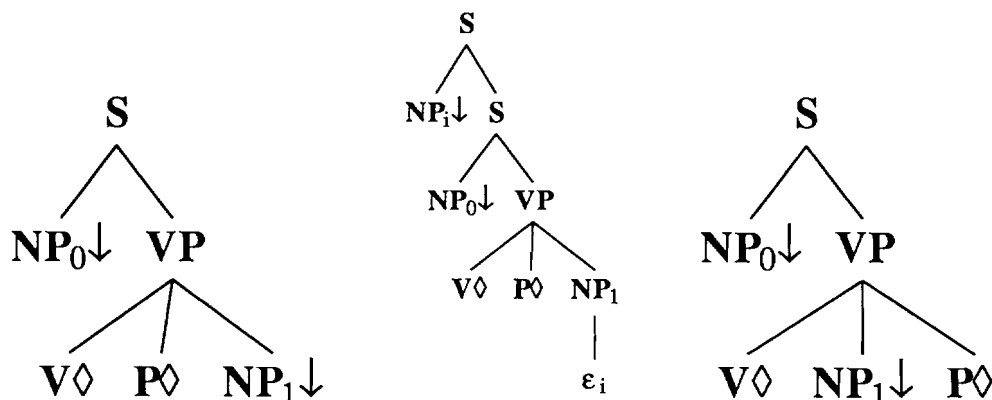
As is the case with argument permutation, the length (or the syntactic 'heaviness') constrains such movement:

- (142) * *I called the girl I like the most up.*
 (143) ? * *John spoke about time that he visited his friend George in the cabin in the mountains with Mary.* (meaning = spoke with Mary about ...)

But particle hopping is more constrained since it is blocked in the case of prepositional complements, and is forced to occur in the case of pronominal complements:

- (144) * *I called up her.*
 (145) *I called her up.*
 (146) * *John looked for the police out.*
 (147) *John looked out for the police.*

In the current state of the grammar, particle movement is represented as another structure within a verb-particle combination family (except when the NP node is empty) as the structures below illustrate:



The movement of the particle can be affected by structural transformations, e.g. dative alternation.

- (148) *John answered back to Mary.*
 (149) *John answered Mary back.*
 (150) * *John answered back Mary.*
 (151) * *John answered to Mary back.*

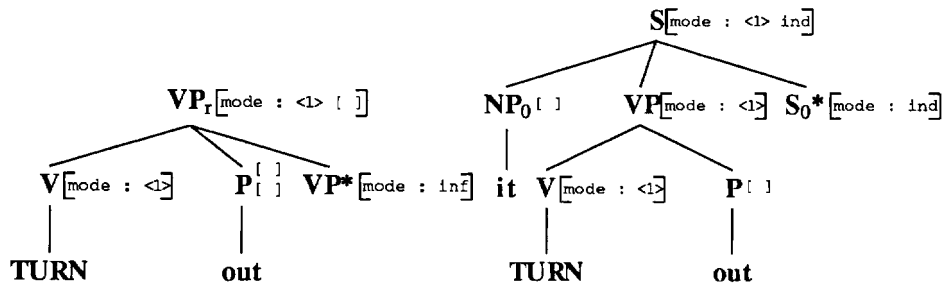
Because they are part of a multi-component anchor, the particles are not expected to contribute any independent semantics to the sentence. The semantically relevant unit is the verb-particle combination.

Having a multicomponent anchor thus avoids the need for reanalysis (or particle incorporation), which may get into trouble in cases of adverb insertion:

(152) *The clock fell apart.*

(153) *The clock fell completely apart.*

Verb-particle combinations which are used as raising verbs are represented along the same lines:



8 Light Verb Constructions

We follow the representation defined, both for English and French, by Abeillé (1988a). Light Verb constructions are sentences with the following property: they contain a non-verbal item that together with the verb control both the structure of the sentence and the selectional restrictions of the arguments of the sentence, ie:

- (154) *The man made a dash across the road.*
 (155) * *The stone made a dash across the road.*
 (156) *The stone/man made a splash in the pond.* (Cattell 1984)

The noun *dash*, when used in a light-verb construction with the verb *to make* only subcategorizes for an animate subject. The noun *splash*, however, in conjunction with the verb, *to make*, subcategorizes for either an animate or an inanimate subject.

In the sentences below, as mentioned by Cattell (1984), it is 'offer' that allows for the dative alternation that 'make' does not allow:

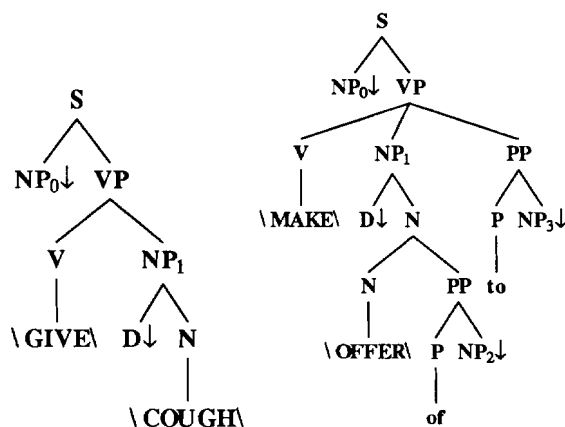
- (157) *They made an offer of money to the police.*
 (158) *They made the police an offer of money.*

The verb seems to provide only person, tense and aspect marking to the sentence. The predicative nominal in the complement position exhibits the same subcategorization frame as when it heads an NP-structure:

- (159) *The man's dash accross the road.*
 (160) *Their offer of money to the police.*

This is why such predicative nouns in complement position are considered to constrain subcategorization and selectional restrictions in light verb constructions. (following M. Gross 1981).

Light verb constructions can be represented naturally in a TAG with only one basic structure. The light verb and the predicative nominal (plus required prepositions) are considered to be a multicomponent anchor of the corresponding sentential elementary tree. For predicative nouns taking complements, such anchors select an expanded elementary tree with a slot for the prepositional complement of the predicative noun.

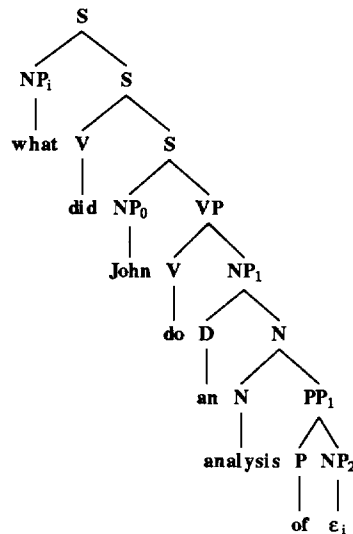
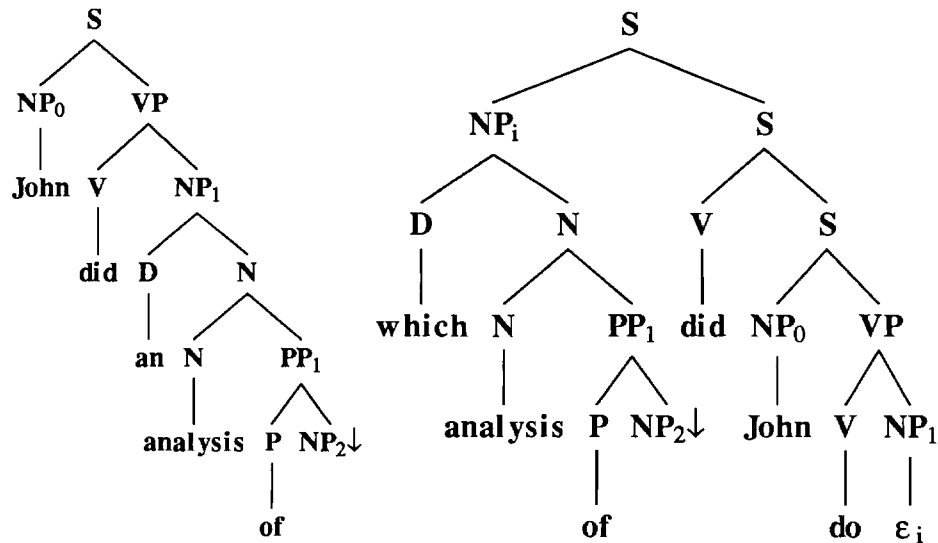


This expansion accounts for the prepositional complement being a complement of a noun but also having all the properties of a complement of a verb. The PP-node below corresponding to *of anaphora* (for example, in 'John did an outdated analysis of anaphora.') will belong to the initial tree, and thus is an argument of the sentence as any verbal complement normally is. At the same time, though, it is dominated by the noun,

analysis, and the properties it exhibits as a nominal complement result. As shown by Abeillé (1988a), the grammar is capable of making syntactic distinctions between strings that are structurally isomorphic, ie [NP VP [NP [PP]]]: the resulting trees are the same, but one is an initial tree, while the other is derived.

One must first bear in mind the grammaticality distinction between a light verb construction and a non-light-verb construction. Let us start with examples of light-verb constructions.

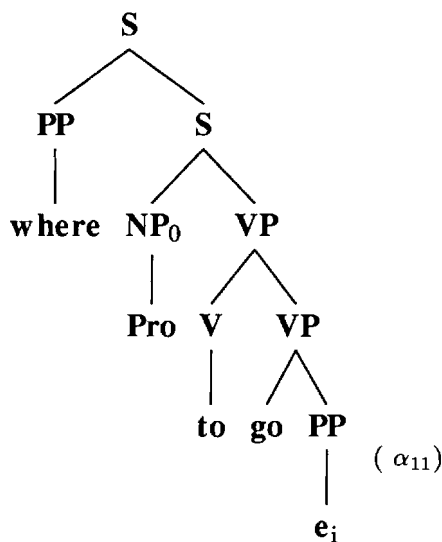
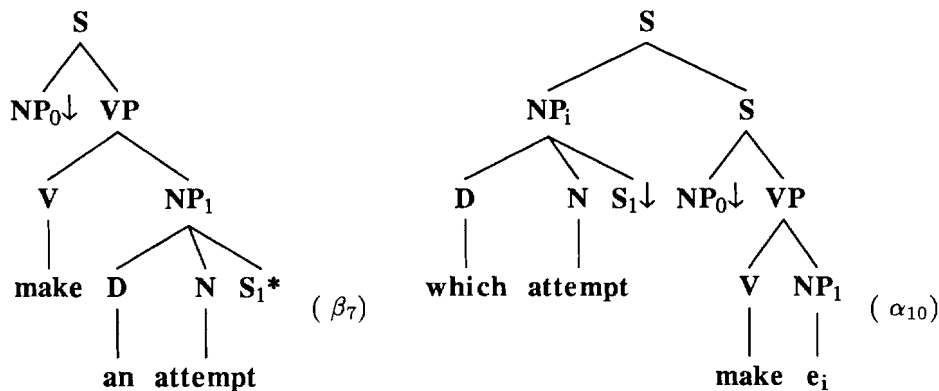
- (161) *John did an outdated analysis of anaphora.*
- (162) *Which analysis of anaphora did John do?*
- (163) *What did John do an analysis of?*
- (164) *This is the phenomenon that John did an analysis of.*



In our grammar, syntactic rules are defined on the arguments of sentential elementary trees. Since *anaphora* in this case is an NP-argument of a sentential tree, wh-movement and topicalization are predicted to apply to it.

