University of Pennsylvania
ScholarlyCommons

Technical Reports (CIS)

Department of Computer & Information Science

August 1988

A Lexicalized Tree Adjoining Grammar for French: The General Framework

Anne Abeillé
University of Paris

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

A Lexicalized Tree Adjoining Grammar for French: The General Framework

Abstract

We present the first sizable grammar written in the Tree Adjoining Grammar formalism (TAG)[1]. In particular we have used 'lexicalized' TAGs as described in [Schabes, Abeillé and Joshi 1988]. We present the linguistic coverage of our grammar, and explain the linguistic reasons which lead us to choose the particular representations. We have shown that a wide range of linguistic phenomena can be handled within the TAG formalism with lexically specified structures only. We first state the basic structures needed for 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 motivate the choice of the head for the different structures and we contrast the treatment of nominal arguments with that of sentential ones, which is particular to the TAG framework. We also give a detailed analysis of sentential complements, because it has lead us to introduce substitution into the formalism, and because TAG makes interesting predictions in these cases. We discuss the different linguistic phenomena corresponding to adjunction and to substitution respectively. We then move on to 'light verb' constructions, in which extraction freely occurs out of the predicative NP. They are handled in a TAG straightforwardly as opposed to the usual double analysis. We lastly give an overview of the treatment of adjuncts,and suggest a treatment of idioms which make them fall into the same representations as 'free' structures.

Comments

A LEXICALIZED TREE ADJOINING GRAMMAR FOR FRENCH: THE GENERAL FRAMEWORK

Anne Abeillé

MS-CIS-88-64
LINC LAB 125

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

August 1988

A Lexicalized Tree Adjoining Grammar for French : the general framework*

Anne ABEILLE†

Department of Computer and Information Science

University of Pennsylvania

Philadelphia PA 19104-6389 USA

Abstract

We present the first sizable grammar written in the Tree Adjoining Grammar formalism (TAG)[1]. In particular we have used 'lexicalized' TAGs as described in [Schabes,Abeillé and Joshi 1988]. We present the linguistic coverage of our grammar, and explain the linguistic reasons which lead us to choose the particular representations. We have shown that a wide range of linguistic phenomena can be handled within the TAG formalism with lexically specified structures only. We first state the basic structures needed for 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 motivate the choice of the head for the different structures and we contrast the treatment of nominal arguments with that of sentential ones, which is particular to the TAG framework. We also give a detailed analysis of sentential complements, because it has lead us to introduce substitution into the formalism, and because TAG makes interesting predictions in these cases. We discuss the different linguistic phenomena corresponding to adjunction and to substitution respectively. We then move on to 'light verb' constructions, in which extraction freely occurs out of the predicative NP. They are handled in a TAG straightforwardly as opposed to the usual double analysis. We lastly give an overview of the treatment of adjuncts,

---

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] and [Joshi 1985] as a formalism for linguistic description. Its detailed linguistic relevance was first shown by [Kroch and Joshi 85]. In this paper, we make use of a 'lexicalized' Tree Adjoining Grammar, as defined by [Schabes,Abeillé, Joshi 1988].

In a TAG, the basic unit is a phrase or a sentence (viewed as an elementary tree), and not a word. Such elementary structures are semantic as well as syntactic units. The category of a word is in fact the tree-structure associated with it : it is a sentential tree in the case of a simple verb, it is a phrasal tree, in the case of a simple noun. The structures are thus viewed as directly asociated with a lexical item, which is at the frontier.

A TAG's basic component is a finite set of elementary trees, each of which defines a domain of locality, and can be viewed as a minimal linguistic structure. A TAG is comprised of two kinds of elementary trees: initial trees (α), which are complete structures, and auxiliary trees (β), which are minimal recursive structures constrained to have exactly one leaf node labeled with a non-terminal of the same category as their root node. In a lexicalized TAG, both kinds of trees are defined as having at least one terminal on the frontier.

These trees can be characterized by the category of their root-node. An S-type initial tree is a particular case since each valid string has to be derived from it. Arguments can be viewed as X-type initial trees whose insertion into frontier nodes of elementary trees (corresponding to predicates) is obligatory.

Sentences of the language of a TAG are derived from the composition of an S-type initial tree with initial trees which are substituted at all frontier nodes bearing non-terminals, 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 other 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) [2].

---

[2]In the feature structure bond TAG, these constraints are implicit in the specification of the

We have added substitution to the formalism, essentially for descriptive purposes[3]. Although adjunction is more powerful than substitution, and can simulate substitution, 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), into an elementary tree. Substitution is always obligatory and only one constraint, selectional substitution (SS), needs to be defined. This improves the descriptive power of the formalism without changing its generative capacity.

Because of the formal properties of adjunction, the formalism is more powerful than Context-Free Grammar, but only mildly so [Joshi 1985]. It is still a matter of discussion as to whether a language like French, or English, needs a more powerful grammar than a context free one. The enlarged domain of locality TAG's enables the representation of dependencies such as agreement, subcategorization, filler-gap, or those between the subject and the determiner of the first complement[4], in a strictly local manner. Further, [Schabes, Abeillé and Joshi 1988] also proved that strict lexicalization cannot be performed in CFG in a linguistically motivated way.

Most of the linguistic advantages of the formalism come from the fact that it factors recursion from local dependencies. We are thus able to localize 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.

Features structures (fs) have become quite popular for writing linguistic constraints. Feature structures can be associated with each node of an elementary tree in a TAG [Vijay-Shanker 1987]. They are defined as having two components: for each, there is a top fs and a bottom fs. Adjunction itself is defined in terms of unification. When a derivation stops, the top and bottom fs of each node must unify. Constraints on adjoining now come implicitly in terms of the fs and the success or failure of unification. Because of the extended domain of locality, unification over arbitrarily large structures is not necessary, hence it can be shown that the full power of unification is not required [Vijay-Shankar 1987, Vijay-Shankar and Joshi 1988].

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

---

feature structures.

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

[4]In 'light' verb constructions, as described in [Gross 1981]; see also section 1.

3

and reciprocal verbs, middle and locative alternations , as well as argument reordering such as scrambling.

We have defined the elementary trees corresponding to the different possible predicate argument structure in French, including compound categories and other idioms. The formalism rightly handles the different extraction properties of these structures. Applying the analysis of unbounded dependencies of [Joshi and Kroch 1985] to French makes subjacency fall out of the formalism in the same way that it does for English.

Some of the major phenomena we have not covered in this paper are pronominalization and coordination, and also word order variation. Recent TAG work has suggested various approaches, especially for word order variation, but we have not pursued them here. A Grammar for English is currently being written with Kathleen Bishop and Sharon Cote along the same lines, and the expected coverage is roughly the same.

In Appendix 1 we present the lists of the tree families with an expanded description of two of them. In Appendix 2 we give the notations adopted for writing the lexicon as well as a sample of the lexical entries.

1   Lexicalizing elementary trees

We view all basic structures as being produced by a lexical item in the lexicon, which serves as the head. These structures are combined together with either substitution or adjunction. Tree structures in TAG correspond to linguistically minimal structures : an elementary tree is a complete argument structure in the case of a predicate; it is the maximal projection of a category in the case of an argument or an adjunct. A predicate yields an initial or auxiliary tree, depending on whether or not it takes sentential complements. An argument yields an initial tree, and an adjunct an auxiliary tree rooted in the category of the node it modifies.

A two steps parsing strategy can now be defined as follows [Schabes, Abeillé, Joshi 1988] :

- select the set of trees (structures) corresponding to the different items of the input string.

- put the structures corresponding to arguments into the structures corresponding to predicates, and adjoin, if needed, the auxiliary trees corresponding to adjuncts to what they select, or are selected for.

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

4

to minimal linguistic structure : sentences (for verbs and predicative nouns or adjectives) or phrases (NP for nouns, AP for adjectives, PP for prepositions yielding adverbial phrases). 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 ; in the same way the predicate adjective imposes a plural subject in 4-5, unless it has a realized complement (6) :

*1) * Ce mot rime.*
2) Ces mots riment.
3) Ce mot rime avec "banane".
*4) *Marie est cousine.*
5) Ces filles sont cousines.
6) Marie est cousine de Paul.[5]

These dependencies cannot be captured by CFG rules such as S→NP VP, or VP→V NP, without the use of feature structures. In a TAG, however, they will be handled directly. [Gross 1981] also has some examples of the fact that adjunction of adverbs or modifiers such as relative clauses, may depend on another element of a given structure than the one they actually modify.

We have another argument for wanting to collapse into the lexicon the traditional CFG core of PS rules such as:
S→NP VP
NP→ D N'
VP→ V NP.

Although these are the most common rules the following are also possible :
S→S VP
NP→N
NP→D N' PP

The rule we want for first rewriting S, for example, has to be stated for each verb (depending on whether it takes a sentential subject). This rule can as well be collapsed at the same level as the rule for rewriting VP (which clearly has to be stated for each verb). The rule we want for NP also has to be stated for each noun, because we have to know if this noun takes complements or not.

Since even the most general looking rules have to be stated for each lexical predicate, they can be lexicalized without redundancy.

---

[5]*This word rhymes./ These words rhyme./ This word rhymes with 'banana'.
*Mary is cousin./ These girls are cousins./ Mary is cousin with Paul.

