Parsing French with Tree Adjoining Grammar: some linguistic accounts

Anne ABEILLE*

Department of Computer and Information Science
University of Pennsylvania
Philadelphia PA 19104-6389 USA

Abstract

We present the first sizable grammar written for TAG. We present the linguistic coverage of our grammar, and explain the linguistic reasons which lead us to choose the particular representations. We show that TAG formalism provides sufficient constraints for handling most of the linguistic phenomena, with minimal linguistic stipulations. We first state the basic structures needed for parsing French, with a particular emphasis on TAG's extended domain of locality that enables us to state complex subcategorization phenomena in a natural way. We then give a detailed analysis of sentential complements, because it has lead us to introduce substitution in the formalism, and because TAG makes interesting predictions. We discuss the different linguistic phenomena corresponding to adjunction and to substitution respectively. We then move on to support verb constructions, which are represented in a TAG in a simpler way than the usual double analysis. It is the first time support verb constructions are handled in a parser. We lastly give an overview of the treatment of adverbs, and suggest a treatment of idioms which make them fall into the same representations as 'free' structures.

Introduction

Tree Adjoining Grammar (TAG) was introduced by /Joshi et al. 1975/ as a formalism for linguistic description. A TAG's basic component is a finite set of elementary trees, each of which is a domain of locality, and can be viewed as a minimal linguistic structure.

A TAG comprises of two kinds of elementary trees: initial trees (α), which are complete structures, usually rooted in S, with preterminals on all their leaves, and auxiliary trees (β), which are constrained to have exactly one leaf node labeled with a non-terminal of the same category as their root node. We have added lexical trees (δ), which are initial trees corresponding to arguments. Their insertion in preterminal nodes of elementary trees, which serves as predicates, is obligatory.

Sentences of the language of a TAG are derived from the composition of an initial tree and any number of auxiliary trees by an operation called 'adjunction'. Adjunction inserts an auxiliary tree at one of the corresponding nodes of an elementary or a derived tree. Recursion is provided by the structure of the auxiliary trees which can adjoin into themselves. Adjunction allows the insertion of a complete structure at an interior node of another complete structure. It appears to be a natural way of handling adverbs and modifiers in natural language. Three constraints can be associated to any node of an elementary tree : null adjunction (NA), obligatory adjunction (OA), and selective adjunction (SA). Because of the formal properties of adjunction, the formalism is more powerful than Context-Free Grammar, but only mildly so /Joshi 85/. Most of its linguistic properties come from the fact that it factors recursion from local dependencies. We are thus able to localize all dependencies such as subcategorization, agreement, and filler-gap relations. Because trees, and not categories, are considered as the units of the grammar, TAGs have a broader domain of locality than usual phrase structure rules.

We have added substitution to the formalism, essentially for descriptive purposes. Although adjunction is more powerful than substitution, and could be used to simulate it, it seems more natural to have substitution itself for lexical insertion and for constructions in which the extra power of adjunction is not needed (section 2). We define a restrictive use of substitution: it inserts an initial tree (or a tree derived from an initial tree), or a lexical tree, into an elementary tree. Substitution is always obligatory and only one constraint, selectional substitution, is defined. This improves the descriptive power of the formalism without changing its generative capacity.

Features structures can be associated with each node of an elementary tree /Vijay-Shanker 87/. They permit the dynamic assignment of constraints. Features are also used for constraining the lexical insertion of items such as prepositions in verbal complements, complementizer of sentential complement or determiner in NP.

Our grammar currently covers the major basic and derived constructions, such as wh-question, relativization or cleft-extraction. We are also able to handle neutral and reciprocal verbs, middle and locative alternations , as well as argument reordering such as scrambling or heavy-NP shift. We refer the reader to /Abeillé 88b/ for a more complete presentation of the grammar. In this paper, we focus on some constructions which are of particular linguistic significance.