In light-verb constructions containing a sentential complement, both wh-movement of the NP containing the predicative noun and unbounded extraction out of the sentential complement are allowed, since both are arguments of the elementary sentential tree.³³

- (165) *John made an attempt to go to Paris.*
 (166) *Which attempt to go to Paris did John make ?*
 (167) *Where did John make an attempt to go to ?*



However, such properties will not always hold for complements of nouns in non-light-verb contexts³⁴ :

³³The example tree structures below need additional adjunctions and substitutions to derive the corresponding sentences. The structure, β_7 , needs to adjoin to a structure representing, "PRO to go to Paris" at the root node, to derive (165). An adjunction of a structure representing the sentence, "PRO to go to Paris" is needed at the interior S node in tree α_{10} , as well as an adjunction of the auxiliary 'do.' to derive (166). After the adjunction of the auxiliary 'do' onto tree β_7 at the root node, that tree structure can be adjoined into α_{11} at the interior S node to derive sentence (167).

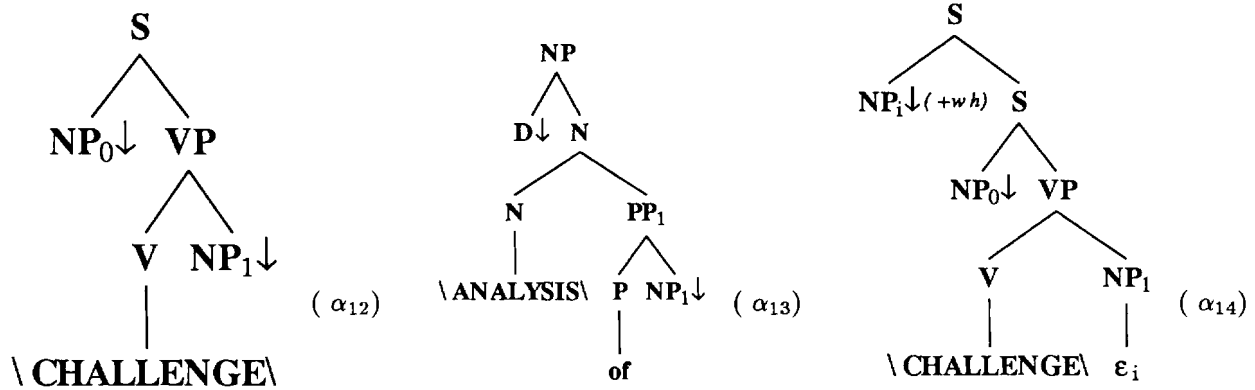
³⁴They would not hold either for modifiers of NP, since modifiers are auxiliary trees that adjoin to NP :

* *What do you see girls without?*

* *These are the kinds of hats you often see girls without*

- (168) *John challenged an analysis of anaphora.*
 (169) *Which analysis of anaphora did John challenge?*
 (170) * *What did John challenge an analysis of?*
 (171) * *This is the phenomenon that John challenged an analysis of.*

Sentence (168) above can be derived by substituting α_{13} into α_{12} at node NP_1 , but, as the reader can verify, there is no derivation using the above tree structures, for the ungrammatical sentences (170)-(171).



As shown by Abeillé (1988c), this representation avoids the need for dual analyses or noun incorporation presented in previous literature. A dual analysis for light verb constructions that allows two different structures for wh-questions (M. Gross 1976, Bach and Horn 1976), or for passives (Bresnan 1983) is unprincipled because the sentences do not exhibit any semantic ambiguity.

Another property of light verb constructions is the constraint that they exhibit on the determiner of the sentence as shown by the contrast below:

- (172) *John took a/this/*his/*my/?*Mary's trip to Spain.* (light-verb)
 (173) *John booked a/this/his/my/Mary's trip to Spain.* (non-light-verb)

Differences in semantic interpretation for the two structures are also possible with this analysis. Since the light-verb predicate-nominal combination defines one elementary tree, it is expected to behave as a semantically composite predicate, although syntactically, the predicate nominal is a direct object of the noun, subject to relativization and wh-question:

- (174) *The trip that he took was wonderful.*
 (175) *He took last year a trip that he will never forget.*
 (176) *Which trip did he take last year?*

It is worthy of note that, in light-verb constructions, standard wh-movement cannot apply to the predicative noun:

- (177) *What did he take?*
 (178) * *A trip to Spain.*

As is to be expected from a component of the anchor, the item corresponding to *trip* must always be lexically present.

The semantic compound is obtained by having the light verb and the predicate nominal be parts of a multicomponent anchor. Their lexical value is co-dependent:

- (179) *Tom gave/*took a sneeze.*
 (180) *Tom *gave/took a snooze.*

For example, the noun, *trip*, can co-occur with the verb *take*, but not with *have* or *give*:

- (181) *John took a trip to Spain.*
 (182) * *John has a trip to Spain.*
 (183) * *John gave Mary a trip to Spain.* (as a light-verb construction)

Having such a semantic compound also predicts the adjective/adverb alternation as shown below:

- (184) *John took a quick trip to Spain = John quickly took a trip to Spain*
 (185) *John made a desperate attempt at leaving = John desperately made an attempt at leaving*

Since the predicate of the light-verb construction is comprised both of the light-verb and the predicate nominal, modifying one or the other should not make any semantic difference. This will not hold of course for non-light-verb constructions. We can thus account for semantic differences between light verb and non light-verb constructions, provided our semantic analysis is based on the derivation trees (by which they differ), not the derived trees.

We are currently exploring the treatment of copular constructions as light verb sentences with the adjective and the copula being the multicomponent anchor of the sentence.

Remaining questions are as follows. We have successfully incorporated into the grammar both the Complex NP-island constraints and the CNPC violations that the light verb constructions would normally allow. We are left with other cases of CNPC violations that cannot be analysed as light verb constructions, such as:

- (186) *Who did you see a picture of ?*
 (187) *What did you write a book about ?*

Another question is how to incorporate verbs that have all the properties of light verbs but seem to have some semantic autonomy :

- (188) *John got the guts to tell Mary the truth.*
 (189) *John sticks to the habit of jogging every morning.*

“Get” and “stick” are usually variants of ‘true’ light verbs, with similar syntactic properties but more semantic contribution :

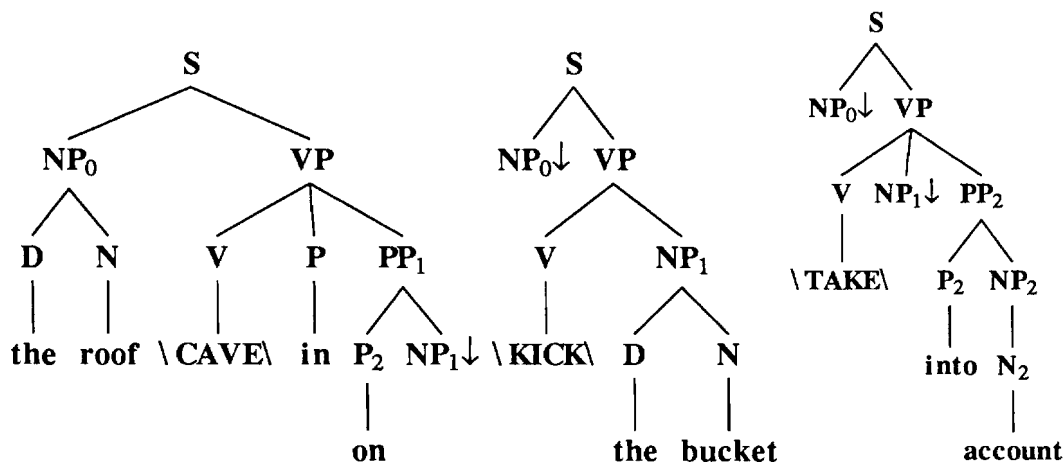
- (190) *John had the guts to tell Mary the truth.*
 (191) *John is into the habit of jogging every morning.*

A solution could be to consider them as substituted, instead of being part of the anchor, into the tree structure of the predicate nominal.

9 Idioms

Following Abeillé and Schabes (1989), an idiom is entered into the lexicon as one unit, rather than being built out of the component words by some rule of idiomatic use. This accounts for the distinguishing characteristic of idioms which is their semantic non-compositionality³⁵ The meaning of *pull one's leg*, for example, cannot be derived from that of *pull* and that of *leg*. It behaves semantically as one predicate, and the whole VP *pull one's leg* selects the subject of the sentence and all its possible modifiers. Semantically it would not make sense to have an idiom's parts listed in the lexicon as regular categories and to have special rules to limit their distribution to this unique context. If the parts of an idiom are already listed in the lexicon, these existing entries are considered only as homonyms. This accounts for the usual ambiguity between literal and idiomatic readings of many idioms.

Such representations are also consistent with the large structural variety that idioms exhibit:



The lexical entries and the tree-families for idioms are like those for regular sentences. Usually, idioms are composed of the same lexical material as ‘free’ sentences. This is captured by having the idiom’s syntactic entry headed by the same morphological entities (existing in the morphological lexicon) as other syntactic entries. The tree structures associated with them are sometimes more expanded than those in regular tree-families, in order to specify a determiner, or any category, that in other cases would be freer.

If one takes the example of ‘NP0 kick the bucket’, it has the following entry in the lexicon, which corresponds to a set of trees more expanded than those for non-idiomatic transitive constructions.

\backslash KICK \backslash ,	V	:	Tnx0Vnx1 (transitive verb)	(a)
\backslash KICK \backslash ,	V; the, D ₁ ; bucket, N ₁	:	Tnx0VDN1 (idiom)	(b)

The three pieces *kick*, *bucket* and *the* are considered to be the multicomponent anchor of the idiom.