1.1 The notion of head

We have the hypothesis that in a non-ambiguous elementary sentence, the basic tree-structure is produced by one lexical item. The head for a structure is the item that :

1 - is of the same category as the root node of the structure (in X-bar notation)[6].

2 - determines the subcategorization of all other nodes in the structure (esp. selection of the subject).

3 - is always lexically present in the structure (and in all its syntactic derivations.)[7]

We follow [Gazdar and Pullum 1981]'s claim that it is necessary to view the notion of head as a primitive, contrary to GB's usual definition of it as 'the' node in a phrase with the same category but at least one bar-level less than the phrase. [Gazdar and Pullum]'s argument is based on examples such as the compound [N-N] phrase, in which you cannot predict the head from the structure of the rule:
N $\rightarrow$ N N
We want to present further arguments : The string D_1 N_1 P D_2 N_2 , corresponding for example to the rules :
NP $\rightarrow$ D N (PP)
PP $\rightarrow$ P NP,
can be of two kinds :

a) la plupart des gens
b) l'opinion des gens

The resulting NP has the gender and number of N_1 in b, of N_2 in a. Since gender and numbers are considered as features passing through the head, it seems logical to have N_2 as the head in a). Semantically also *la plupart des gens* can be analysed as a modification of *les gens*. Example a) is not isolated, and lots of nouns in N_1 position produce the same kind of phrase, with in general being ambiguous with structure b :

7) la masse des gens sont contents/ ?est contente
8) la majorité des gens sont contents / ?est contente
9) un tas de jouets est/sont sur la table.[8]

Usually, the predicate is the verb in sentential structures, but if one considers the

---

[6]S is a special case of course.

[7]It could be added that if the structure is an S, extraction out of all complements of the head has to be permitted.

[8]the mass of the people are/?is happy
the majority of the people are/is ? happy
a pile of toys are/is on the table.

case of nominalizations such as *Jean a de l'amour pour Marie*[9], or *Jean a l'espoir de partir demain*[10], that involve a predicate nominal and the light verb *avoir*, nouns can also be considered heads of sentences; (see section 4). We also consider predicative adjectives as heads of sentences for copular constructions, as is generally agreed (section 5). [11]

The heads for sentential idioms are the same as for 'free' sentences, the items for which the lexical value is specified by the head might be more numerous, without changing the structure of the value or the procedure of value assignment as a whole. In the case of discontinuous heads, elementary trees are considered as having more than one terminal at their frontier.

The other items of a sentence are, linguistically speaking, arguments or adjuncts. The adjuncts are represented as auxiliary trees bearing the same category at their root node (and their foot node) as the node they modify (and to which they are adjoined). The arguments are generally substituted at the leaves bearing preterminals in the elementary trees produced by the heads. But for some cases of sentential arguments, the argument is represented as an initial tree and the predicate structure as an auxiliary tree which is adjoined to it (see section 3).

We duplicate an entry if it has different argument structures. This seems a sound basis for further semantic distinction of the different meanings of a given predicate. For example, the verb *arriver* clearly has , at least, three structures :

a) *NP arriver prep NP* (locative prep) = to arrive somewhere
b) *NP arriver a faire qqc* (prepositional sentential complement) = to succeed in doing something
c) *S (or NP) arrive (à NP)* = to happen (to someone) [12]

1.2 Subcategorization and domain of locality

It follows from the previous discussion that we do not use subcategorization frames as such: instead of having phrase structure rules in the grammar on one hand and

---

[9]Jean has love for Marie.

[10]Jean has a hope to leave tomorrow

[11]It can be argued that a preposition can be a sentential head also in a copular construction such as : a) 'Jean est contre la peine de mort', to be contrasted with the locative : b) 'La pelle est contre le mur' in which the head will be the verb 'be'[Gunnarson 1986]. Sentence a), but not b), can be abbreviated as 'Jean est contre', and 'etre contre' can be seen as one semantically complex predicate.

[12]We do not thereby claim that there is a perfect match between syntactic and semantic classes, but that serious differences in meanings usually show up in the argument structures (and the selectional restrictions, which are kind of syntactic features).

argument structures in the lexicon on the other hand, we collapse both together and associated a tree structure that is a minimal syntactic structure, and a maximal argument structure to the predicative items.

It falls out from the formalism that subcategorization is satisfied in some locality domain, without the need for a special principle for that. Furthermore, because of the extended domain of locality of TAG (compared to CF rules) we are able to have the subject subcategorized because it can be defined in the same domain of locality as the verb, although it is not a sister of V. A TAG in this case will collapse the two CF rules S →NP VP, and VP→V NP (for example) into one (which is a tree of depth 2).

As in LFG [Bresnan et al. 1982], HPSG [Sag and Pollard 1988] and [Gross 1984]'s Lexicon-Grammar, we assume that the subject is subcategorized by the verb the same way the objects are. But in LFG and HPSG, contrary to TAG, the subject is not in the same domain of locality as the verb, and this prevents subcategorization from being formally defined over one locality domain. [Pollard and Sag 1988] present the following arguments for subject subcategorization in English :

- the verb assigns thematic role to the subject ('please' vs 'like')

- the verb may select certain lexical forms as subjects (the dummy : 'there' and 'it')

- certain verbs semantically compatible with sentential subjects rule them out :
  * That we invested in time made us rich.
  The fact that we invested in time made us rich. (J.Fodor)

- in certain languages, different verbs assign different cases to their subjects (e.g icelandic 'dative' subject [Zaenen 1986]).

We add the two following arguments, one concerning the selection of sentential subjects, and the second one concerning the constraint on plural subjects.

Let us take the two verbs *parler* and *dire*, that are closely related in meaning (to speak/to tell). Only the second one can take a sentential subject :

10) * Aller à la pêche parle à Jean
11) Aller à la pêche dit à Jean.[13]

Adjectives too select for sentential subjects in copular constructions , as shown by :

---

[13]Going fishing appeals to John. It is noticable that 'dire' has this meaning 'to appeal' only with a sentential subject: * La pêche dit à Jean. (* The act of fishing 'says' to Jean)

8

*12) * Aller à l'école est incertain.*
13) Aller à l'école est important.[14]

Some verbs exhibit a further selection on the mode of the verb in the sentential subject, the same way they select for the mode of the verb in sentential complements :

14) Que Marie soit amoureuse a percé/transpiré dans la presse.
*15) * Aimer Bob a percé/transpiré dans la presse.[15]*

The same selection holds for compound verbs :

16) Que Marie soit amoureuse saute aux yeux de/va droit au coeur de Luc
*17) * Aimer Suzanne saute aux yeux de/va droit au coeur de Luc.[16]*

Adjectives in copular constructions exhibit a similar selection of the mode of the sentential subject :
18) Que Jean a tué Marie est vrai/faux.
*19) * Avoir tué Marie est vrai/faux.[17]*

20) Aller à New York est cher/bon marché.
*21) * Que les gens aillent à New York est cher/bon marché. [18]*

Another argument is the constraint on plural subjects. It is well known that certain verbs such as 'entasser', 'réunir' or 'séparer' [19] select only for plural objects (or objects with a collective meaning). They have been gathered in [BGL 1976]'s table 32PL. Verbs such as 'pulluler' or 'abonder'[20] require plural (or collective) subjects; and there is not always an ergative relation between verbs in table 32PL and verbs requiring a plural subject; they thus constitute a class on their own[21].

---

[14]* Going to school is uncertain.
Going to school is important.
[15]That Mary was in love was leaked to the press.
*For Mary to be in love was leaked to the press.
These verbs are part of [Gross 1975]'s table 5 that is characterized by the constructions : S_0 V P N_1 = il V P N_1 S_0
[16]That Mary is in love is clear for/speaks directly to Luc. * To love Marie is clear for/ speaks directly to Luc.
[17]That Jean killed Marie is true/false. * For John to have killed Marie is true/false.
The adjective 'faux' can take an infinitival subject but with a totally different meaning : hypocrit
[18]Going to New York is cheap/expensive. * That people go to New York is cheap/expensive.
[19]to pile up, to gather, to set apart
[20]to swarm, to abound
[21]It can be noted, both for plural objects and plural subjects, that generic terms fulfil the constraint as well :
Le lapin de garenne abonde dans ces landes.

1.3 Arguments and adjuncts

It is not always easy to distinguish essential complements from adjuncts although our formalism requires a clear-cut distinction. This distinction has been discussed for a long time, and although it is always crucially assumed, the set of working criteria for distinguishing between an argument and an adjunct may be sometimes fuzzy. We use the following ones that are commonly agreed upon :

- an argument is subcategorized by the head.

- there is dischargement of the role after insertion of the argument, although adjuncts can be reiterated. [Bresnan 1982]

- word order variation between two arguments (if possible) does not affect the meaning of the sentence , differently from between two adjuncts where scoping might vary [Pollard and Sag 1988] . More generally, adjuncts vary more freely as to at what place they can occur. And there might be a change of meaning (so called 'scope') when change of place occurs.

- some arguments may be optional and some adjuncts may be obligatory, but an obligatory element is more likely to be an argument. When an argument is optional the meaning of the sentence is exactly the same without it; this is not the case with optional adjuncts.

- when an element yields a clitic pronoun, it is an argument.