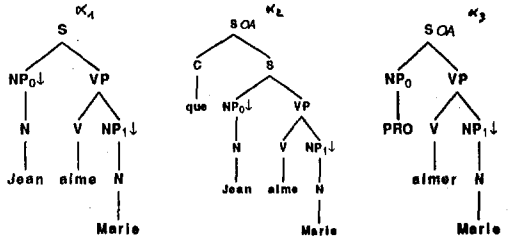
*Visiting from University of Paris VII. This work was partially supported by a J. W. Zellidja grant, and also by ARO grant DAA29-84-9-007, DARPA grant N0014-85-K0018, NSF grants MCS-82-191169 and DCR-84-10413 to the University of Pennsylvania. Thanks are due to Aravind Joshi, Anthony Kroch, and Yves Schabes.

1 Elementary trees and minimal linguistic structures

Our framework is that of a lexicon-grammar /Gross 75/ and /Gross 81/. We view all basic structures as being produced by a lexical item in the lexicon. We adopt the notation of /Gross 75/ and /Boons, Guillet, Leclère 76/. In this framework, as in a TAG, the linguistic unit is the sentence. We define 40 basic structures for French : 12 for verbs taking NP essential complements, and 28 for verbs taking sentential complements.

1.1 Elementary trees for basic constructions

Each of the first 12 structures are represented in the grammar by three initial trees corresponding to declarative sentences, complement clauses, and infinitive clauses. Corresponding to N_0 V N_1 , we have [1]:



NP's are substituted at the proper nodes in the trees. The structures $\alpha 2$ and $\alpha 3$, which would otherwise yield incomplete sentences, bear an obligatory adjunction constraint on their root-nodes. We have to differentiate trees with infinitive from trees with tensed verb, because lexical insertion is defined on already inflected items, and because French does not allow lexical subject in infinitive clauses. We thus state this constraint as a basic structure of the grammar : in $\alpha 3$, the subject has to be non lexical (PRO).

A verb is thus defined by its syntactic argument structure, and the corresponding set of trees are associated with it. We refer to a given argument structure as a tree-family. The optionality of a given argument and the lexical value of the preposition (for verbs taking prepositional complements) are noted as part of the argument structure. A verb with more than one possible argument structure will be duplicated. /Gross 81/ shows that French verbs have no more than three essential arguments, including the subject.[2]

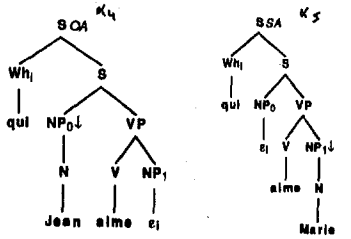
Having such trees associated with the lexical items, instead of a standard argument structure in the form of a list (or of a feature) and rules for sentence formation, provides us with an extended domain of locality that has interesting linguistic consequences. We do not manipulate basic categories, but tree-structures corresponding to minimal sentences (for a verb, or a predicative noun) or complete constituents (NP, for a non-predicative noun, AP, for a modifying adjective). We are thus able to state cross-level dependencies often overlooked in grammars, because they can only be defined on the sentence as a basic unit. For example, the value of the determiner of the subject may depend on the verb, as shown in 1-2, but it also depends on the presence of a verbal complement in 3 ; the adjunction of the right adjective on the nominal complement depends on the lexical value of the verb in 4-5 :

*1) * Ce mot rime.*
2) Ces mots riment.
3) Ce mot rime avec "banane".
*4)*Jean mange un futur gâteau. /Gross 81/*
5) Jean prépare un futur gâteau.[3]

Lexical insertion, or adjunction of adverbs or modifiers such as relative clauses, depend on each element of the elementary tree, and not on just the immediately dominating node. They are difficult to capture by CFG rules such as S→NP VP, or VP→V NP.