Some idioms allow some lexical variation, usually between a more familiar and a regular use of the same idiom, for example in English, *NP₀ have a cow*, *NP₀ have a bird*, *NP₀ have a fit*. (to get very upset). This is represented by allowing disjunction on the string that belongs to the multicomponent anchor of the idiomatic tree. *NP₀ have a cow/bird/fit* will thus be one entry in the lexicon, and we do not have to specify that *cow*, *bird*, and *fit* are synonymous (and restrict this synonymy to hold only for this context).

Some idioms select elementary tree structures that are identical to those for ‘free’ sentences. For example, the idiom *NP₀ sees red* has the same structure as any verb taking an adjectival complement (ex: *NP₀ is*

³⁵For a discussion of the relationship between syntactic rules applying to some idioms and the compositionality of the same idioms (originally proposed by Wasow, Sag and Nunberg 1982), see Abeillé and Schabes 1990, and Abeillé 1990.

AX_1), except that *red* is part of the multicomponent anchor of the idiom, whereas a ‘free’ AX is inserted in AX_1 in the case of ‘is’.

Discontinuities that may arise in idioms are accounted for straightforwardly (Abeillé, Schabes 1990). They may be part of the very definition of the idiom (in the case of holistic predicates such as : to take NP into account) or arise from the regular adjunction of modifiers (as in : The roof **will soon** cave in on John). They are directly encoded in the topology of the elementary tree for idioms as they are in any elementary tree. It is only at the level of the derivation tree that all idiomatic parts will appear as lexically one (contiguous) unit.

Since idioms select regular syntactic structures, i.e., elementary trees that are already in the grammar or that could have been derived from structures already existing in the grammar, they are predicted to be subject to the same syntactic constraints as non-idiomatic sentences. Lexical and syntactic rules are also liable to apply, although they might exhibit more lexical idiosyncrasies (Abeillé 1990 forthcoming).

An advantage of TAG’s is that they allow the definition of multicomponent anchors for idiomatic structures without requiring them to be contiguous in the input string. The formalism also allows the direct access of different elements of the compound without flattening the structure. For example, as opposed to CFGs, direct dependencies can be expressed between arguments that are at different levels of depth in the tree without having to pass features across local domains.

Taking more advantage of Tag’s extended domain of locality for semantic analysis is currently being investigated. Research is being done along the lines of “Synchronous Tags” as proposed by Shieber and Schabes 1990, which define pairs of trees corresponding syntactic structures on one side and trees corresponding to semantic representation (eg logical form) on the other side. This allows to keep semantic analysis strictly monotonic, in the sense that any combination of syntactic structures amounts to combining semantic ones as well at the same time, while allowing for semantic non compositionality of certain natural language phenomena such as idiomatic structures.

10 Ongoing Work

10.1 PRO, Case, and Anaphora

Although PRO has not yet been added to our LTAG, a simple extension has been defined. Because PRO is not an overtly present lexical item, it can not be added to the lexicon in the same way as other words. One possible but inelegant solution would be to make a duplicate of all of the relevant S-trees, with 'PRO' subjects and the appropriate features directly inserted. In other words, there would be a different representation for every possible occurrence of PRO in English. In addition to the fact that this would ignore the relation of case to subject type, it is an approach that would probably have unfortunate effects when we begin to build a compositional semantics for the grammar.

An alternative solution would be to treat PRO as a special type of lexical item which defines a simple elementary NP tree available for insertion into subject position. This tree will be available when evaluating any sentence but does not necessarily have to be used in order for the evaluation to be successful. The PRO tree will, however, be substituted into subject position when no other lexical item can fill that position (and the features on the NP node are compatible with the features of PRO.) If neither a 'normal' lexical item nor PRO can fill the subject position, we will say that the sentence is incomplete and, hence, not acceptable as output of the grammar.

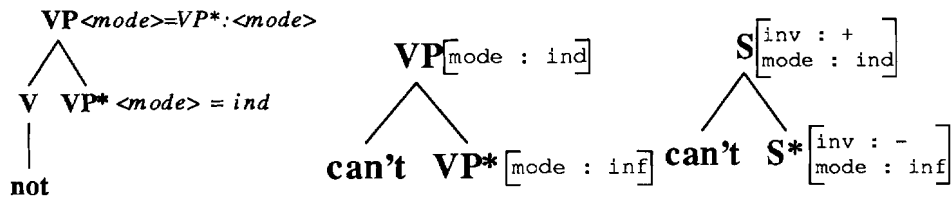
The ⟨CASE⟩ feature, the details of which are currently being worked out, will play a crucial role in preventing nominative subjects in tenseless sentences. Another feature, such as ⟨LEXSUBJ⟩, might be used to prevent PRO from occurring after ECM verbs like *believe*. *believe* would be specified as ⟨LEXSUBJ⟩ = + while verbs like *expect* would be unspecified for the feature ⟨LEXSUBJ⟩ and thus able to act as both ECM verbs and 'Control' verbs. Alternatively, we could use the ⟨CASE⟩ feature to handle the ECM/Control verb distinction as well if we allow a special feature value, say NONE, which has the property of only unifying with itself (i.e., it will not unify with an unspecified value.) PRO would be marked ⟨CASE⟩ = NONE and pure ECM verbs would constrain their S-complements to be ⟨CASE⟩ = ACC. Neither of the above alternatives has been worked out completely but they each seem to have advantages. In general, however, they achieve the same results.

Bishop (1988) describes a possible feature system for dealing with PRO control both in sentential complements and in relative clauses in TAGs. Full advantage is taken of lexicalization because constraints on PRO control are sensitive to the lexical value of the main clause anchor. TAG's extended domain of locality is also needed in this system to enable reference to a predicate and all of its arguments at one level. Bishop (1988) relies on an association of thematic roles to the arguments but, since thematic roles have not yet been implemented in the grammar, this approach has not been put to a practical test. An analysis of arbitrary 'PRO' has not been developed.

Work is also being done currently on binding anaphors and restricting co-reference on pronouns.

10.2 Negation

Some LTAG structures for simple negation, such as the one below, have recently been developed by Raffaella Zanuttini:



This will allow for sentences such as:

- (192) *I could not have been not informed of your coming.*
- (193) *He might have forgotten not not to do it.*

And will rule out sentences like:

- (194) * *John not loves Mary.*

Negative contractions have been incorporated into the auxiliary verb structures.

Polarity features and constituent negation have received some consideration but need to be worked out.

11 Future Topics

11.1 Copula Constructions and Small Clauses

The most recent proposal for copulas is to treat their sentences as raising constructions. The *be* auxiliary would thus be adjoined to a ‘small clause’ headed by either a noun, an adjective or a preposition. In order to be able to always use the same *be* structure, the interior node it adjoins to should always have the same category, for example PredP. If we redefined all the VP’s in our grammar (which are always interior nodes since we do not allow for VP complements) as PredP we would then have a larger generalization about what sentences are. Perhaps it is even more natural to have this PredP for sentences anchored not only by verbs but also by prepositions, or nouns (as in the verb-particle and light verb cases).

Trees for raising verbs and auxiliaries would then also have to be re-rooted in PredP instead of VP. The above-mentioned trees for negation, and the trees for adverbials would undergo the same renaming.

Additional constraints on non-verbal PredPs would then of course have to be added to prevent, for example, the insertion of modals and *do* into adjectival PredPs unmarked for the *mode* feature. Further constraints would also have to be set for distinguishing verbs that take any kind of sentential complement from verbs that will take only the S-complements that are not small clauses. The same distinction would have to be made between raising verbs that adjoin into any PredP, and those that adjoin only into non-small clauses. In general, it is not yet clear that this is the right approach to handling the small clause phenomena.

11.2 Adverb Movement

Limited treatment has been given to adverbs at this point. The most basic cases are simply handled by an adjunction to VP, AP, PP or S of an adverbial auxiliary tree. Lexical restrictions can be used to constrain

whether left or right-adjoining trees (or both) are available for a particular adverb.

It is worth mentioning here that, although adverb movement has not yet been given serious treatment in our grammar, it appears that our analysis of prepositional phrases will have some predictive effect on this. For PP adverbials adjoined to verb phrases, preposition stranding is ruled out in English, as shown in the contrast below:

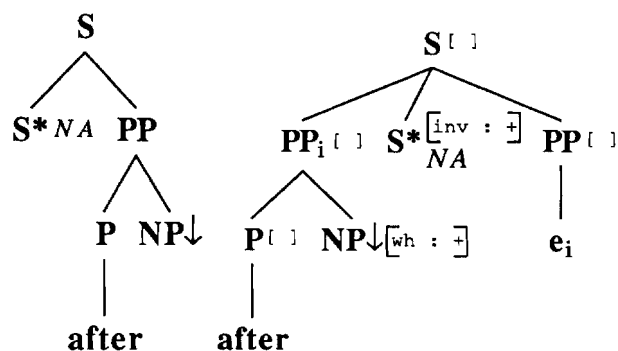
- (195) *Bill slept without a pillow.*
 (196) ?? *What did Bill sleep without?*

LTAG can prevent this construction naturally since the adverbial adjunction tree can not be rooted in both VP and S.

However, because of expanded PP nodes, PP adverbials adjoined to S can still get preposition stranding as in:

- (197) *Which park did Glen and Rhonda have their picnic in?*

The adverbial scope distinctions for PP that we are making here are correct, and it seems that there are some unacceptable extractions out of S modifiers as well. These exceptions have to be specified by lexical restrictions on the types of adjunct structures available.



If we exclude the auxiliary tree with preposition stranding from the lexical entry for *after*, allowing only trees similar to those shown above, then we rule out:

- (198) ?? *Which meal did John sleep after?*

but allow:

- (199) *After which meal did you sleep?*

as well as the echo-question:

- (200) *You slept after which meal?*

As mentioned before, the discussion of adverbials is quite informal at this point; classification of adverbial types, except for very simple cases, has yet to be made.

Wh-movement of adverbials is an important area for future study. One possibility is that the treatment of these constructions will depend upon the analysis of optional arguments.

11.3 Optional Arguments

Where an argument to a verb is optional (under some definition of optionality that we have yet to choose), the lexical entry will have parentheses around that argument within the tree-family name. This will in turn associate two sets of trees with that entry: one set with the argument filled, and one set with the argument not present. In this fashion we group together structures that have the same meaning, but differ only on the presence or absence of one argument.