- when wh-question of an element uses 'qui'/'que'/'quoi', it is an argument.

The question of locative complements, and that of obligatory 'adjuncts' such as manner adverbials in 'Jean va bien' or 'Jean se porte mal' is discussed later.

2 Sentential structures with verbal heads

2.1 Elementary trees for basic constructions

We use [Gross 1975] and [Boons, Guillet, Leclère 1976] for collecting the linguistic data. In their framework, the linguistic unit is the sentence. We thus have 35 basic structures for French : 9 for verbs taking NP arguments, 26 for verbs taking sentential arguments. We give below some examples of the structures with nominal arguments, using a simplified notation:


Les pêcheurs entassaient le maquereau frais au fond du bateau.
(The wild rabbit abunds in these moors./ The fishermen were piling the fresh mackerel in the bottom of the boat.) see [Boons, Guillet, Leclère 1976] for futher details.

1- N_0 V : Jean marche
2- N_0 V N_1 : Jean mange une pomme
3- N_0 V Prep N_1 : Jean part de Paris
4- N_0 V N_1 Prep N_2 : Jean vend une maison à Marie
5- N_0 V Prep N_1 Prep N_2 : Jean parle de son travail à Marie
6- N_0 V N_1 N_2 : On élit Jean président
7- N_0 V Adj : Jean est heureux
8- N_0 V Adj Prep N_1 : Ce départ semble heureux à Marie
9- N_0 V N_1 Adj : Marie rend Jean heureux

The lexical coverage varies for each of these structures : 3000 verbs for (2), 300 for (3), 20 for (5) for example. [Gross 1981].

We now give an idea of the representation of these structures in a TAG. Each of the first 9 structures are represented in the TAG grammar by a set of initial trees Among them are the trees corresponding to declarative sentences, complement clauses, and infinitive clauses. For structure (2) : N_0 V N_1 , we thus have [22]:
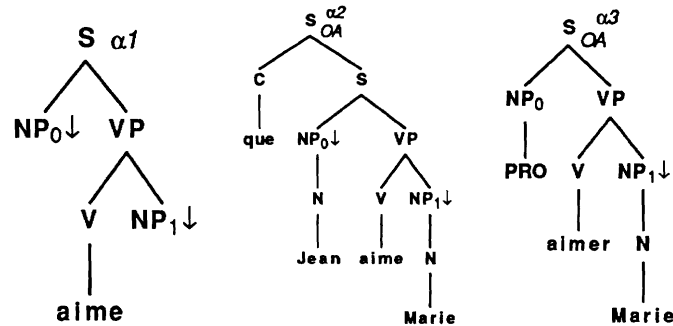
NP's are substituted at the proper nodes in the trees. The structures $\alpha 2$ and $\alpha 3$, which would otherwise yield incomplete sentences, bear an obligatory adjunction constraint on their root-nodes. We have to differentiate trees with an infinitive verb from trees with a tensed verb because French does not allow lexical subject for 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. We do not have subcategorization frames or features as such : the name of the tree family is an indicator that points out to the structures in which the verb can occur.

The optionality of a given argument and the lexical value of the preposition (for

---

[22]For simplification, we do not put all the adjunction constraints these trees bear at their different nodes.↓ is the mark for substitution, and () that for optionality.

verbs taking prepositional complements) are noted as part of the argument structure. A verb with more than one possible argument structure will be duplicated. It can be shown that French verbs have no more than four essential arguments, including the subject, and cases with four are quite rare : in such examples as :

22) Jean parie 100 F à Marie que Pierre viendra. [23]

So [Gross 1981] considered the upper boundary to be three, with some possibility for reanalyis.

2.2   Trees Families for Derived Constructions

In a TAG, as in HPSG, there is only one level of syntactic representation. The standard derived constructions are represented as elementary trees of the grammar. But they do not have to be specified for each lexical item if one views the argument structure associated to a given verb as a set of trees instead of a single tree. Let us refer to this set as a Tree-Family. A tree family consists of elementary trees, which, as such, must be complete structures and have their gaps bounded in the same tree they appear in. The principles that are used for designing such families correspond to syntactic rules or to lexical rules in derivation-based theories of grammar. Let us take the example of wh-question and relativation.

Wh-questions give 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 :
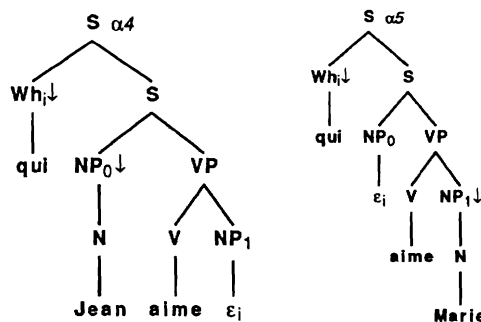
The different local constraints (obligatory adjunction or not) account for the asymmetry between subject and object movement. $\alpha 5$ can be an autonomous sentence, whereas $\alpha 4$ is only an indirect question, and must have an auxiliary tree such as *Je sais S* adjoined to it.

Relative clauses are represented as auxiliary trees rooted in NP which can then

---

[23]Jean bets Marie 100FF that Pierre will come.

adjoin to the NP node they modify. Each elementary tree, corresponding to a declarative sentence, has thus corresponding auxiliary trees rooted in NP.

It should be noted that neither Comp nor relative pronouns are the heads of complement clauses and relative clauses respectively. That accounts for the optionality of the complementizer in many languages other than French [24] complement clauses; see also complement or relative clauses in English :

The man I love is a great person.

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 (linked to another elementary tree in a tree family), provides us with strong predictions: wh-movement is defined to apply only to elements present in an elementary tree, that is to arguments of our basic linguistic structures, and not to adjuncts. The *C'est ...que..* extraction is generated from the elementary trees :
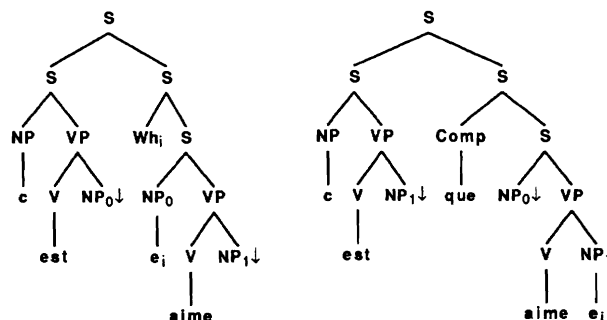
Adjunction is allowed at the root-nodes of α'1 or α"1 (and would be obligatory for the cleft extraction trees corresponding to infinitival (α3) or complement clauses (α2)) , and the grammar produces the grammatical sentences :

23) Je pense que c'est Jean qui aime Marie.[25]

To the cleft extraction on infinitival clause is also associated a complement clause tree, allowing the derivation of 24):

C'est Marie que Jean pense aimer[26] ,

but not to that corresponding to the complement clause, and the following ungrammatical sentence is prevented :

---

[24]But *que* is usually optional in French québécois

[25]I think that it is Jean who loves Marie.

[26]It is Marie that Jean thinks he loves .

*24a) * C'est Marie que je pense que Jean aime.*[27]

The composition of the tree family is independent of the lexical item that will actually yield this family : as soon as the argument structure is specified for this item, it is considered as the head of all the trees of the family.

Along the same lines, features can be associated with a tree structure independently of the lexical value of its head. It is the case for subject-verb agreement for example. It is stated as a feature on the sentential tree. One should notice that in TAG there is no conflict between agreement on the subject and agreement on the object (as arises in pure feature passing along CFG rules in HPSG or GPSG , that need two different principles for that) because both nodes are present in the same domain of locality.

But how do we account for lexically dependent derivations, otherwise known as lexical rules ? Consider that the actual composition of a Tree family has a central core, and some elements which are optional. Lexically dependent derivations consist of middle, ergative, passive, or locative alternation. They are represented as features associated with the proper verb, and correspond to trees that will be considered as part of the tree family of the verb. For example, *casser*[28] will have the feature [erg = +], yielding *NP1 casse*, and its derived constructions. The default value of these features is minus, and they won't be activated for the verb *manger*[29], for example.

Given this formulation, one has actually the choice of seeing the passive, or the middle alternation for example, as a syntactic or lexical derivation. The difference is not in whether it leaves a trace or not, but in whether it is totally regular throughout the lexicon or not. The answer for French, as often discussed (see [Gross 1981], [Boons, Guillet, Leclère 1976], among others), is that passivation is sensitive to the lexical value of the verb. It will be treated then as a feature associated to the tree family for verbs with a direct complement.

One should notice that the lexical item to be marked is in fact a pair (lexical entry, argument structure). For example, *regarder*[30] has, at least, four argument structures, that is to say four entries :

a) NP_0 *regarde* NP_1
b) NP_0 *regarde* NP_1 (S_1 ⟨mode⟩=inf)
c) *(NP_0 + S_0) regarde* NP_1
d) NP_0 *regarde* S_1 (subj)

---

[27]* It is Marie that I think that Jean loves.
[28]to break
[29]to eat
[30]to watch

Only *regarder*(a) has a passive.

For mere 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 1987].

2.3   The treatment of complement clauses

Following [Gross 1975]'s classification, we have the following additional constructions for French :

10- S_0 V : Qu'il neige en Août arrive.