1.2 Elementary Trees for Derived Constructions

In a TAG, the standard derived constructions are represented as elementary trees of the grammar. They are part of the tree family associated with a verb. Their properties are those of elementary trees, which must be complete structures and have their gaps bounded in the same tree they appear in, plus the properties given by adjunction and substitution, respectively. If one considers the principles that are used for designing such families, these principles will correspond to syntactic rules, or 'transformations', in derivation-based theories of grammar. Wh-question gives rise to the corresponding wh-elementary trees for each of the arguments of an elementary tree. For the initial tree $\alpha 1$, corresponding to the structure N_0 V N_1, we have for example :



The different local constraints account for the asymetry between subject and object movement. $\alpha 5$ can be an autonomous sentence, whereas $\alpha 4$ is only an indirect question, and must have an auxiliary tree such as *Je sais S* adjoined to it. Relative clauses are represented as auxiliary trees rooted in NP which can then adjoin to the NP node they modify. Each elementary tree, corresponding to a declarative sentence, has thus corresponding auxiliary trees rooted in NP. Cleft-extraction is also represented by elementary trees. To say that a tree with a wh-element, or a relative pronoun, must be an elementary tree, derived from another elementary tree, provides us with strong predictions: wh-movement is forced to apply only to elements present in an elementary tree, that is to arguments of our basic linguistic structures, and not to adjuncts.

[1] For simplification, we do not put all the adjunction constraints these trees bear at their different nodes.↓ marks substitution.

[2] Leaving apart such examples as *Jean parie 100 F à Marie que Pierre viendra*, which can undergo some kind of reanalysis. It is not always easy to distinguish essential complements from adjuncts although our formalism requires a clear-cut distinction.

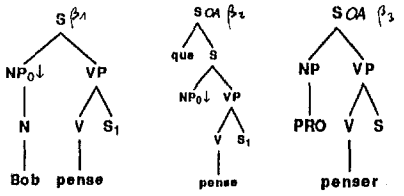
[3] *This word rhymes. These words rhyme. This word rhymes with 'banana'. *Jean is eating a future cake. Jean is making a future cake.

Lexically dependent derivations comprise of middle, ergative, passive, or locative alternation. They are represented as features associated with the proper verb, and correspond to sets of trees to be added to the tree family of the verb. One should notice that the verbal item to be marked is in fact a pair (lexical entry, argument structure). For example, *regarder* has, at least, four argument structures, that is to say four entries :
a) NP_0 *regarde* NP_1
b) NP_0 *regarde* NP_1 *(V-inf W)*
c) *(*NP_0 *+* S_0*) regarde* NP_1
d) NP_0 *regarde que P (subj)*
Only *regarder*(a) has a passive.

For more surface reordering, we have the possibility of defining linear-precedence rules associated either with a tree-family, or with a specific tree, as described in /Joshi 87/.

2 The treatment of complement clauses

The representation of a verb taking a sentential argument can be viewed as the composition of two sentential structures. The standard way of composing two structures in a TAG is to have one adjoined to the other. Complement clauses can thus be represented as elementary trees, with 'matrix' sentences being auxiliary trees adjoined to them, or vice versa. Following /Kroch and Joshi 85/ we prefer the former in order to account for wh-movement out of a complement clause, and to have unbounded dependencies falling out of the formalism. N_0 V S_1, for example, is represented by :