This implementation requires a careful study of the lexicon to determine the appropriate semantic basis for these groupings.

But there is also a question that remains to be answered. An auxiliary tree cannot have an optional foot node, by definition. We thus have to make sure that no sentential complement, or more precisely no sentential complement out of which extraction is allowed (as these are the complements that we add by adjunction rather than by substitution), can be optional. This seems to be the case so far:

(201) *John said that he was late.*

(202) * *John said.*

But this observation requires further study. When there is an option between considering the optional element to be an NP or an S, we will of course favor choosing an NP.

11.4 Syntactic and Lexical Rules

The distinction (made by Wasow (1977)) between syntactic rules, which are regular and are meaning-, category- and argument structure- preserving, and lexical rules which have lexical idiosyncrasies and can be argument structure- and category- changing as well as meaning altering, is easy to represent in a TAG grammar. Structures built by the former are not marked by any special features in the tree family, and every verb with the corresponding argument structure selects for them. Structures associated with the latter produce trees that may either be kept in a larger tree family and called upon lexically by the relevant verbs with the proper feature constraints, or correspond to separate syntactic entries (different tree families) for the same verbs, the possible semantic relations between the two verb forms being stated at another level of grammar. It seems to us that it makes sense to keep in the same tree family the resulting structures of lexical rules that do not change the syntactic category of the head.

For example, where a verb can undergo an ergative alternation, its lexical entry may contain the feature $\langle \text{ERG} \rangle = +$, which matches with an $\langle \text{ERG} \rangle = +$ tree in the transitive tree family.³⁶ The lexical entry for *bake* will include an extra tree structure as a result of this.

Passive and dative alternation follow the same convention. A feature on a verb in the lexicon allows it to take or not take certain tree-structures in the tree family(ies) it selects.³⁷ The feature $\langle \text{DAT} \rangle = +$ in the lexical entry of 'give' will select the set of trees that correspond to the basic argument ordering, 'John gave Mary some books.' The lexical entry of 'roll' will have the feature $\langle \text{DAT} \rangle = -$, so that these tree structures

³⁶Work is currently being done on the implementation of the ergative construction by Beth Ann Hockey.

³⁷Work on the implementation of these features is being done by Megan Moser and Beth Ann Hockey, respectively.

will not be able to be instantiated with 'roll' as the head. (*John rolled the wagon to the driveway/* John rolled the driveway the wagon*)

When a verb in a certain structure can be passivized, its lexical entry includes (for the corresponding tree family) a feature to indicate that the corresponding set of trees (with the passive surface order, and the appropriate change of features) are possible with that verb as the head.

We also have to add into each family tree structures for cleft-extraction and topicalization. We are exploring the possibility of using the same structures for wh-questions and topicalization. The final characteristic of the sentence will come from having an NP (wh)=+ or (wh)= - in Comp position. The only complication with this approach is the need for different inversion features in the two cases, i.e. the elementary tree for the wh-question must force the adjunction of an inverted S or auxiliary, and the topicalization must not allow such an adjunction.

We want to explore how to state more explicitly which principles govern the well-formedness of a tree family, given the corresponding predicate-argument structure. We might view such principles as metarules actually allowing to generate most elementary trees out of some more basic ones. However, it seems to be the case that some predicates (especially idiomatic ones) will select directly such generated trees without selecting for the corresponding 'basic' ones. This would argue for rather considering such metarules as non-oriented rules.

More generally, we want to explore more thoroughly the borderline between syntactic and lexical rules and the correct way as how to state the distinction. It seems for example that lexical idiosyncracies can be found for most of the so-called 'syntactic' rules (Abeillé 1990). This is especially true if one wants to use the same grammar for light verb constructions, idioms and 'free' sentences. It seems for example that even wh-question and topicalization should be lexically marked for most idiomatic constructions.

12 Conclusion

The English grammar we have described here was built using a lexicalized, feature-based form of Tree Adjoining Grammar (LTAG). The grammar is basically organized as a lexicon, and elementary tree structures are grouped together into sets called tree families. We have described principles that affect the size and structure of elementary trees, especially the benefits of capturing dependencies within an elementary tree. We have shown that extraction properties can thus be accounted for as constraints on the structure of the elementary trees. In particular, we have shown how extraction out of S-complements is naturally represented in our framework without the use of intermediate traces, and how extractions are ruled out for relative clauses and indirect questions. We have described a straightforward analysis of pied-piping and preposition-stranding in prepositional complements, and have shown the types of structures used to handle auxiliaries in inverted and non-inverted structures. We have described how similar structures can also be used to handle various types of raising constructions. Using multi-component anchors, we have also provided a natural account of verb-particle constructions, ‘light-verb’ constructions, and idioms.

We have been able to take advantage of the distinction available in the TAG formalism between derivation trees and derived trees (differently from Context free Grammars). We have thus assigned more extended elementary trees with multicomponent anchors to complex predicates while making them generate the same derived trees as more simple predicates. As a syntactic consequence of these larger trees, we have accounted for ‘pseudo passive’ or CNPC violations in the case of verb-preposition combinations or light-verb constructions, while avoiding the need for ‘reanalysis’. As a semantic consequence, we have been able to make non-compositional predicates fit naturally into a linguistic representation which is compositional by elementary structures.

The last two sections detailed some of the ongoing and future work in this project. The discussion of ongoing work included treatments of negation, PRO and case assignment, and anaphors. The future work section outlined some areas of interest for future research such as copular constructions and small clauses, optional arguments, adverb movement, and the general nature of syntactic and lexical rules.

We have attempted overall to give a clear description of the fundamentals of the English LTAG. In doing so, we have outlined some of the current research areas and the analyses being considered within these areas. These analyses should be suggestive of the kinds of generalizations we expect to be able to make within the LTAG framework.

In general, we expect to continue our exploration of the theoretical benefits of TAG’s extended domain of locality and of the lexicalization of the grammar. In particular, we need to develop a formal definition of lexical anchors that clarifies their uniqueness from both traditional syntactic ‘heads’ and semantic ‘functors’, or that, more properly, explains the extent to which anchors are the union of these two entities.

In particular we are working on the development of a semantic representation, which will benefit from TAG’s compositionality with respect to elementary structures rather than words, and which exploits the difference between derivation structures and derived structures. We intend to explore the extent to which such a semantics can be lexicalized. Related to this, we will continue our exploration of the uses of multi-component anchors.

We believe that the expansion of the set of abstract elementary trees, as well as the development of new features of various types, will undoubtedly be influenced by the ongoing expansion of the lexicon. Many supposed idiosyncrasies of the grammar may turn out to be patterns that can be handled naturally within our framework.

A APPENDIX 1: Organization of the syntactic lexicon

The full syntactic categories of the lexical items are the tree (or set of tree) structures they select. More traditional categories such as verb, noun, preposition, adverb or adjective are used to refer to the address of the head in such elementary trees. They are also used to link the morphological and the syntactic lexicons. They could be used for more theoretical linguistic generalizations as well.

For now the size of the syntactic lexicon is as follows:

Nouns : 350
 Verbs: 750
 Prepositions: 40
 Adjectives :1500
 Adverbs : 50

Our perspective has been so far to enter the most common words in priority, using (Francis and Kucera 1982) as the reference for ranking.

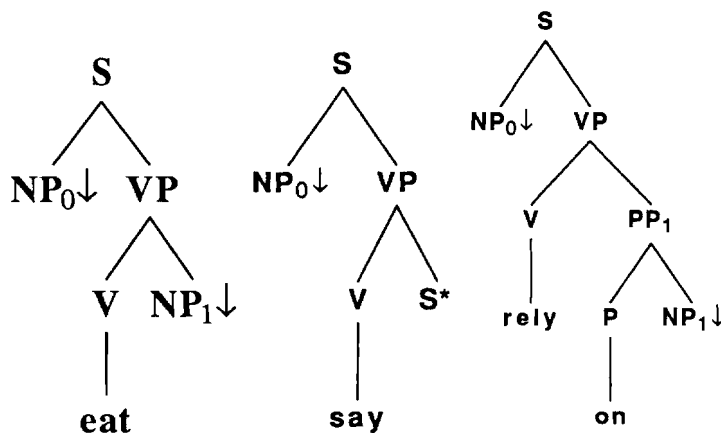
Idioms have to be added, and the different families corresponding to their more expanded elementary trees have to be added too. A sizable coverage of idiomatic phrases is crucial if one wants to parse current written texts. Idioms occur in real texts much more frequently than is usually thought (and nearly as much as free sentences, if one takes light verb constructions into account) as has been shown by M.Gross 1984.

A.1 Verbs

The classifications for main verbs, ignoring features, are the (sets of) sentential structures that they head. Two verbs have the same basic classification if they introduce the same tree structures. Generally, this means that they have the same subcategorization frame. More specific classification is determined by features. Examples of verb entries, without features, are as follows:

\EAT\, V: Tnx₀V, Tnx₀Vnx₁.
 \THINK\,V: Tnx₀VS₁.
 \RELY\,V: Tnx₀Vpnx₁

Examples of trees are the following:



A.2 Auxiliaries

Auxiliary verbs are defined as auxiliary structures without arguments of their own. The basic structures used are very similar to those used for adverbs and other modifiers. The constraints on the use of each auxiliary are represented by features (see Appendix 2 for a list.) Below are some examples of lexical entries for auxiliaries:

Morphological Lexicon

```
does:\DO\, V{V.b:<mode>= ind; :V.b:<agr pers>= 3, V.b:<agr num>=singular,
V:<agr 3rdsing>= +}. did:\DO\, V{V.b:<mode>= ind} doing:\DO\, V{V.b:<mode>= ger}
```

Syntactic Lexicon

```
\DO\, V{V.b:<mode>= ind} :  $\beta$ vS {
Sr.b:<mode>= ind,
Sr.b:<tense>= pres,
S.b:<mode>= base,
S.b:<conditional>= -,
S.b:<passive>= -,
S.b:<perfect>= -,
S.b:<progressive>= -,
S.b:<agr>= V.b:<agr>}.

```

```
\DO\, V :  $\beta$ vVX {
VPr.b:<mode>= ind,
VPr.b:<tense>= pres,
VP.b:<mode>= base,
VP.b:<conditional>= -,
VP.b:<passive>= -,
VP.b:<perfect>= -,
VP.b:<progressive>= -,
VPr.t:<agr>= V.b:<agr>}.