11- S_0 V N_1 : Qu'il neige réjouit Jean.

12- N_0 V S_1 : Jean pense qu'il fait froid.

13- S_0 V S_1 : Qu'il neige prouve qu'il fait froid.

14- N_0 V Prep S_1 : Jean cherche à ce que Marie reste.

15- S_0 V Prep N_1 : Que Marie reste compte pourJean.

16- S_0 V Prep S_1 : Que Jean parte équivaut à ce qu'il reste.

17- N_0 V N_1 S_2 : Jean voit Marie partir.

18- N_0 V S Prep N_2 : Jean dit qu'il fait froid à Marie.

19- N_0 V S_1 Prep S_2 : Jean déduit qu'il fait froid de ce qu'il neige.

20- N_0 V N_1 Prep S_2 : Jean convaint Marie de ce qu'il fait froid.

21- S_0 V S_1 Prep N_2 : Que Marie parte montre qu'elle a froid à Jean.

22- S_0 V N_1 Prep S_2 : Qu'il neige habitue Jean à ce qu'il fasse froid.

23- S_0 V N_1 Prep N_2 : Aller à la chasse donne des boutons à Marie.

24- N_0 V Prep S_1 Prep N_2 : Travailler chez Jean revient à ne rien faire pour Marie.

25- N_0 V Prep N_1 Prep S_2 : Jean compte sur Marie pour que ça aille mieux.

26- S_0 V Prep N_1 Prep N_2 : Que Jean soit venu équivaut à une tuile pour Marie.

27- S_0 V Prep N_1 Prep S_2 : Qu'il neige suffit à Jean pour qu'il achète des skis.

28- S_0 V Prep S_1 Prep N_2 : Aller à New-York a abouti à devenir journaliste pour Marie.

29- N_0 V Prep S_1 Prep S_2 : Jean s'aperçoit de ce qu'il fait froid à ce qu'il neige.

30- S_0 V S_1 Prep S_2 : Qu'il neige ajoute qu'il y ait des morts à ce qu'il y a des sans abri.

31- S_0 V Prep S_1 Prep S_2 : Aller loin diffère d'aller près en ce qu'il faut une voiture.

32- S_0 V Adj S_1 : Que Marie reste rend possible que Jean soit heureux.

33- S_0 V Adj : Que Marie reste est heureux.

34- N_0 V Adj S_1 : Jean trouve heureux que Marie reste.

35- S_0 V N_1 Adj : Que Marie reste rend Jean heureux.[31]

The representation of a verb taking a sentential argument can be viewed as the composition of two sentential structures. Two operations are defined for composing

---

[31]In 18-31 the order of the complements can usually be switched

structures in TAG, namely adjunction and substitution. Nominal, or prepositional arguments of a verb are substituted in the tree-structure produced by the verb. Why not use substitution for sentential arguments ? We refer the reader to [Kroch and Joshi 85] who originally proposed a representation of sentential complements using adjunction only. This allowed them to treat unbounded dependencies as localized dependencies and have subjacency fall out of the formalism as a constraint on the well formedness of elementary trees.

Because adjunction premits the insertion of a structure at an interior node it is possible to have the wh element (that has to be fronted) and the corresponding gap in the sentential complement as part of the same elementary tree. Filler-gap relations are thus preserved no matter how many intervening clauses there are. Substitution could not be used because it allows the insertion of a structure at a leaf node only.

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 1985] we prefer the former in order to be able to have the same structure (namely an initial tree) when the structure is a matrix or independent clause:
Qui ei aime Marie
for example can be either a direct or an indirect question.
   Verbs taking sentential complements thus yield a set of auxiliary tree :
N_0 V S_1, for example, is represented by :
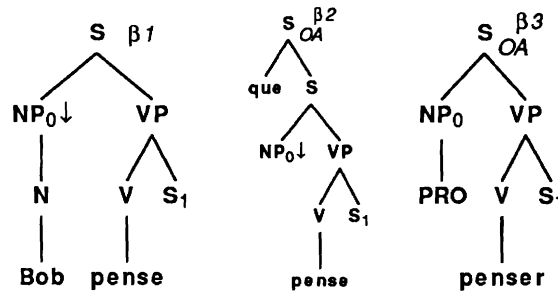
$\beta 1$ is adjoined to 'que Jean aime Marie' ($\alpha 2$) to produce:
25) Bob pense que Jean aime Marie[32].
$\beta 2$ and $\beta 3$ are cases of recursive adjunction; 7 is derived from $\beta_1 \rightarrow \beta_2 \rightarrow \beta_2 \rightarrow \alpha_2$:

26) Bob pense que Paul pense que Max pense que Jean aime Marie[33]

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 1985] and [Kroch 1986]
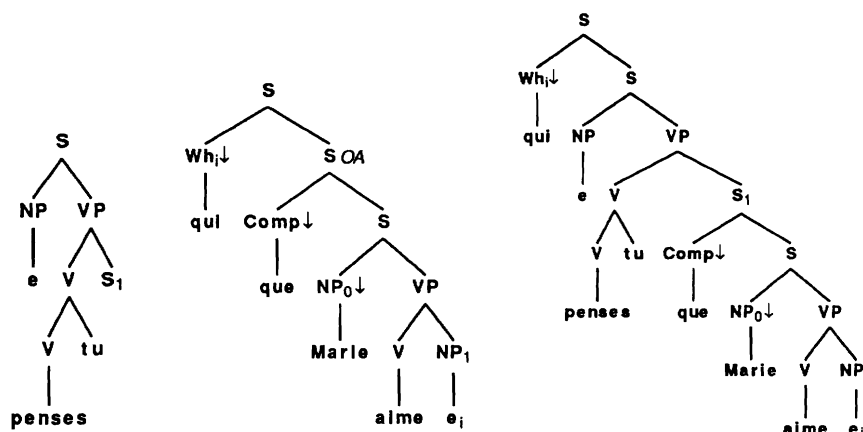
---

[32]Bob thinks that Jean loves Mary
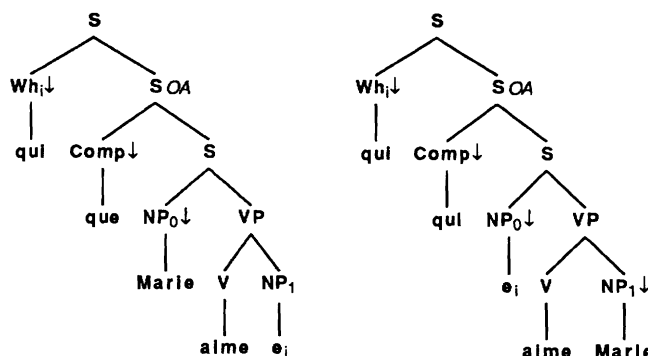[33]Bob thinks that Paul thinks that Max thinks that John loves Mary

: adjunction is not limited and does not destroy the gap-filler relations stated in the initial trees. For example:

27) Qui$_i$ penses-tu que Marie aime e$_i$?[34]

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.

```
       S                      S                             S
      / \                    / \                          /   \
    NP   VP              Wh_i↓   S OA                   Wh_i↓    S
    |   / \                |    /   \                     |    /   \
    e  V   S_1            qui Comp↓  S                    qui NP    VP
       /\                  |    /  \                       |   |   /  \
      V  tu               que NP_0↓ VP                     e   V     S_1
      |                        |   /  \                       /\    /  \
    penses                   Marie V  NP_1                   V  tu Comp↓ S
                                   |    |                    |     |   /  \
                                 aime  e_i                 penses que NP_0↓ VP
                                                                      |   /  \
                                                                    Marie V  NP_1
                                                                          |    |
                                                                        aime  e_i
```

As is shown in [Kroch and Joshi 1985], having the complementizer as part of the complement clause tree permits an easy way for stating the 'that-trace' effect in English. For French, we do not want to say that the second "qui" is a complementizer, and it cannot be a true wh-element because it cannot move to the front of the sentence. Following [Pesetski 1982], we consider this 'qui' as a pronominal element. The copresence of the complementizer node and the wh-element in the same tree, though, allows us to express the wanted restriction, whatever category is assigned to the second 'qui'.

```
           S                               S
         /   \                           /   \
      Wh_i↓   S OA                    Wh_i↓    S OA
        |    /   \                      |    /   \
       qui Comp↓  S                    qui Comp↓  S
            |    /  \                        |   /  \
           que NP_0↓ VP                     qui NP_0↓ VP
                |   /  \                          |   /  \
              Marie V  NP_1                       e_i V  NP_1↓
                    |    |                            |    |
                  aime  e_i                         aime Marie
```

As also pointed out by [Kroch and Joshi 85], this representation also allows one to state the so-called stylistic inversion in French. This subject-verb inversion occurs in subordinate clauses, and , as shown by [Kayne and Pollock 1981], it depends on

---

[34]Who do you think that Marie loves ?