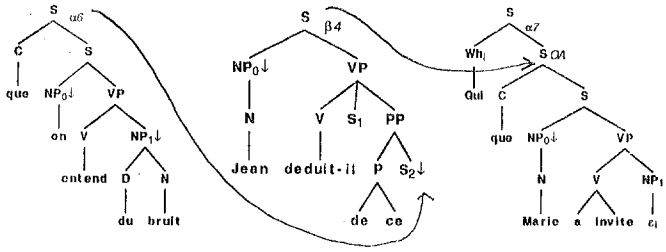
$\beta 1$ is adjoined to $\alpha 3$ to produce:
6) Bob pense que Jean aime Marie[4].
$\beta 2$ and $\beta 3$ are cases of recursive adjunction; 7 is derived from $\beta_1 \rightarrow \beta_2 \rightarrow \beta_2 \rightarrow \alpha_3$:
7) Bob pense que Paul pense que Max pense que Jean aime Marie[5]
The wh-element and the corresponding gap are always in the same basic structure. Unbounded dependencies, which have always been a problem for generative grammar, are thus represented in a straightforward way /Kroch, Joshi 85/ and /Kroch 86/ : adjunction is not limited and does not destroy the gap-filler relations stated in the initial trees. For example:
8) Qui$_i$ *penses-tu que Marie aime* e_i *?*[6]
is derived from *Qui*$_i$ *que Marie aime* e_i *?*, which is one of the Wh-trees corresponding to the initial tree : *que Marie aime Jean*, and *penses-tu* is adjoined to it. The Wh-island constraint is no longer a constraint on movement, but becomes a constraint on the structure of the elementary trees of the grammar. No elementary tree with two wh-elements is defined, and there is no means to derive 9 because there is no elementary tree corresponding to 10:
*9) *Qui*$_i$ *te demandes-tu comment Jean a rencontré* e_i*?*
*10) *Qui*$_i$ *comment Jean a rencontré* e_i *?* [7]

This simple account fails short in the case of verb taking two sentential arguments, such as *Jean préfère perdre Marie à perdre son âme*, because an auxiliary tree is constrained to have exactly one foot-node, and cannot adjoin to two initial trees at the same time. We use for this purpose substitution as an alternative operation. It replaces the leaf node of an elementary tree with an initial, or a lexical, tree (or a tree derived from an initial tree), provided it has a root-node of the same category as that of the leaf-node of the elementary tree.

Let us compare the linguistic properties derived from substitution and adjunction respectively. Substitution represents embeddment as the insertion of a complement clause at a leaf node of the matrix clause. Adjunction views it as the insertion of a matrix clause at any node of a complement clause. Constraints on the derivation are put in the matrix clause, when using substitution, and in the complement clause when using adjunction. Complement clause which undergo wh-movement must be composed with their matrix clause by adjunction, because the matrix clause has to be inserted at an interior node (between the Wh-element and the complementizer). If one uses substitution, on the other hand, insertion at an interior node will be blocked, and wh-movement out of the complement clause will be ruled out. Both operations are therefore complementary; in order to know whether to use one or the other, one has to ask whether wh-movement out of the embedded clause is possible or not.

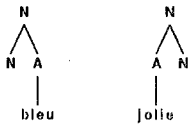
In the case of verbs taking both a sentential subject and a sentential object, we use substitution to represent the subject clause. This makes the well-known sentential-subject island constraint fall out from the formalism. We generate for example 11 and rule out 12:
11) Que Marie aille en Grèce ennuie Jean
*12) *Où*$_i$ *que Marie aille* e_i *ennuie-t-il Jean ?* [8]
The verb *ennuyer* is associated with the argument structure S_0 V NP_1, which is represented as an initial tree[9]. For verbs taking two sentential complements , wh-movement is normally allowed only out of one of the S-complements, usually the direct one.
13) Jean déduit que Marie a fait venir Bob de ce qu'on entend du bruit. [10]
14) Qui$_i$ *Jean déduit-il que Marie a fait venir* e_i *de ce qu'on entend du bruit ?*
*15) * Que*$_i$ *Jean déduit-il que Marie à fait venir Bob de ce qu'on entend* e_i *?*
Using adjunction for the clause subject to extraction and substitution for the other one rightly predicts the ungrammaticality of 15).

[4] Bob thinks that Jean loves Mary
[5] Bob thinks that Paul thinks that Max thinks that John loves Mary
[6] Who do you think that Marie loves ?
[7] *Who$_i$ do you wonder how Jean met e_i ?
[8] That Mary is going to Grece bothers Jean.
[9] To account for the constraint in its full generality we substitute sentential complements even in structures with no other sentential argument.
[10] John deduces that Mary invited Bob from hearing noise.