```

```
\HAVE\, V{V.b:<mode>= ind} :  $\beta$ vS {
Sr.b:<mode>= ind,
Sr.b:<tense>= pres,
Sr.b:<perfect>= +,
S.b:<mode>= ppart,
S.b:<passive>= -,
S.b:<agr>= V.b:<agr>}.

```

```
\HAVE\, V :  $\beta$ vVX {
VPr.b:<mode>= ind,
VPr.b:<tense>= pres,
VPr.b:<perfect>= +,
VP.b:<mode>= ppart,
VP.b:<passive>= -,
VPr.t:<agr>= V.b:<agr>}.

```

```
to, V :  $\beta$ vVX {
VPr.b:<mode>= inf,
VP.b:<mode>= base}.

```


A.3 Complementizers

Complementizers are incorporated in the current state of the grammar as auxiliary structures that adjoin to the S trees for sentential arguments. The basic tree structure for a complementizer is βCompS . Complementizers assign $\langle\text{COMP}\rangle$ features to both the root and the foot S nodes, and $\langle\text{MODE}\rangle$ to the foot S node. These features handle the co-restrictions between the complementizer and the mode of the sentence it adjoins into, as well as those between the matrix verb and the complementizer of the complement clause it adjoins into. (Such features are discussed in the next subsection). The optionality of *that* in sentences such as *John thinks (that) Mary is lying* is also represented straightforwardly without having to deal with empty complementizers.

However, we are not completely committed to this representation, especially for semantic reasons. It is not clear that any natural semantics can be assigned to either *that* or *for*. If we wish to argue that an elementary tree is both a syntactic and a non-empty semantic unit, we will want to directly attach these complementizers to an elementary tree that already has some semantics. In this case we would include ‘that’ and ‘for’ complementizers in matrix clause structures, which would thus have multicomponent anchors: the complementizer and the verb. (A similar representation could be used for relative pronouns, with trees for relative clauses having multicomponent anchors consisting of the verb and the relative pronoun.) The alternative is to allow an ‘empty’ semantics for these complementizers.

A.4 Nouns and Pronouns

Nouns all select initial trees rooted in NP. They are further based on whether or not³⁸:

- The noun subcategorizes for complements
- The noun requires a determiner
- The noun can be modified by an adjective

When a noun requires a determiner, it selects the tree structure ‘ αNPdn ’, when it does not, it selects the structure ‘ αNPn ’. Plural forms of noun select both. For singular forms, the distinction has to be lexically specified. *Water*, for instance, can take both, while *chair* only takes the first one (at least in the concrete, non human sense of *chair*).

Wh-terms like *who* and *what*, pronouns and names are defined as ‘ αNP ’, which means that they take neither determiners nor adjectives nor complements of any kind, but they can occur wherever any NP can occur and be coordinated with any NP.³⁹ Having only an NP node available for names also prevents restrictive relative clauses from adjoining to them.

Nouns taking complements select not only NP trees but also sentential trees. The structural classifications for these predicate nominals are essentially equivalent to the set of tree families for verbs, and depend on their subcategorization frame. The V node in these structures is filled by a ‘light’ verb, as explained in section 4.

Below are some lexical entries for nouns, along with some examples of the trees that these entries represent:

house, N: αNPdn .

water, N: αNPdn , αNPn .

he : $\alpha\text{NP}\{\text{NP.t:}\langle\text{PRO}\rangle=+\}$

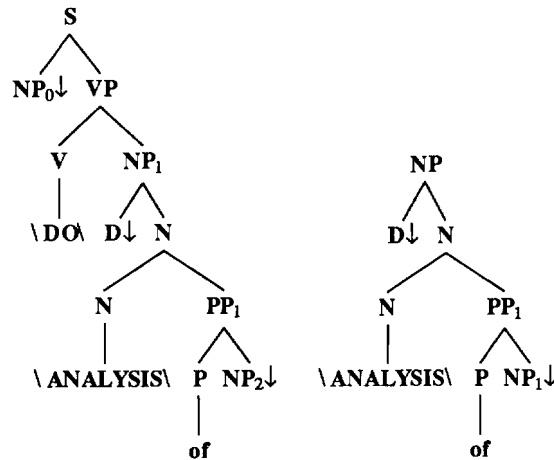
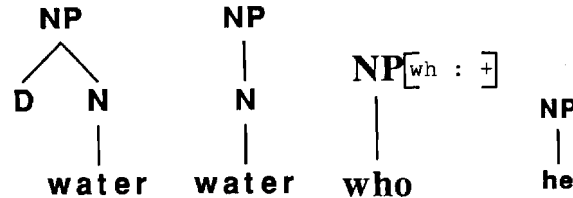
³⁸We are ignoring compound nouns for the moment, although a similar treatment as the one proposed for idioms could be considered.

³⁹They may also differ from other nouns by features $\langle\text{WH}\rangle=+$ and $\langle\text{Pro}\rangle=+$

who : $\alpha\text{NP}\{\text{NP.t:}\langle\mathbf{WH}\rangle=+\}$

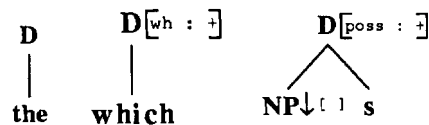
Mary: αNP

analysis,N; \DO\,V: Tnx₀lVdn₁pnx₁₁[P1=of].



A.5 Determiners

Some lexical items have very little structure. Determiners⁴⁰, for instance, have simple syntactic definitions in our grammar because the tree that they introduce is only a determiner node. The structures corresponding to ‘the’, ‘which’ and ‘s’ are shown below.



A.6 Adjectives

Adjectives are classified on three different properties⁴¹ :

- Their ability to add a modifier structure to an NP or to fill an AP position in copula constructions and ‘small clauses’,

⁴⁰Leaving aside quantifiers and partitives for the moment

⁴¹Only Property 1 has been entered so far in the TAG Lexicon

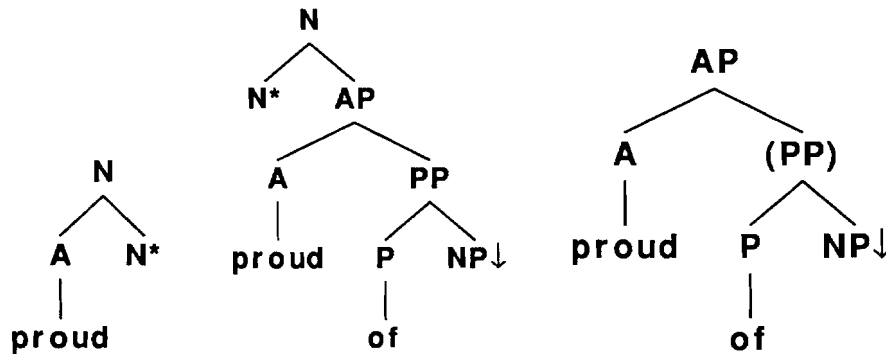
- Their ability to take complements or not, and
- The nature of the complement they can take.

Most adjectives can modify nouns. They are therefore defined in the lexicon as βaN , which is an auxiliary tree that adjoins at an N-node, with the adjective preceding the noun. An adjective with a realized complement, however, cannot occur in this position; it will thus select other auxiliary trees (βN_{apn1} , or βN_{aps1} depending on whether they take nominal or sentential arguments.)

(203) * *A proud of himself man*

(204) *A man proud of himself*

If the adjectival complement is optional, as is most often the case, the adjective will also select βaN ⁴². Examples of trees are the following :



Certain adjectives cannot be used as modifiers. *ill*, for example, cannot be used in this context.⁴³:

(205) * *I saw an ill man*

Adjectives of this type are perfectly acceptable in copula constructions and small clauses, and they are thus defined as αAP_a .

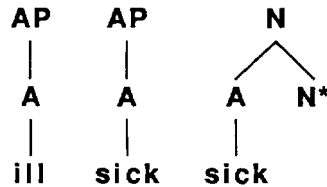
(206) *John is ill.*

Many adjectives, of course, fall into both classifications. For example:

pink, A: βaN , αAP_a .

(*Mary likes pink flowers, The roses are pink.*)

Examples of trees are:



⁴²The only adjectives with obligatory complements we know of are : loath to, rich with???, able to ??? (with the same meaning)

⁴³Except in the compound: *an ill-effect*

The adverbs adjectives can take might depend on whether the adjective is used as a predicate or as a modifier:

- (207) *The task is easy.*
- (208) *It is an easy task.*
- (209) *This task is quite easy.*
- (210) * *This a quite easy task.*

- (211) *This suggestion is utterly absurd.*
- (212) * *This an utterly absurd suggestion.*

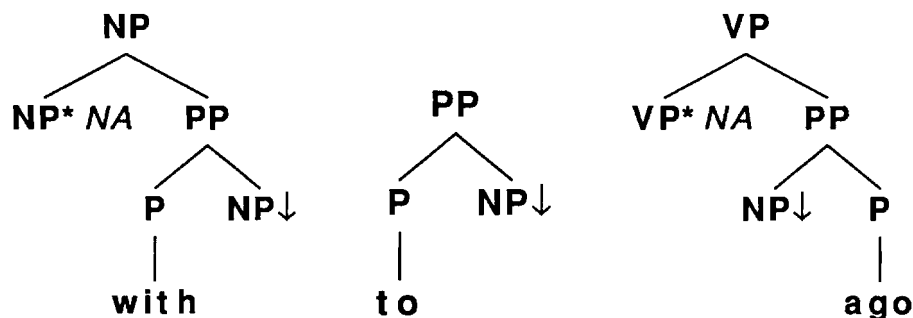
This will be done by having different adjunction constraints associated with modifiers trees and predicate trees.

A.7 Prepositions

Prepositions may also be defined by simple lexical trees, when the PP complement is subcategorized for by the verb. But, for semantic reasons, we may sometimes consider them as directly attached⁴⁴. For prepositional adjuncts, the preposition, as the head of a PP, introduces the entire auxiliary structure. The structures are differentiated depending on:

- Which category (or categories) the PP modifies i.e. attaches to.
- Whether the so-called ‘preposition’ follows or precedes its NP argument
- Whether the preposition takes an NP argument, an S argument or both.