17

the presence of a Wh-element in the same clause at some point of the derivation. In TAG, this phenomenon will be handled directly in the structure of the initial tree. The copresence of the wh-element and the inverted subject is kept no matter how many adjunctions take place, and the constraint is represented straightforwardly : our grammar yields the grammatical 27 and rules out the ungrammatical 28 and 28a:

27) *Que$_i$ pensez-vous qu' a mangé Marie e$_i$?*
28) * *Vous pensez qu' a mangé Marie une pomme.*
28a) * *A mangé Marie une pomme.*

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 in Comp nodes defined, and there is no means to derive 29 because there is no elementary tree corresponding to 30 :

29) * *Qui$_i$ te demandes-tu comment Jean a rencontré e$_i$?*
30) * *Qui$_i$ comment Jean a rencontré e$_i$?* [35]

One should notice that extraction out of infinitive complement clauses, which is usually considered as a problem in French, receives a straightforward treatment. The basic facts are the following :

- Wh-question of the subject is allowed, although there is no corresponding declarative sentence :
31) *Qui penses-tu avoir aimé Marie ?*
32) * *Tu penses Jean avoir aimé Marie.*
   - Relativation of the subject is allowed in the same way, but not indirect questions :
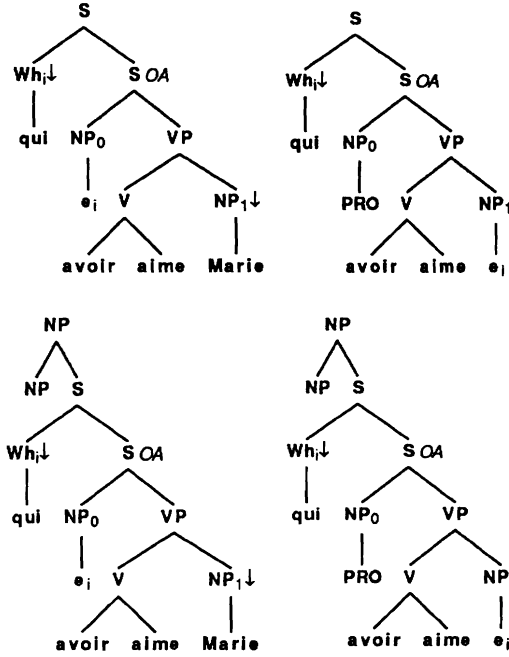33) *Jean, que Max pensait avoir aimé Marie, s'est révélé un fieffé menteur*
34) **Tu sais qui avoir aimé Marie.*
   - No other Wh-questions are allowed with an NP-subject :
35) **Qui penses-tu Jean avoir aimé ?*

In TAG , we have only one initial tree *PRO aimer Marie* for declarative infinitive clauses, and * *Tu penses Jean avoir aiméMarie* will never be generated. The infinitive Wh-trees are derived from this infinitive tree. They must correspond to the arguments of the initial structure, we thus have :

---

[35] * Who$_i$ do you wonder how Jean met e$_i$?

18

S
Wh$_i$↓   S OA
qui NP$_0$ VP
        e$_i$ V   NP$_1$↓
           avoir aime Marie

S
Wh$_i$↓   S OA
qui NP$_0$ VP
       PRO   V   NP$_1$
          avoir aime e$_i$

NP
NP S
Wh$_i$↓   S OA
qui NP$_0$ VP
        e$_i$ V   NP$_1$↓
           avoir aime Marie

NP
NP S
Wh$_i$↓   S OA
qui NP$_0$ VP
       PRO   V   NP$_1$
          avoir aime e$_i$

A tree corresponding to qui_i *Jean aimer* e_i will never be generated. and the ungrammaticality of **Qui penses-tu Jean avoir aimé ?* is predicted. On the other hand we correctly predict the grammaticality of :

Qui$_i$ penses-tu avoir aimé e$_i$?

The auxiliary trees corresponding to relative clauses receive the same treatment. Furthermore, the gap on the subject is properly bounded in the αW0, and no coreference with an element of the matrix sentence is required. In *PRO aimer Marie* , however, αW1 obeys the general rules of PRO-binding. *Jean$_i$ veut PRO$_i$ aimer Marie* and *Max voit Jean$_i$ PRO$_i$ aimer Marie* will thus be properly generated.

This simple account falls short in the case of verbs taking two sentential arguments, such as :

36) Jean préfère perdre Marie à perdre son âme.

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 [36].

---

[36]We cannot subsitute auxiliary trees because we would run the risk to end up with ill-formed trees with more than one foot-node.

Let us compare the linguistic properties derived from substitution and adjunction respectively. Substitution represents embedding 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 37 and rule out 38:

37) *Que Marie aille en Grèce ennuie Jean*
38) **Où$_i$ que Marie aille e$_i$ ennuie-t-il Jean ?* [37]

The verb *ennuyer* is associated with the argument structure S_0 V NP_1, which is represented as an initial tree[38]. This reprsentation also accounts for the intermediary status of sentential subjects which are considered as half way between sentential and nominal arguments. [Koester 1986]. As sentential arguments they are initial trees, but they are substituted and not adjoined into as nominal arguments.

For verbs taking two sentential complements , wh-movement is normally allowed only out of one of the S-complements, usually the direct one:

39) *Jean déduit que Marie a invité Bob de ce qu'on entend du bruit.* [39]
40) *Qui$_i$ Jean déduit-il que Marie a invité e$_i$ de ce qu'on entend du bruit ?*
41) ** 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 41.

---

[37]That Mary is going to Greece bothers Jean. * Where does it bother Jean that Marie goes ?

[38]To account for the constraint in its full generality we substitute sentential subjects even in structures with no other sentential argument.

[39]John deduces that Mary invited Bob from hearing noise.
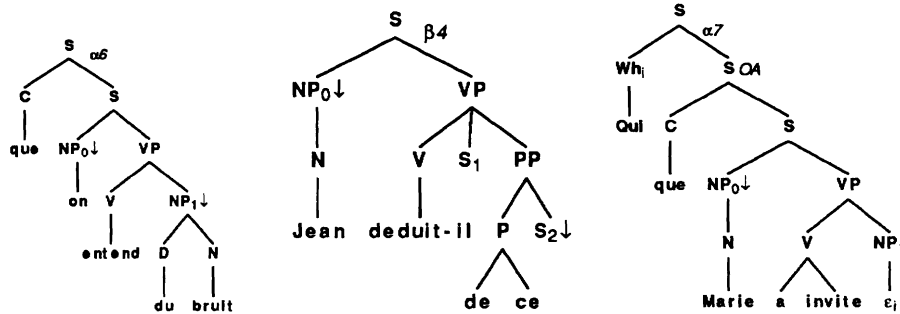
We also have a representation for handling marginal cases where extraction seems to occur freely out of either of the two complement clauses, such as :

42) Où$_i$ Jean préfère-t-il aller e$_i$ plutôt que de se casser une jambe?
43) ? Où$_i$ Jean préfère-t-il se casser une jambe plutôt que d'aller e$_i$? [40]

Two structures must be defined : in 42) the first sentential constituent results from adjunction and the second one from substitution, and vice versa in 43).
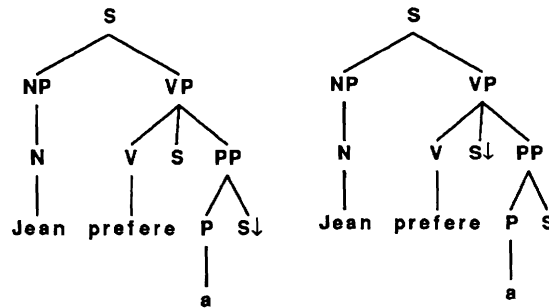
Such a device is linguistically motivated : extraction never occurs in both S-complements at the same time, it is necessary to block it in one of the sentential complement for each structure.

*44) * Où quelle jambe Jean préfère-t-il aller plutôt que de se casser ?*
*45) * Quelle jambe où Jean préfère-t-il se casser plutôt que d'aller ?* [41]

This constraint is properly handled by substitution. The formalism prohibits exactly what has to be avoided; namely concurrent extraction out of both sentential complements at the same time, which would be a case of concurrent adjunctions.

The case of double gaps, that is to say extraction of the same constituent out of both clauses, is to be analysed as parasitic gaps. They are properly handled by the assymetry between adjunction and substitution.

---

[40]Where does Jean prefer going to over breaking his leg ?
? Where does Jean prefer breaking his leg over going to ?

[41]* Where e$_i$ which leg e$_j$ does Jean prefere going to e$_i$ over breaking e$_j$? * Which leg e$_i$ where e$_j$ does Jean prefer breaking e$_i$ over going to e$_j$?

46) Qui Jean préfère-t-il oublier plutôt que détester ?
46a) Qui Jean préfère-t-il oublier plutôt que le/la détester?
46b) *Qui Jean préfère-t-il l'oublier plutôt que détester? [42]

α1 is thus adjoined to a wh-initial tree : [Qui$_i$ Pro oublier e$_i$], and the elementary tree : [Pro détester e$_i$] is substituted at the relevant open node. The elementary tree : [Pro détester e$_i$] is defined as an initial tree that must be substituted as an argument in a structure. The proper constraints on substituting such a tree are stating in terms of features on gaps : it is ruled out unless the main clause has a sentential complement with a gap in the same position:

47) *A qui préfères-tu téléphoner e$_i$ plutôt qu'oublier e$_i$?
48) A qui préfères-tu téléphoner e$_i$ plutôt qu' écrire e$_i$? [43]

In 47 the first gap stands for a PP and the second for an NP (direct object), and the sentence is ill-formed. In 48, the two gaps correspond to the same PP (prep = a).