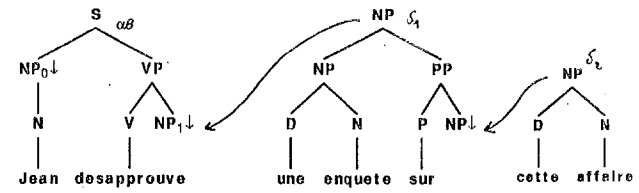
3 The structure of NP : support verb constructions

Modifiers of NP are treated like adjuncts in respect to sentential structures. Adjectives, for example, are represented as auxiliary trees rooted in N, and they adjoin to the node they modify, either before or after the noun :
16) Jean voit un camion bleu.
17) Jean voit une jolie femme [11]



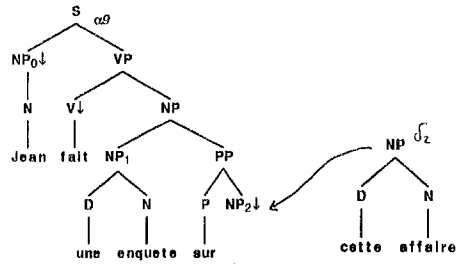
16 and 17 are derived respectively from *Jean voit un camion*, and *Jean voit une femme*. Adjectives produce then two types of structures, one for their modifying nouns, and one for their being arguments of a sentence structure, such as NP_0 V NP_1 A :
18) Jean trouve Marie jolie [12]
They are listed twice in the lexicon, except for so-called relational adjectives, which can only be modifiers :
19) C'est une décision ministérielle
*20) *Cette décision est ministérielle.*[13]
Prepositional phrases modifying NP receive the same treatment, and *Jean voit une femme sans fard*[14] is derived from the adjunction of *sans fard* to *Jean voit une femme*.

Complements of nouns can be either prepositional phrases or sentential complements. They can be viewed as a node in the lexical tree yielded by the head-noun (to be substituted at any NP-node in any elementary tree). This is what we do for sentences such as :
21) Jean désapprouve une enquête sur cette affaire [15].
The PP can only be moved together with the head noun it modifies, and extraction is ruled out for it. Because *cette affaire* is an NP substituted in it, extraction is blocked in the correct way.



The derived constructions, such as wh-movement or cleft-extraction, are defined on the nodes present in the elementary tree. They are thus defined only for NP_1 *enquête*, with or without its complement since the complement is optional, but not on the PP *sur cette affaire*. We thus rule out:
*22) *Sur quoi Jean désapprouve-t-il une enquête ?*
*23) *C'est sur cette affaire que Jean désapprouve une enquête.*

But sentences can be found which are of the same surface structure as 21) but in which the PP exhibit different syntactic properties: it seems to have properties of a nominal and of a verbal complement as well:
24) Jean fait une enquête sur cette affaire. [16]
25) C'est une enquête sur cette affaire que Jean fait.
26) Sur quoi Jean fait-il une enquête ?
27) C'est sur cette affaire que Jean fait une enquête.
These constructions have been called 'support verb' sentences by /Gross 81/, because the verb gives only person and tense marking to the sentence (with optionally some aspectual variation). The noun is the predicative head of the sentence and subcategorizes the subject. /Gross 76/ proposed to have two basic structures associated with these constructions, although they are not ambiguous, and they are problems for most formalisms /Abeillé 88a/. However, they can be represented in a TAG in a natural way with only one basic structure. We consider the PP-node corresponding to *sur cette affaire* as belonging to the initial tree, which makes it an argument of the sentence as any verbal complement. But it is as the same time dominated by the noun *enquête*, and this accounts for its properties as nominal complement.



The difference between 22-23 and 26-27 comes from the fact that wh-movement and cleft-extraction are defined only on the arguments (nodes) of elementary structures rooted in S. In $\alpha 2$ both NP-1 and the PP are available for movement. We are thus able to handle, in the grammar, differences in syntactic properties concerning sentences which are exactly of the same string : (NP VP (NP (PP))). The resulting trees are the same, but one is an initial tree, while the other one is derived .