Examples of trees for prepositions are the following:



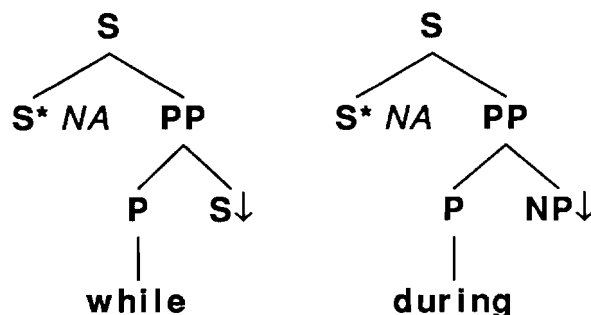
Examples of ‘postposed’ prepositions are *notwithstanding*, or *ago*, but *notwithstanding* can occur both ways. Currently, we do not consider subordinating conjunctions to be a separate category. We only specify which prepositions take NP arguments, which take S arguments, and which take both, eg.:

- (213) *John took a bath while Mary was eating / * while the evening.*
- (214) *John took a bath * during Mary was reading / during the evening.*
- (215) *John has been taking a bath since Mary started reading/ since the beginning of the evening.*

Sample trees for prepositional modifiers of S are shown below⁴⁵.

⁴⁴This is mostly likely with non-locatives. See section 3-4 for further discussion

⁴⁵The motivation for having the sentential argument of the preposition substituted is explained later.



Different elementary structures for a specific word may correspond to a difference in meaning. The temporal *since*, for example, subcategorizes for either NP or S, whereas the causal *since* subcategorizes for S only. *since* will therefore be listed as two different lexical entries.

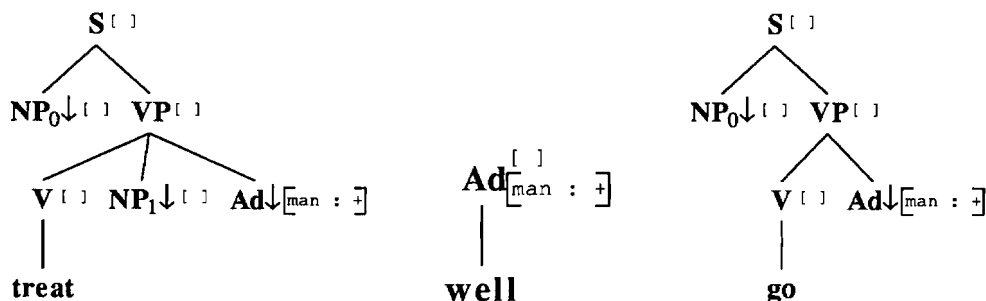
A.8 Adverbs

Adverbs can be used either as arguments, subcategorized for by such verbs as *go* or *treat*, and more commonly as modifiers, classified by the type of adjunct structure they can add to a tree. The first type are obligatory and have a fixed position:

- (216) *John treats Mary rottenly.*
- (217) * *Rottenly John treats Mary.*
- (218) *The project is going smoothly.*
- (219) * *Smoothly the project is going.*

Without the adverb, (218) is ungrammatical and (216) has a totally different meaning.⁴⁶

- (220) * *The project is going*
- (221) *John treats Mary* (he pays for her)



Adverbs used as modifiers, on the other hand, are optional and can usually occur in various positions. They are represented as auxiliary trees rooted with the category of the node they modify. Currently, the structures we are using for the English lexicon are limited to right and left attachment to S, and right and left attachment to VP. These trees are meant to provide the proper scope for each adverb. Additional structures will be added to handle attachment to APs and other categories. Some adverbs may be marked with tense

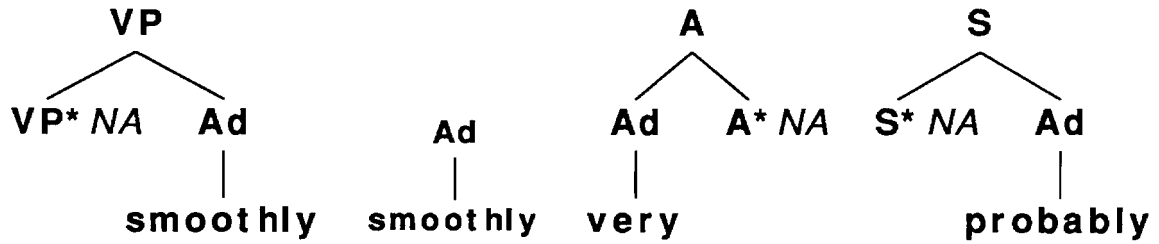
⁴⁶The feature *man* would indicate whether an adverb is a “manner” adverb or not.

or aspect information which will have to unify with the node it adjoins to. Most adverbs, however, do not carry any features at this point. In order to allow for word order variations, rules such as those defined in Joshi (1987) may have to be used. For the moment, a typical entry for a sentential adverb is as shown below:

smoothly, Ad: α Ad, β VPad

probably, Ad: β Sad, β adS.

very, Ad : β adA



B APPENDIX 2 : List of the Feature Structures used

In the lexicon, a feature may be assigned by a lexical item to the top or bottom feature structure of any node in its definition (i.e., anywhere in the structure that lexical item introduces.) Usually, bottom features are more 'lexical' and come from the head of the structure, while top features are more syntactic (or semantic) and are constrained by the tree structure itself (usually making an adjunction obligatory).

This section provides a description of the most important features currently used in the lexicon of the English LTAG. We have not felt a need for recursive feature structures, either as reentrance fs associated with nodes in the elementary trees, nor as the result of unification after different trees have been combined together.

Portions of this lexical feature system still have to be developed. In particular, we need to decide whether or not to include thematic roles or more specific semantic roles (as in HPSG). A restricted set of thematic roles was considered in Bishop (1988) as a possible approach to handling control phenomena.

B.1 The features \langle MODE \rangle and \langle COMP \rangle

When a predicate term takes a sentence as an argument, the features \langle MODE \rangle and \langle COMP \rangle are used to constrain the type of sentence acceptable in that position for that predicate term. The current possible values for \langle MODE \rangle are: indicative (ind), infinitive (inf), base (bse), past participle (ppart), and gerundive (ger). ' \langle MODE \rangle = ind' is used if the predicate term requires a tensed subject. ' \langle MODE \rangle = inf' is used if a to-infinitive is needed. ' \langle MODE \rangle = bse' indicates the need for a tenseless sentence; this includes bare infinitives and the so-called english subjunctive. Finally, ' \langle MODE \rangle = ger', when used in this context, indicates that a gerund should be used. (More generally, this feature is used to signal the presence of a verb in -ing form. Because the \langle MODE \rangle value of the top verb is coindexed with the S node, marking a sentential complement with ' \langle mode \rangle = ger' effectively indicates a gerund.) \langle MODE \rangle is assigned to the bottom feature structure of a node, preventing it from passing up through adjunction.

The feature \langle COMP \rangle specifies what complementizer, if any, may be adjoined to an S node. When \langle COMP \rangle is specified in the lexicon by a predicate term, it serves as an additional constraint on sentential complements. The possible values for \langle COMP \rangle are : that/whth (whether)/if/for/nil/none.⁴⁷ For example, the verb *think* requires either no complementizer or *that* and this information is represented in the lexicon as:

\backslash THINK\ \backslash , V: Tnx₀Vs₁{s₁.b: \langle Comp \rangle = that/none}, Tnx₀Vps₁ {p=of}.

For verbs taking prepositional sentential complements, there are no lexical variations regarding \langle Comp \rangle and \langle Mode \rangle . Their value (resp. none and gerund) are thus stated directly at the level of tree families without appearing in the lexical entry of the matrix verb.

The difference between the feature values 'nil' and 'none' is that, while 'none' is used to represent an S node that can not have a complementizer adjoined to it, 'nil' represents one which does not have a complementizer but is a possible adjunction node for comp.

Mode is also selected by prepositions in subordinate clauses. *Because*, for example, requires that the mode of the clause be indicative, while *when* allows ind or ger, and *in order* allows inf:

(222) *John is happy because he got a job.*

(223) * *John is happy because getting a job.*

⁴⁷Notice that wh-terms other than *whether* are not treated as complementizers. As described above, these other wh-terms are substituted into an NP node in an elementary tree.

- (224) * *John is happy because to get a job.*
 (225) *When he killed Mary, John was unhappy.*
 (226) *When killing Mary, John was unhappy.*
 (227) *John came in order to see Mary.*

B.2 Tense and Related Features

Both simple tenses and the complex tense/aspect structures formed with auxiliary verbs are described with lexical features. These features are also used to prevent ungrammatical auxiliary structures. The feature $\langle \text{TNS} \rangle = \text{pres/past/fut}$ is marked on all and only those verbs in the indicative mode ($\langle \text{MODE} \rangle = \text{ind}$). See also the section of this report on auxiliaries. Corestrictions between the tense of the matrix clause and that of adverbial clauses or temporal adverbials such as *today* or *tomorrow* are stated with similar features.

Here's a summary of all the features which the auxiliary system affects:⁴⁸

Feature	Possible Values	Meaning	What requires it
$\langle \text{mode} \rangle$	base,ind,inf,ger,ppart	verb form	aux verbs, governing verb
$\langle \text{passive} \rangle$	+,-	passive arg structure	perfect,passive aux
$\langle \text{conditional} \rangle$	+,-	conditional mood	possibly higher verb
$\langle \text{perfect} \rangle$	+,-	perfect aspect	possibly higher verb
$\langle \text{progressive} \rangle$	+,-	progressive aspect	possibly higher verb
$\langle \text{tense} \rangle$	past,pres	for $\langle \text{mode} \rangle = \text{ind}$	possibly higher verb
$\langle \text{agr} \rangle$	(complex)	person,number	subject-verb agreement
$\langle \text{inv} \rangle$	+,-	inverted S	only 1 aux fronted, only matrix S

B.3 Agreement Features

In English, agreement features are marked on verbs, nouns, pronouns, and quantifiers, as well as on some determiners and adjectives (eg., “various”). Quantifiers and determiners are marked only for number agreement. Agreement in the English LTAG is marked by the following lexical features:

- $\langle \text{AGR num} \rangle = \text{sing/plur}$
 $\langle \text{AGR pers} \rangle = 1/2/3$
 $\langle \text{AGR gen} \rangle = \text{fem/masc}$

The feature types, $\langle \text{AGR pers} \rangle$ and $\langle \text{AGR num} \rangle$ are used to limit lexical choice. For most verbs, a distinction is only made in the present tense and then only between those verbs that are both $\langle \text{AGR pers} \rangle = 3$ and $\langle \text{AGR num} \rangle = \text{sing}$, and those that are not (i.e., $\langle \text{AGR pers} \rangle = \sim 3$ or $\langle \text{AGR num} \rangle = \text{plur}$).