Substitution is also needed when the verb taking a sentential argument is inside a relative clause, for example [44]:
49) Je connais la femme qui pense que tout le monde la déteste. [45]

Because a relative clause is an auxiliary tree rooted in NP the tree corresponding to [NP qui ei pense S] would be ill formed with an S-footnode for the sentential complement. We thus have to consider that in relative clauses the S-footnode turns into a substitution node. This further rightly predicts that no extraction out of the complement clause can occur and that the relative clause is an island for extraction :

50) C'est l'homme qui pense qu' il est supérieur à Marie.
50a) *[A quelle femme]$_i$ est-ce l'homme qui pense qu'il est supérieur e$_i$? [46]

---

[42]Who does John prefer forgiving to hating ?
Who does John prefer forgiving to hating him/her ?
*Who does John prefer forgiving her/him to hating ?
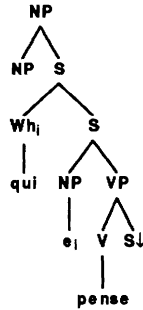[43]* Whom do you prefer giving a call over forgetting ?
Whom do you prefer giving a call over sending a letter ?
[44]We thank Anthony Kroch for pointing this problem out to us
[45]I know the woman who thinks that everyone hates her.
[46]This is the man who thinks that he is superior to Marie.
* Which woman is this the man who thinks he is superior to ?

This also is an argument for having the relative pronoun in the same elementary tree as the relative clause, and the main verb, and not the relative pronoun, as the head of the relative clause.

As a whole, the core-grammar for sentential complement-verb sentences is comprised of more than 150 elementary trees.

3 The structure of NP : light verb constructions

Nouns in general produce initial trees rooted in NP (to be substituted at any NP-node in any elementary tree).

3.1 Complements and modifiers of N

Complements of nouns can be either prepositional phrases or sentential complements. They can be viewed as a node in the intial tree yielded by the head-noun. This is what we do for sentences such as :

51) Jean désapprouve une enquête sur cette affaire [47]

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 the NP-tree, it is not an argument of the sentential tree yielded by the verb. No elementary tree corresponding to wh-movement is therefore defined for it, and extraction out of NP is blocked in the correct way.


---
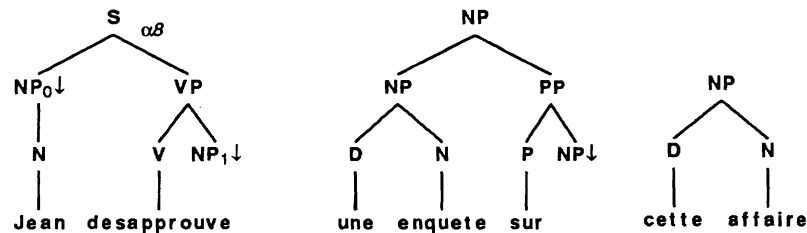
[47]Jean disapproves of an inquiry into this affair.

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:

51a) *Sur quoi Jean désapprouve-t-il une enquête ?
51b) *C'est sur cette affaire que Jean désapprouve une enquête. [48]

Prepositional phrases modifying NP are not nodes present in the NP tree from the beginning. They are auxiliary trees that have to be adjoined to the NP tree.

52) Jean voit une femme sans fard[49]
    is derived from the adjunction of sans fard to Jean voit une femme :
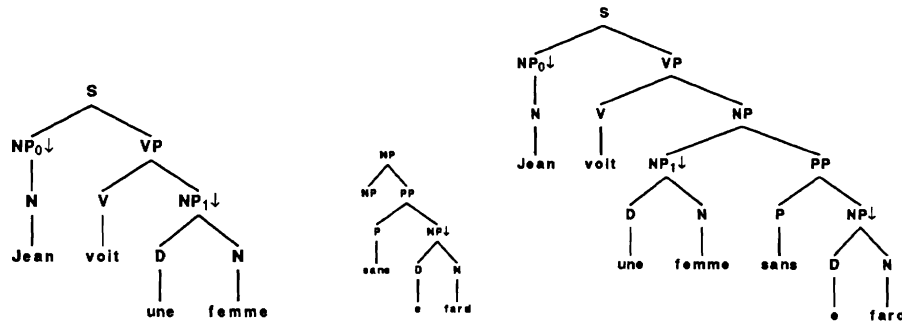
The head of the auxiliary tree for the modifier is the preposition, and one can distinguish between prepositions introducing NP modifiers and those introducing VP modifiers (see section 5).

One should notice that what makes the difference between complements and modifiers of nouns is whether the node for insertion is part of the tree structure of the noun or not; and the operation involved, namely adjunction or substitution, is therefore different. It is not the bar level at which the insertion takes place that matters, and actually the question has been the matter of much controversy.

Extraction out of the modifier is blocked in the right way :

* Sans quoi as-tu vu une femme ?[50]
could only come from the simultaneous adjunction of 'sans quoi' at the top S-node of 'as-tu vu une femme' and of an empty PP (coindexed with 'sans quoi') at the NP node 'une femme'. Such an empty PP would be ill formed by definition of a lexicalized TAG. Extraction of complements out of NP is addressed in the following section that deals

---

[48]* What does Jean disapprove of an inquiry into ? * This is this affair that Jean disapproves of an inquiry into.
[49] Jean sees a woman without make-up
[50]* What did you see a woman without ?

with the privileged case of 'light verb' constructions. For extraction of complements of NP in non light verb sentences, see [Kroch 1988] and [Abeillé 1988(a)] but the question is still under discussion.

3.2 Light verb constructions

As pointed out by [Gross 1976] 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:

53) Jean fait une enquête sur cette affaire.
53a) Sur quoi Jean fait-il une enquête ?
53b) C'est une enquête sur cette affaire que Jean fait.
53c) C'est sur cette affaire que Jean fait une enquête. [51]

These constructions have been called 'support verb' sentences by [Gross 1981], because the verb gives only person and tense marking to the sentence (with optionally some aspectual variation). [52] The noun is the predicative head of the sentence and subcategorizes the subject. [Bach and Horn 1976] for English, and [Gross 1976] for French, proposed to have two basic structures associated with these constructions : NP V (NP (PP)) and NP V (NP) (PP). But these sentences are not ambiguous, and there is no clear evidence to support such a dual analysis. Sofar thes constructions remain a problem for most formalisms [see Abeillé 1988a]. 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.


---
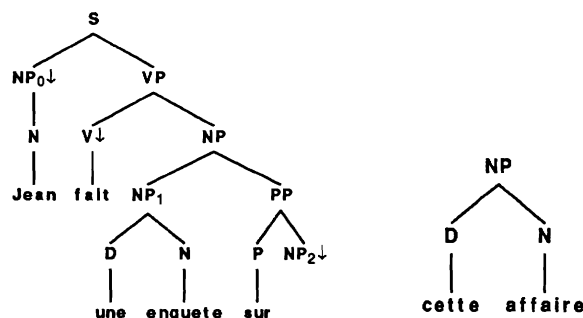
[51] Jean makes an inquiry into this affair.
What does Jean make an inquiry into ?
It is an inquiry into this affair that Jean makes.
It is this affair that Jean makes an inquiry into.
[52] We use for English Jespersen's name 'light verb', Jespersen considered them only as stylistic variation of 'heavy verb ': I hope that ...= I have a hope that....

The difference between 51a-53a and 51b-53b comes from the fact that wh-movement and cleft-extraction are defined only on the arguments (nodes) of elementary structures rooted in S :
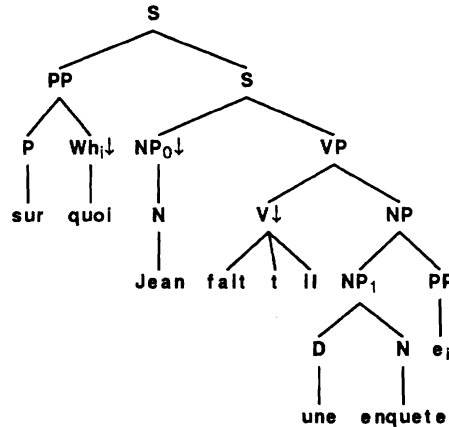
In $\alpha 2$ both NP-1 and the PP are available for movement. We are thus able to handle, in the grammar, differences in syntactic properties concerning sentences which are exactly of the same string : (NP VP (NP (PP))). The resulting trees are the same, but one is an initial tree, while the other one is derived .

Heavy-NP shift for (4), as opposed to (6), is handled , because it is stated between two arguments copresent in the same elementary tree.

The difference between the relativization of simple verb and light verb constructions, for example, follows from the only constraint that modifiers cannot be adjoined to empty nodes. We have two relative clauses for (51) but only one for (53) :

51d) Je connais l'enquête que Jean fait sur cette affaire.
51e) Je connais l'enquête (sur cette affaire) que Jean fait.
53d) *Je connais l'enquête que Jean désapprouve sur cette affaire.
*53e)* Je connais l'enquête (sur cette affaire) que Jean désapprouve. [53]

In (51) the PP is an argument of the initial tree. It can be treated as arguments and left behind when Wh-movement applies to N_1 :

---

[53] I know the inquiry that Jean made about this affair
I know the inquiry (about this affair) that Jean made
* I know the inquiry of which Jean disapproves into this affair
I know the inquiry (into this affair) of which Jean disapproves.