We also find support verb constructions with nouns taking sentential complements of NP, and we find pairs similar to 21-24 :
28)Jean a le projet d'aller à New-York
29)Jean critique le projet d'aller à New-York.[17]
In 28, the S-node corresponding to the sentential complement of NP is part of the elementary tree, and the string *Jean a le projet de S* is represented as an auxiliary tree. In 29, there is only one NP-node as direct complement of

[11] Jean sees a blue truck. Jean sees a pretty woman.
[12] Jean finds Mary pretty.
[13] This is a departmental decision. *This decision is departmental
[14] Jean sees a woman without make-up
[15] Jean disapproves of an inquiry into this affair.
[16] Jean makes an inquiry into this affair.
[17] Jean has a plan to go to New York/John opposes a plan to go to New York

critique, and the complex NP is treated as a lexical tree, the sentential complement being substituted in it, before insertion in the complete sentence. Thus, extraction is made possible for 28 and not in 29:
30) Où Jean a-t-il le projet d'aller ?
*31) *Où Jean critique-t-il le projet d'aller ?*
To represent support verb constructions with sentential complements as auxiliary trees accounts for unbounded dependencies:
32) Où$_i$ as-tu l'impression que Jean nous a donné l'idée de faire la proposition ... d'aller e_i ? [18]

We consider all nouns taking complements as having corresponding support verbs that they subcategorize. They thus yield a tree family, just as verbs, which comprises of the trees for the support verb construction, and of the complex NP lexical tree as well, which correspond to non-light verb constructions. So, a predicative noun will not be listed twice. Verbs, on the other hand, will be listed twice, as predicate for their 'plain' use, and as arguments for their support verb use, except for a few verbs which appear to be always support verbs : *pratiquer, perpétrer* or *commettre*.

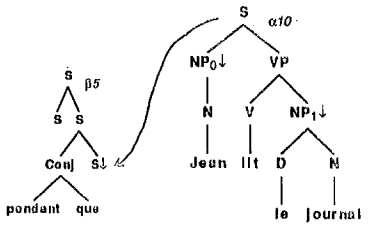
Such a representation can be extended to Verb-Adj-PP constructions, such as:
33) Jean est content de son nouveau chapeau.
34) Jean est content que tout le monde le regarde.[19]
They are considered as S-initial trees yielded by the predicative adjective, and the node for complement, out of which extraction is possible, is present in it. Adjectives taking complements produce then three kinds of tree structure (sentential, attributive, modifying).

We thus extend the set of elementary trees of our grammar to the support verb constructions. They are projections of the noun, or the adjective, in the lexicon, and add 40 basic structures in our grammar.

4 The adjunction of adverbs

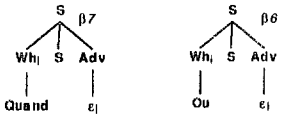
Adverbs can be :
- 'lexical' adverbs : *souvent, rarement*
- PP : *à huit heures*
- NP : *ce jour-là*
- subordinate clauses : *pendant que Jean lit le journal.*[20]

lexical adverbs, PP introduced by prepositions, and subordinate clauses introduced by conjunctions, are represented by the proper auxiliary tree(s) in the lexicon. The prepositional adverbs are listed under the value of their preposition; the bare-NP adverbs under that of their noun, and are considered cases of compound adverbs (see section 5). The subordinating conjunctions are represented as auxiliary trees rooted in S, in which sentential trees (derived or initial) are substituted :



The use of substitution, which forces the insertion of a sentential structure to take place at its root-node, predicts that extraction is ruled out out of an adjunct:
35) Marie regarde la télé pendant que Jean lit le journal.
*36) *Qu$_i$ 'est ce que Marie regarde la télé pendant que Jean lit e_i ?*[21]