In the English LTAG, $\langle \text{AGR gen} \rangle$ it is marked only on pronouns, possessive determiners, and a few select nouns (eg., “mother”, “boy”, and perhaps “boat”), and will be used to constrain lexical choice for bound anaphora :

*John is proud of himself/*herself*

and for obligatorily possessive determiners in sentences such as:

- (228) *John gave his/*her word to Sophie that he will help her.*
 (229) *Mary gave her/*his word to Sophie that she will help her.*

⁴⁸This list was compiled by Megan Moser and appears in her report on the addition of passives to the auxiliary system.

B.4 Selectional Restrictions

Selectional restrictions currently have a unique status in the grammar. Although they are represented like other features, they are not used to constrain a parse. If the unification of two selectional restrictions fails, a sentence is given a marked semantics but the parse still succeeds. The three features which now represent the selectional restrictions are grouped together in the grammar under the heading **<RESTR>**. The specific features and their possible values are:

<RESTR hum> = +/- (used to distinguish human from non-human)
<RESTR anim> = +/- (used to distinguish animate entities from inanimate ones)
<RESTR conc> = +/- (used to distinguish concrete objects from abstract concepts)

A shorthand has been developed for the possible combinations assignable to a noun in the lexicon:

<RESTR>=HUMAN is
<RESTR hum>=+, **<RESTR anim>**=+, **<RESTR conc>**=+.

<RESTR>=CRE is
<RESTR hum>=-, **<RESTR anim>**=+, **<RESTR conc>**=+.

<RESTR>=OBJ is
<RESTR hum>=-, **<RESTR anim>**=-, **<RESTR conc>**=+.

<RESTR>=ABS is
<RESTR hum>=-, **<RESTR anim>**=-, **<RESTR conc>**=-.

Because a verb will often select for some combination of these possibilities in its arguments, the distinguishing feature may be the simplest representation in this part of the lexicon. For example:

fall, V: Tnx₀v {nx₀.t:**<RESTR conc>**=+}. (not - **<RESTR>**=HUMAN/ CREAT/OBJ.)

We need more selectional features in order to properly constrain the adjunction of modifiers such as adjectives or adverbs.

C APPENDIX 3 : Samples of structures handled

The current size of the grammar in terms of tree-families is 82 basic groupings. The basic breakdown of these groupings is (See below for a complete listing):

- Sentences headed by verbs taking nominal or prepositional arguments (= 8 Tree Families)
- Sentences with verbs taking sentential arguments (= 21 Tree Families)
- Light Verb-Noun constructions (= 17 Tree Families)
- Light Verb-Adjective constructions (= 6 Tree Families).
- Verb-Particle combinations (= 15 Tree Families)
- idioms (= 15 Tree Families)

Expanded versions of most of these tree families will need to be added to handle more idioms. Each of these tree families currently contains between 3 and 12 trees .

C.1 Notational Conventions

A tree-family's name corresponds to a basic surface ordering of its arguments in the declarative form. Nominal arguments are denoted by nx_i . These are distinguished from sentential arguments, which are marked by s_i . An adjectival argument is marked by ax_i , and a prepositional phrase that is a basic argument of the verb (as in "John clings to Mary"), is marked by pnx_i or ps_i . In all cases, i is the number of the argument position filled by that node.

The canonical subject, for instance, is always assigned '0' (eg. nx_0 or s_0) The family, " Tnx_0vnx_1 ," denotes all the structures for a transitive verb. The family, " Tnx_0vpnx_1 ," denotes all the structures for any verb that subcategorizes for a single prepositional argument. Normally, direct objects are marked with a '1' and higher numbers are used for other complements. In tree families where arguments are expanded, the numbering on the additional structure is preceded by the number of the argument it is attached to. For example, in a light-verb construction where the predicate nominal nx_1 requires a prepositional argument of its own, this is described as nx_1pnx_{11} . Since names of individual tree structures also represent their surface order, these numbers will not always be sequential. (For example, our analysis of dative movement places the tree $\alpha nx_0vnx_2nx_1$, in the tree family, " $Tnx_0vnx_1pnx_2$.")

The following list provides the notational information necessary to read or write the names of tree structures and tree families in an LTAG: (Note that x means "maximal projection.")

- nx = Noun phrase
- dn = Noun phrase expanded to N with a determiner
- n = Noun phrase expanded to N without a determiner
- v = verb
- vx = verb phrase
- ad = adverb
- a = adjective
- ax = adjective phrase
- p = preposition (or particle)
- pnx = prepositional phrase (or 'pdn' or 'pn' if the noun phrase is expanded)
- s = sentence (used for sentential arguments, etc...)
- C = Complementizer
- R = Extraction in a relative clause
- W = Wh-extractions
- I = It-cleft Constructions

In addition, we use the following affixes:

T = prefix attached to the name of a tree family

α = prefix attached to the name of an initial tree

β = prefix attached to the name of an auxiliary tree

In the examples, anchors are put in bold face.

C.2 List of Tree Families

C.2.1 Verbs with nominal arguments

Tnx0V

John **walks**.

Tnx0Vnx1

John **watches** Mary.

Tnx0Vpnx1

John **departed** from Philadelphia.

Tnx0Vnx1pnx2

John **sold** a book to Mary.

Tnx0Vpnx1pnx2

John **spoke about** Linguistics to Mary

Tnx0Vnx1nx2

Mary **called** her son Jim.

Tnx0Vax1

John **looks** confused.

Tnx0Vnx1ax2

John **makes** Mary happy.

C.2.2 Verbs with sentential arguments

Ts0V

That John likes Mary **stinks**.

Ts0Vnx1

Sending letters to the Congress **helps** the cause.

Tnx0Vs1

John **proved** that this could be done.

Ts0Vs1

Setting a good example **shows** that you care.
Ts0Vpnx1
Waging war **leads to** destruction.
Tnx0Vps1
John **thinks about** going skiing.
Ts0Vps1
Living in Paris **differs from** living in New York.
Tnx0Vsls2
John **considers** working in New York City living in hell.
Ts0Vnx1nx2
That he is an honest man **makes** Bill a good President.
Tnx0Vnx1s2
John **persuaded** Mary that the world should be changed.
Tnx0Vnx1ps2
Bill **prevented** Mary **from** doing the dishes.
Tnx0Vslps2
John **equates** going to Epcott Center **with** going to Europe.
Tnx0Vslpnx2
John **assigned** writing the introduction **to** Bill.
Ts0Vslpnx2
That Mary rolled her eyes **indicated** that she was disgusted to Bill.
Ts0Vnx1ps2
Living in a monastery **prevents** John **from** getting into trouble.
Ts0Vnx1pnx2
For Mary to give up now would **put** all the responsibility **on** Bill.
Tnx0Vps1pnx2
John **talked to** Mary **about** going swimming.
Ts0Vpnx1pnx2
That John suffered without retaliating **spoke about** his character **to** everyone .
Ts0Vs1ps2
Being tired **turns** going next door **into** running a marathon.
Ts0Vax1
That John saw a flying saucer **looks** incredible.
Ts0Vnx1ax2
That John returned home **left** Mary speechless.

C.2.3 Light-verb constructions

These constructions are headed by a multicomponent anchor comprising the light verb and the predicate nominal. Any argument of the predicate nominal is thus an argument of the sentential structure, and extraction out of it occurs freely.

Light-Verb Predicative nouns combinations with Nominal arguments

Tnx0VNX1
John **gave a cry**. (John's cry)
Tnx0VPNX1
John **is in a good mood**. (John's (good) mood)
Tnx0VNX1pnx11
John **has an influence on** Mary. (John's influence on Mary)
Tnx0VNX1pnx11
John **gave a talk on** the passive. (John's talk on the passive)

Tnx0VPNx1pnx11

John **is in contact with** the media. (John's **contact with** the media)

Tnx0VNX1pnx11pnx2

This book **gives** a new **analysis of** the world to its readers. (This book's **analysis of** the world)

Light-Verb Predicative noun combinations with sentential arguments

Ts0VNX1pnx11

Losing his driver's license **had an effect on** John.

(The **effect of** losing his license (on John))

Ts0VPNx1pnx11

Helping the poor **is in** the best **interest of** everyone.

Ts0VNX1pnx11pnx2

Voting in the election **gives hope for** progress to the country's people.

Tnx0VNX1s11

Mary **has a feeling that** Bill loves her. (Mary's **feeling that** Bill loves her)

Ts0VNX1s11

Giving someone your housekeys **makes the point that** you trust him.

Tnx0VNX1ps11

John **had a say in** starting the new company. (John's **say in** starting the company)

Ts0Vnx1ps11

Taking a test **bears some resemblance to** walking the plank. (The **resemblance of** taking a test to walking the plank)

Tnx0Vpnx1ps11

John **is in the habit of** quitting. (John's *habit of* quitting)

Tnx0Vpnx1s11

John **is in the mood to** dance.

Ts0Vpnx1ps11

Eating fish raw **is in the process of** becoming quite popular.

Ts0Vpnx1s11

Indicting Exxon **is on the presidential agenda to** crack down on polluters.

Light-verb Predicative adjective combinations

Tnx0VA2pnx21

John **is proud of** his results

Ts0VA1pnx11

Lying to one's friend **is equivalent to** a crime.

Tx0VA1s11

John **is happy that** everyone likes him.

Ts0VA1s11

That it will snow in January **is liable to** happen.
apartment.

Ts0VA1ps11

Ignoring racism **is influential in** propagating it.

Tnx0VA1ps11

Mary **feels tired of** living in a small apartment.

C.2.4 Verb-particle combinations

The particle is a preposition which does not head any PP sub-tree. It usually is part of the clausal anchor with the verb.

Verb-particle combinations with nominal arguments

Tnx0VP1

The clock **fell apart**.

Tnx0VP1nx2

John **wolfed down** a hot-dog

Tnx0VP1pnx2

Mary **found out about** the party.

Tnx0Vp1nx2pnx3

John **handed over** his gun **to** the police.

Tnx0VP1pnx2pnx3

John **lectured on about** Linguistics **to** Mary.

Tnx0VP1ax2

The souffle **came out** perfect.

Tnx0VP1ax2pnx3

The results **showed up** negative on the screen.

Verb-particle combinations with sentential arguments

Ts0VP1

That John arrived early **worked out** somehow.