26
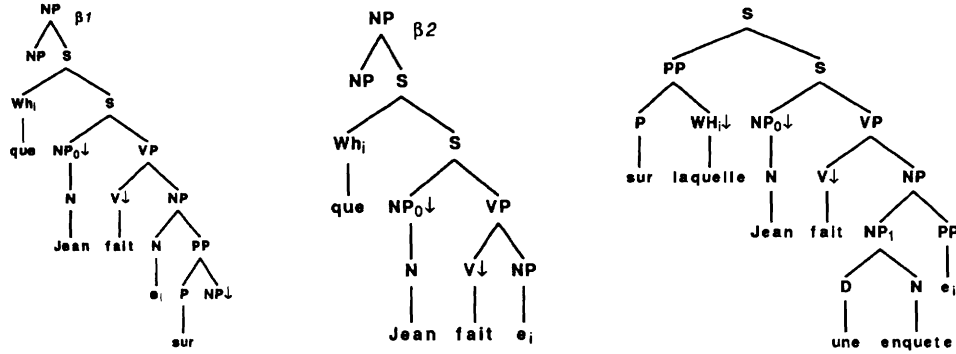
In the basic structures corresponding to NP_0 fait une enquête sur NP, NP is an argument of the initial trees. A relativized tree is thus defined on it , and it can stay behind when relativization applies to 'enquête' . In 53), on the other hand, the PP 'sur cette affaire' is not an argument of the basic structure NP_0 désapprouve NP_1, and it has no way of being inserted in the elementary tree corresponding to the relative clause with a gap on the NP_1 'enquête'.

We also find light verb constructions with nouns taking sentential complements of NP, and we find pairs similar to 51-53:

54) Jean a le projet d'aller à New-York
55) Jean critique le projet d'aller à New-York.[54]

In 54, 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 55, there is only one NP-node as direct complement of *critique*, and the complex NP is treated as an initial tree, the sentential complement being substituted in it, before insertion in the complete sentence. Thus, extraction is made possible for 54 and not in 55:

54a) Où Jean a-t-il le projet d'aller ?
*55a) * Où Jean critique-t-il le projet d'aller ?* [55]

---

[54]Jean has a plan to go to New York/John opposes a plan to go to New York

[55]Where does Jean have a plan to go ? * Where does Jean criticize a plan to go ?
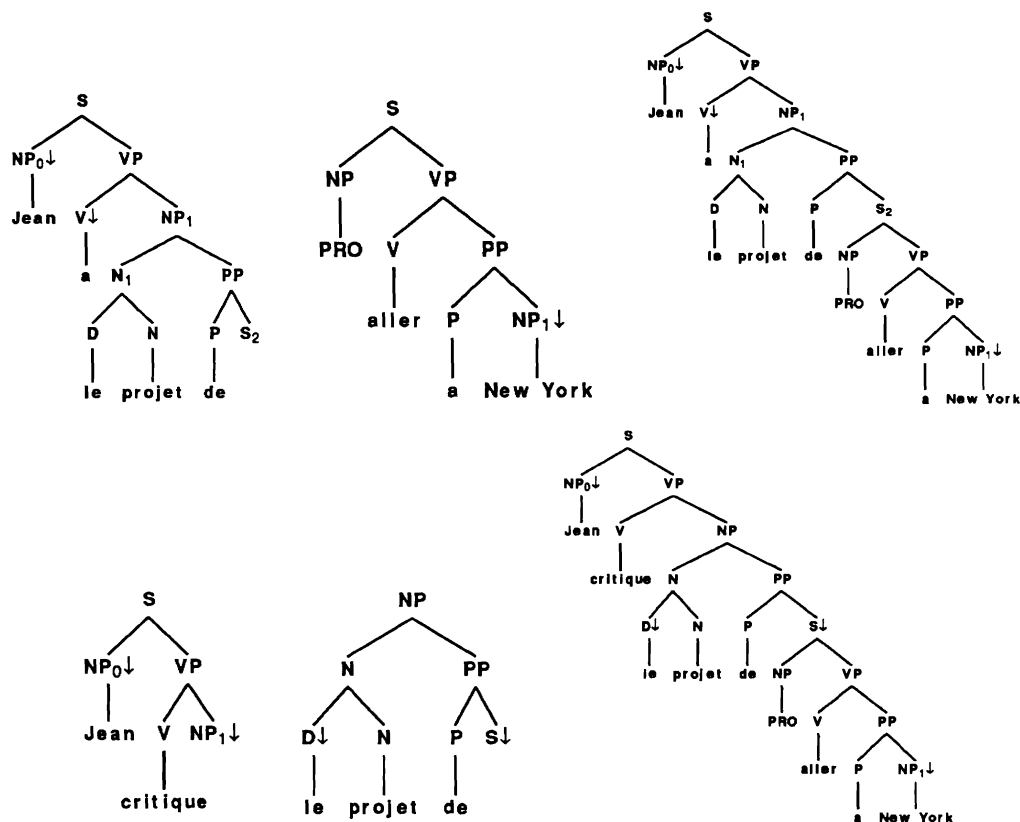
To represent light verb constructions with sentential complements as auxiliary trees accounts for unbounded dependencies:

56) Où$_i$ as-tu l'impression que Jean nous a donné l'idée de faire la proposition ... d'aller e$_i$? [56]

We want to have the predicative noun as the syntactic head of these structures. It has been proposed that S be a projection of a noun (or an adjective) for copular sentences such as 'John is a Republican' [???], and independently for small clauses [Stowell 1981]. We extend this to the light verb constructions on the following grounds :

- subcategorization of the subject by the predicate nominal :
The house/ The explosion gave a cry*
The house/The explosion gave a push
- 'subcategorization', if one might say, of the verb by the predicate nominal :
*John gave/*took a sneeze*
*John *gave/took a snooze.*
- It is noticeable that in light verb construction, as in some idioms, standard

---

[56]Where$_i$ do you have the impression that Jean gave us the idea to make the suggestion....to go to e$_i$?

Wh-movement cannot apply to the predicative noun:
* *Que fait-il ? - Un cadeau a Marie.*

The only transformations which apply are relativization and quel-movement :
Le cadeau qu'il fait .../ Quel cadeau fait-il ?

Pronominalization is accepted only in topicalized, or coordinate, structures :
Ce formidable cadeau, il l'a fait à Marie (pas à Jeanne)

That is to say the item 'cadeau' is always lexically present (M.Gross ; personal communication).

We have several arguments for treating the verb as two entries in the lexicon : one for the 'light ' verb (treated as an argument), and one (at least) for the 'heavy' constructions. First of all, the selectional restrictions are not the same. *prendre* as heavy verb, selects non abstract objects for example, whereas as light verb it will cooccur with *vacances, temps,* or *plaisir* for example.

Another argument is coordination : as pointed out by [Danlos 1985], you cannot coordinate two complements if in one VP the heavy verb is used whereas it is the light verb in the other one :

Avant de sortir, Jean prend son chapeau et Jean prend le temps de le brosser.
Avant de sortir, Jean prend son chapeau et le temps de le brosser.

In certain languages, morphology may differ: In English *have* does not have a passive when it is the predicate of the sentence:

John had a car
A car was had by John.

But it can have one when it is used as a 'light verb' :
Everyone had a good time .
A good time was had by everyone.

We thus extend the set of elementary trees of our grammar to the light verb constructions. All nouns taking a complement produce at least one light verb construction with *être* as light verb. Their NP-type initial tree, for their appearing in non light verb constructions, is considered to be part of the same tree family, so they do not have to be listed twice in the lexicon, unless they have a different meaning :

57) *Jean n'a pas eu la liberté de choisir.*
58) *La liberté de penser doit être protégée partout dans le monde.*
59) *La liberté est un des grands principes de 89.*[57]

---

[57]Jean did not have the freedom to choose.
The fredom of thinking must be protected everywhere.
Liberty was one of the principles of 1789.

The predicative noun *liberté* has the same meaning in the light verb construction 57 and the non light verb one 58. But in 59 the meaning is more general and philosophical, and it will be a separate entry.

Light verb constructions are projections of the noun in the lexicon, and 20 basic structures are thus added in our grammar.