Adverbials are represented as auxiliary trees usually rooted in S or in VP. Leaving aside the case of negation, which is a discontinuous constituent, corresponding to a tree rooted in V (because of the word-order), we consider most the adverbs to be rooted in S, in order to have a correspondence with such Wh-trees as $\beta 6$ and $\beta 7$, which have to be rooted in S :



Although the formalism rules out extraction out of adjuncts /Kroch 86/, it does not rule out wh-movement of the adverbial as a whole. It further predicts that only S-rooted adverbials give rise to wh-question:
37) Jean a déploré la destruction de Beirouth Est le 4 Juin.[22]
37 is analyzed as being ambiguous, between an S- and an NP- attachment of the adverbial. But the fronted *Quand Jean a-t-il déploré la destruction de Beirouth Est ?* is correctly disambiguated, because *quand* can only be adjoined to S.

The various positions of an adverb in a string, with the same attachment, is handled by linear precedence rules associated with the tree-structure it adjoins into /Joshi 87/.

For adverbs which are obligatory in a sentence, such as *Jean va bien.* [23], there are two possibilities: either to put an obligatory adjunction constraint in a structure such as *Jean va*, or to treat the adverb as an argument of the elementary tree. We choose the latter, in order to maintain our claim that elementary trees correspond to semantic, as well as syntactic units.

5 The representation of idioms

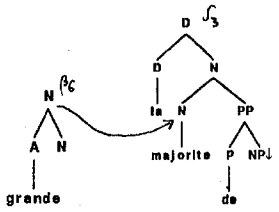
Because in a TAG the linguistic unit is the sentence, not the word, entries comprising of several words can easily be defined. Compound phrases, which can be discontinuous constituents, are assigned a head that is usually either the item of the same category as the whole, or the most significant item. The head produces the subtree corre-

[18] Where$_i$ do you have the impression that Jean gave us the idea to make the suggestion....to go to e_i ?
[19] Jean is happy about his new hat. Jean is happy that everybody admires him
[20] often, seldom/ at eight o'clock/ that day/while Jean is reading the paper
[21] Marie is watching TV while Jean is reading the paper. *What is Mary watching TV while Jean is reading e_i ?
[22] Jean deplored the destruction of East Beirut on June, 4th
[23] Jean is doing fine.

sponding to the compound phrase, which will itself yield a tree-family in the case of a compound predicate (e.g. a compound verb).

The internal structure of sentential idioms is expanded more than that of 'free' sentences. For example, the NP subject is usually noted as an NP-node, open for substitution; if part of it is frozen, the corresponding node (D or N) is directly in the basic tree, and its lexical value is subcategorized by the verb. The heads for sentential idioms are the same as for 'free' sentences. For example, *Jean voit un canard*, which is a free sentence, is a tree of depth 1 : (NP (V NP)), whereas *Jean chasse le canard*, with the meaning of *to hunt*, has a frozen verb-determiner combination, and is represented by a tree of depth 2 : (NP (V (D N))). The verb *chasser*produces also a tree of depth 1, for its occurrence in free sentences, with the meaning of *to chase*. The parser will give two analyses, one corresponding to the idiomatic sense, the other to the literate interpretation.

As for compound categories, we view basic categories as nodes which can be expanded if needed. If it is a simple category, it will be treated as a preterminal, if it is a compound one, its internal structure will be specified. To have the precise internal structure is important in the case of idioms allowing some variations, or insertion. We thus have a unified representation for the complex determiners *la majorité de* and *la grande majorité de* : the adjective *grande* is adjoined to the noun *majorité* as to any N.



Conclusion

Choosing the TAG formalism for parsing French has both computational and linguistic advantages. The linguistic stipulations are minimized and the general organization of the grammar is simplified: all structures are stated in terms of surface structures, and there is a direct matching between the lexical information and the tree structures. The implementation of such a grammar leads to a new parsing strategy developed in /Schabes, Abeillé, Joshi 88/.