Ts0VP1nx2

That it could still snow in March **brings me down**.

Tnx0VP1s2

John **found out** that this could be done.

Ts0VP1s2

That the rain never stops **goes on** perplexing Bill.

Tnx0VP1ps2

John **gave up on** asking Mary to date him.

Tnx0Vp1nx2(ps3)

The new dishwasher **freed up** John **from** doing the dishes.

Light-verb particle combinations

Tnx0VP1Nx2pnx21

John **worked out** a study of ellipsis.

Tnx0VP1Nx2ps21

John **took over the burden of** explaining the situation to Mary.

C.2.5 Idioms

TDN0VdN1

The world is your oyster.

TDN0VP1Pnx2

The roof caved in on John.

Tnx0VDN1

John kicked the bucket.

Tnx0VdN1

John got Mary's goat.

Tnx0VDAN1

John talks a good game.

Tnx0VAN1

John's past is open territory.

Tnx0VPDN1

The project went to the dogs.

Tnx0VPN1

John is treading on eggs.

Tnx0VDN1A2

John painted the town red.

Tnx0VN1Pnx2

John took umbrage at the project.

Tnx0VDN1Pnx2

John thumbed his nose at Mary.

Tnx0Vnx1PDN2

John took Mary to the cleaners.

Tnx0Vnx1PN2

John took Mary's words into account.

Tnx0Vnx1A2

John sold Mary short.

Tnx0VN1PDN2

John is building castles in the air.

Tnx0VDN1PDN2

John took the bull by the horns.

References

- Abeillé, Anne, 1990. Lexical Constraints on Syntactic Rules in a Tree Adjoining Grammar. To appear in the proceedings of the 28th Meeting of the Association for Computational Linguistics (ACL'90).
- Abeillé, Anne, August 1988a. Parsing French with Tree Adjoining Grammar: some Linguistic Accounts. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING'88)*. Budapest.
- Abeillé, Anne, 1988b. *A Lexicalized Tree Adjoining Grammar for French: the General Framework*. Technical Report MS-CIS-88-64, University of Pennsylvania.
- Abeillé, Anne, 1988c. Light Verb Constructions and Extraction out of NP in Tree Adjoining Grammar. In *Papers from the 24th Regional Meeting of the Chicago Linguistic Society*. Chicago.
- Abeillé, Anne and Schabes, Yves, 1989. Parsing Idioms in Tree Adjoining Grammars. In *Fourth Conference of the European Chapter of the Association for Computational Linguistics (EACL'89)*. Manchester.
- Abeillé, Anne and Schabes, Yves, 1990. Non Compositional Discontinuous Constituents in Tree Adjoining Grammar. In *Proceedings of the Symposium on Discontinuous Constituents*. Tilburg, Holland.
- Bach, E. and Horn, G., 1976. Remarks on Conditions on Transformations. *Linguistic Inquiry* 7(1):265-299.
- Bishop, Kathleen, 1988. Pro Control in a Lexicalized Tree Adjoining Grammar. In *Proceedings of Escol'88*. Philadelphia.
- Bishop, Kathleen, 1990. *Dictionary Interface User Manual*. Technical Report, University of Pennsylvania.
- Borsley, R., 1984. On the Non-existence of VPs. In Geest, W.de and Putseys, Y. (editors), *Sentential Complementation*. Foris, Dordrecht.
- Bresnan, J., 1982. Complementation and Control. *Linguistic Inquiry* 13(3).
- Bresnan, J., 1983. The Passive in Lexical Theory. In Bresnan, J. (editor), *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge MA.
- Bresnan, J. and Kaplan, R., 1983. Lexical-Functional Grammar: A formal system for grammatical representation. In *The Mental Representation of Grammatical Relations*. M.I.T. Press, Cambridge, MA.
- Cattell, R., 1984. Composite Predicates in English. In *Syntax and Semantics Vol 17*. Academic Press, New York, NY.
- Chomsky, N., 1981. *Lectures on Government and Binding*. Foris, Dordrecht.
- Culicover, Peter W. and Wilkins, Wendy K., 1984. *Locality in linguistic theory*. Academic Press.
- Ertshik, N., 1971. *On the Nature of Island Constraints*. PhD thesis, MIT.
- Francis, W. Nelson and Kučera, Henry, 1982. *Frequency Analysis of English Usage*. Houghton Mifflin, Boston.
- Freckelton, P., 1984. Des that-clauses prépositionnelles en anglais. *Linguisticae Investigationes* 8-2.
- Gazdar, G.; Klein, E.; Pullum, G. K.; and Sag, I. A., 1985. *Generalized Phrase Structure Grammars*. Blackwell Publishing, Oxford. Also published by Harvard University Press, Cambridge, MA.
- Gross, Maurice, 1976. Sur quelques groupes nominaux complexes. In et M. Gross, J.C. Chevalier (editor), *Méthodes en Grammaire Française*. Klincksieck.
- Gross, Maurice, 1981. Les bases empiriques de la notion de prédicat sémantique. *Langages* 63.
- Gross, Maurice, 2-6 July 1984. Lexicon-Grammar and the Syntactic Analysis of French. In *Proceedings of the 10th International Conference on Computational Linguistics (COLING'84)*. Stanford.

- Joshi, Aravind K., August 1969. Properties of Formal Grammars with Mixed Type of Rules and their Linguistic Relevance. In *Proceedings of the International Conference on Computational Linguistics*. Sanga Saby.
- Joshi, Aravind K., 1973. A Class of Transformational Grammars. In M. Gross, M. Halle and Schutzenberger, M.P. (editors), *The Formal Analysis of Natural Languages*. Mouton, La Hague.
- Joshi, Aravind K., 1985. How Much Context-Sensitivity is Necessary for Characterizing Structural Descriptions—Tree Adjoining Grammars. In Dowty, D.; Karttunen, L.; and Zwicky, A. (editors), *Natural Language Processing—Theoretical, Computational and Psychological Perspectives*. Cambridge University Press, New York. Originally presented in a Workshop on Natural Language Parsing at Ohio State University, Columbus, Ohio, May 1983.
- Joshi, Aravind K., July 1987. Word-order Variation in Natural Language Generation. In *AAAI 87, Sixth National Conference on Artificial Intelligence*, pages 550–555. Seattle, Washington.
- Joshi, Aravind K. and Vijay-Shanker, K., 1989. Unbounded Dependencies in TAGs and LFG: functional uncertainty is a corollary in TAGs. In *27th Meeting of the Association for Computational Linguistics (ACL'89)*. Vancouver.
- Joshi, Aravind K., Vijay-Shanker and Weir, David, forthcoming, 1990. The Convergence of Mildly Context-Sensitive Grammatical Formalisms. In Sells, Peter; Shieber, Stuart; and Wasow, Tom (editors), *Foundational Issues in Natural Language Processing*. MIT Press, Cambridge MA.
- Joshi, Aravind K.; Levy, L. S.; and Takahashi, M., 1975. Tree Adjunct Grammars. *J. Comput. Syst. Sci.* 10(1).
- Kaplan, R. and Bresnan, J., 1983. Lexical-functional Grammar: A Formal System for Grammatical Representation. In Bresnan, J. (editor), *The Mental Representation of Grammatical Relations*. MIT Press, Cambridge MA.
- Karttunen, Lauri, 1986. *Radical Lexicalism*. Technical Report CSLI-86-68, CSLI, Stanford University. Also in *Alternative Conceptions of Phrase Structure*, University of Chicago Press, Baltin, M. and Kroch A., Chicago, 1989.
- Kroch, A., 1988. Unbounded Dependencies and Subjacency in a Tree Adjoining Grammar. In Manaster-Ramer, A. (editor), *Mathematics of Language*. John Benjamins, Amsterdam.
- Kroch, A. and Joshi, Aravind K., 1985. *Linguistic Relevance of Tree Adjoining Grammars*. Technical Report MS-CIS-85-18, Department of Computer and Information Science, University of Pennsylvania.
- MacCawley, J., 1975. The category status of English Modals. *Foundations of Language* 12.
- Perlmutter, D., 1970. On the two verbs Begin. In *Readings in English transformational grammars*. Jacobs and Rosenbaum (eds).
- Pollard, Carl and Sag, Ivan A., 1987. *Information-Based Syntax and Semantics. Vol 1: Fundamentals*. CSLI.
- Rosenbaum, P., 1967. *The Grammar of English Predicate Complement Constructions*. MIT Press.
- Ross, J., 1967. *Constraints on Variables in Syntax*. PhD thesis, MIT.
- Schabes, Yves, 1989. *TAG System User Manual*. Technical Report, University of Pennsylvania.
- Schabes, Yves and Joshi, Aravind K., June 1988. An Earley-Type Parsing Algorithm for Tree Adjoining Grammars. In *26th Meeting of the Association for Computational Linguistics (ACL'88)*. Buffalo.
- Schabes, Yves and Joshi, Aravind K., August 1989. The Relevance of Lexicalization to Parsing. In *Proceedings of the International Workshop on Parsing Technologies*. Pittsburgh. To also appear under the title *Parsing with Lexicalized Tree adjoining Grammar in Current Issues in Parsing Technologies*, MIT Press.

- Schabes, Yves; Abeillé, Anne; and Joshi, Aravind K., August 1988. Parsing Strategies with 'Lexicalized' Grammars: Application to Tree Adjoining Grammars. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING'88)*. Budapest.
- Shieber, Stuart and Schabes, Yves, 1990. Synchronous Tree Adjoining Grammars. To appear in the 13th International Conference on Computational Linguistics (COLING'90).
- Steedman, M., 1987. Combinatory Grammars and Parasitic Gaps. *Natural Language and Linguistic Theory* 5:403-439.
- Vijay-Shanker, K., 1987. *A Study of Tree Adjoining Grammars*. PhD thesis, Department of Computer and Information Science, University of Pennsylvania.
- Vijay-Shanker, K. and Joshi, Aravind K., 1985. Some Computational Properties of Tree Adjoining Grammars. In *23rd Meeting of the Association for Computational Linguistics*, pages 82-93.
- Vijay-Shanker, K. and Joshi, A.K., August 1988. Feature Structure Based Tree Adjoining Grammars. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING'88)*. Budapest.
- Wasow, T., 1977. *Transformations and the lexicon*. Reidel.
- Wasow, T.; Sag, I.; and Nunberg, G., 1982. Idioms: An Interim Report. In *Proceedings of the XIIIth International Congress of Linguists*. Tokyo.