We have shown that TAG formalism is suited for building a sizable grammar for a natural language, and furthermore it allows one to state more local dependencies than other formalisms. We show that constraints on extraction out of complement clauses and syntactic properties of support verb constructions are handled in a natural way. We are using our current approach to build a TAG grammar for English along the same lines.

The overall size of the French grammar amounts to 80 basic structures (tree-families), which correspond to simple verbs (12), verbs with sentential complements (28) support verb-noun combinations (20), and support verb-adjective combinations (20). An average tree-family comprises of 15 trees, and the whole size of the grammar is roughly 1200 trees. One should notice that what crucially matters is the number of tree-families, which is closed, if we have been exhaustive. We have not incorporated yet pronominalization and coordination, the two major remaining phenomena. We have added selectional restrictions features to each predicate. We know how to limit the future growth of the grammar: if the derivation we want to add amounts to word-reordering, it is stated by adding a rule to the set of linear precedence rules associated either to the tree-family, or to one of the trees /Joshi 87/. If it is a lexical rule, a feature will be added to the predicative entries. In both cases, the size of the tree-database remains unchanged. If it is a syntactic rule, it adds the proper number of trees to at most each tree-family, so the multiplying factor is 80 in the worst case.

Our grammar has been implemented in an Earley-type parser as defined in /Schabes and Joshi 1988/, and uses a dictionary which comprises of more than 4000 lexical items, that are the most common for French.

References

A. Abeillé (a), "Light Verb Constructions and Extraction out of NP in Tree-Adjoining Grammar", in *Papers of the 24th Regional Meeting of the Chicago Linguistic Society*, 1988.

A. Abeillé (b), "A French Tree Adjoining Grammar", *Technical Report*, Univ. of Pennsylvania, Philadelphia, 1988.

J.P. Boons, A. Guillet, C. Leclère,*La Structure des Phrases Simples en Français : Constructions Intransitives*, Droz, Genève, 1976.

J.P. Boons, A. Guillet, C. Leclère,*La Structure des Phrases Simples en Français : Classes de Constructions Transitives*, Rapport de Recherche du LADL, Univ. Paris VII, 1976.

M. Gross, "Sur quelques groupes nominaux complexes", in *Méthodes en Grammaire Française*, J.C. Chevalier et M. Gross (eds), Klincksieck, 1976.

M. Gross, *Méthodes en Syntaxe*, Paris, Hermann, 1975.

M. Gross, "Les bases empiriques de la notion de prédicat sémantique", *Langages*, n°63, Larousse, Paris, 1981.

M. Gross, "Les limites entre phrases libres, phrases figées et phrases à verbe support, *Langages*, Larousse, Paris, 1988.

A. Joshi, "How much Context-Sensitivity is necessary for characterizing Structural Descriptions: Tree Adjoining Grammars", in D. Dowty et al. eds, *Natural Language Processing: Psycholinguistic, Computational and Theoretical Perspectives*, New-York, Cambridge University Press, 1985.

A. Joshi, L. Levy, M. Takahashi, "Tree Adjunct Grammars", *Journal of the Computer and System Sciences*, 10:1, pp.136-163, 1975.

A. Joshi, "Word-order variation in Natural Language Generation", in *AAAI 87, Sixth National Conference on Artificial Intelligence*, pp 550-555, Seattle, July 1987.

A. Kroch, "Unbounded Dependencies and Subjacency in a Tree Adjoining Grammar", in *The Mathematics of Language*, New-York, Benjamins.1986

A. Kroch, A. Joshi, "Some Aspects of the Linguistic Relevance of Tree Adjoining Grammar", in *Technical Report CIS 85-18*, University of Pennsylvania, 1985.

Y. Schabes, A. Abeillé, A.Joshi, "Parsing Strategies with 'Lexicalized' Grammars: Application to Tree Adjoining Grammars", in Proceedings of the *12th International Conference on Computational Linguistics*, Budapest, 1988.

Y. Schabes, A. Joshi, "An Earley-type Parsing Algorithm for Tree Adjoining Grammars", in *Proceedings ACL'88*, 1988.

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