36- N_0 V' [N_1 Prep N_2] : Jean fait une enquête sur cette affaire.
37- N_0 V' [Prep N_1 Prep N_2] : Jean est dans l'attente d'un important événement.
38- N_0 V' [N_1 Prep N_2] Prep N_3 : Cette lettre donne de l'amour pour Marie à Jean.
39- N_0 V' [Prep N_1 Prep N_2] Prep N_3 : Jean parle en faveur de Luc à Marie.
40- N_0 V' N_1 [Prep N_2 Prep N_3] : Cet exploit met Jean en admiration devant Marie.
41- N_0 V' [N_1 S_2] : Jean a l'espoir que Marie vienne.
42- S_0 V' [N_1 S_2] : Qu'il rougisse est une preuve que Pierre ment.
43- N_0 V' [N_1 Prep S_2] : Jean a tendance à exagérer.
44- N_0 V' [Prep N_1 S_2] : Jean est d'avis que Marie vienne.
45- S_0 V' [N_1 Prep N_2] : Qu'il neige donne envie de vacances.
46- S_0 V' [N_1 Prep S_2] Qu'il manque une place est signe de ce que Pierre part.
47- S_0 V' [Prep N_1 Prep N_2] Prep N_3 Qu'il aide Jean parle en faveur de Luc à Marie.
48- N_0 V' [Prep N_1 S_2] : Max est sur le point de partir.
49- S_0 V' [Prep N_1 Prep N_2] : Qu'il aide Marie plaide en faveur de Jean.
50- $_0$ V' [Prep N_1 Prep S_2] : Habiter Paris est en passe de devenir à la mode.
51- N_0 V' [N_1 S_2] Prep N_2 : Bob donne l'impression qu'il dort à Luc.
52- S_0 V' [N_1 S_2] Prep N_2 : Que Bob ferme les yeux donne l'impression qu'il dort à Luc.
53- N_0 V' [N_1 Prep S_2] Prep N_3 : Marie donne l'idée de partir à Jean.
54- S_0 V' [N_1 Prep N_2] Prep N_3 : Rencontrer un ami donne le courage de ses opinions à Jean.
56- S_0 V' N_1 [Prep $N_2 N_3$] : Aimer Max met Marie en contradiction avec elle-même

These constructions are characteristic of nouns taking complements (either sentential or nominal ones). But there are a few "intransitive" predicative nouns which will be described as heads of T-structures (because they do not take any NP-complement). For example "sport" will be head of "NP_0 fait du sport" and receive the category : Tn1(V=*faire*). Our main hypothesis, which seems to hold true so far, is that the head of the predicate is always lexically present in any of the constructions it yields. Since a noun is the head of a sentence, standard Wh-movement cannot apply :

* Que fait-il ? - Du tennis
Que fait-il? - Il fait du tennis.

The only transformations which apply are relativization and quel-movement :

Le sport qu'il fait est inoffensif.
Quel sport fait-il ? - Du tennis.

For more details on light verb constructions in French, we refer the reader to [Gross 1976], [Giry-Schneider 1978], [Danlos 1980], [Gross 1981] and [Giry-Schneider 1987]; for more details on a TAG analysis of the phenomena see [Abeille 1988a].

4 Adjectives as predicates, arguments or modifiers

4.1 Predicative adjectives

Adjectives behaving as predicates can be found in copular constructions, among which are constructions analogous to the light verb constructions just discussed. Let us consider the following sentences :

60) Jean est content de son nouveau chapeau.
61) Jean est content d'aller à Paris.
61a) De quoi Jean est-il content ?
62a) Où Jean est-il content d'aller ?[58]

These sentences can be contrasted with :

63) Marie trouve Jean content de son nouveau chapeau.
*63a) * De quoi Marie trouve-t-elle Jean content ?*[59]

We say that 61-62 are 'light verb' constructions, while 63 is not. We use in TAG a representation similar to that of nouns in light verb constructions[60] . The same arguments hold for wanting them as heads of the structures. And we can check that movement of the predicative adjective is not natural, differently from 'plain' adjective as color adjectives for example :

64) Comment est Jean ? - Grand.
65) ? Comment est Jean ? - Content de son travail.*[61]

---

[58]Jean is happy about his new hat.
Jean is happy to go to Paris.
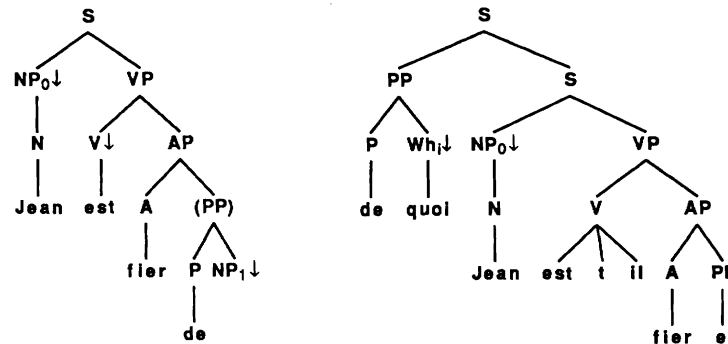What is Jean happy about ?
Where is Jean happy to go to ?
[59]Marie finds Jean happy about his new hat. * What does Mary find Jean happy about ?
[60]For more details on adjectives taking sentential complements see [Picabia 1974].
[61]How is John? - Tall.
?* How is John ? - Happy about his work.

These sentences 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. The light verbs here are : *être, devenir, paraître....*

We thus add the 10 following structures in the grammar :

57- N_0 V' [Adj Prep N_1] : Max est fier de ce resultat.

58- N_0 V' [Adj Prep N_1] Prep N_2 : Jean est content de ses résultats pour Marie.

59- N_0 V' N_1 [Adj Prep N_2] : Cette lettre rend Jean fou de Marie.

60- N_0 V' [Adj S_1] : Jean est content que tout le monde le regarde

61- S_0 V' [Adj Prep N_1] : Tuer son prochain est passible d'emprisonnement à vie.

62- S_0 V' [Adj S_1] : Qu'il fasse beau est susceptible de faire des heureux.

63- N_0 V' [Adj S_1] Prep N_2 : Jean est content que Bob aille mieux pour Marie.

64- S_0 V' [Adj Prep S_1] : Qu'il neige est nécessaire pour pouvoir skier

65- S_0 V' [Adj Prep S_1] Prep N_2 : Aller à N-Y est nécessaire pour pouvoir skier pour Max.

66- S_0 V' N_1 [Adj Prep N_2] : Qu'elle l'ait embrassé sur le front a rendu Jean fou de Marie.

67- N_0 V' N_1 [Adj S_2] : Jean rend Marie heureuse de vivre.

68- S_0 V' N_1 [Adj S_2] :Qu'il fasse beau rend Marie contente d'aller en vacances.

Differently from English, for which the adjective adjoins after the noun when it has a complement but before if it has not (or if it is not realized), in French we can have the same representation for the adjective with and without its optional complement, exactly as for nouns with optional complements ; an AP tree with optional node for its complement.[62]

66) Marie est une femme fière.
66a) Marie est une femme fière d'elle même.
Mary is a proud woman.
Mary is a woman proud of herself
Mary is a proud of herself woman.

What about copular constructions for adjectives that do not take complements ?

---

[62]As for nouns, some adjectives have obligatory complements, example 'être sujet à NP'.

The selection of the subject seems to be done by the adjective as far as sentential subjects are concerned. Color adjectives take only NP subject, whereas adjectives which refer to a state of mind can usually take both :

68) * Aller à la pêche est bleu
69) Aller à la pêche est ennuyeux.

But if we have the adjective as the head of all copular constructions, we cannot account for the wh-movement that it undergoes (when the subject is an NP) without violating the postulate that the head is always lexically present :

70) Jean est joyeux.
70a) Comment est Jean ?[63]

But the contrast with the following sentences can serve as an insight that in fact two different constructions are here at stake :

69a) *Comment est aller à la pêche ?
69b) Comment aller à la pêche est-il ?

This is also supported by the impersonal derivation : $il\ est\ A\ S_0$ that has no equivalent : $*\ il\ est\ A\ NP_0$:

71) Il est ennuyeux d'aller à la pêche.
72) * Il est joyeux un homme.[64]

We define two different structures to 69 and 70. In 70 , the head is the adjective, whereas in 69 the head is the verb. We thus define initial trees for adjectives as arguments, and these trees are also used for sentences such as :

Marie rend Jean joyeux

---
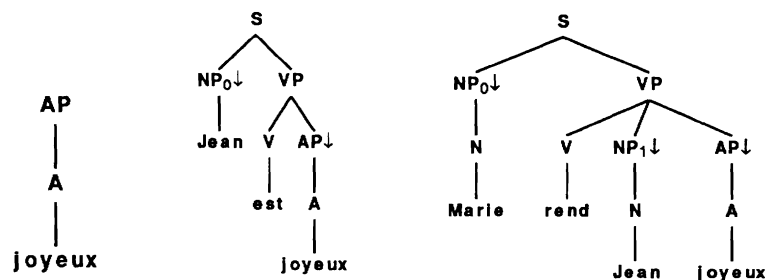
[63]Jean is happy.
How is Jean ?
[64]It is bothersome to go fishing. * It is happy a man.

33

4.2 Adjectives as modifiers or arguments

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 :

73) Jean voit un camion bleu.
74) Jean voit une jolie femme [65]
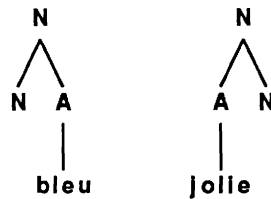
73 and 74 are derived respectively from *Jean voit un camion*, and *Jean voit une femme.*

Adjectives produce then two types of structures, one for their modifying nouns, and one for their being arguments of a sentence structure (noted 7,8,9 and 32, 34, 35 in section 2) for example, NP_0 V NP_1 A :

75) Jean trouve Marie jolie. [66]

To have these adjectives listed twice in the lexicon is not redundant because some adjectives are only one type or the other. The so-called relational adjectives can only be modifiers :
76) C'est une décision ministérielle
*77) *Cette décision est ministérielle.* [67]

In English [Fillmore 1968] gives the reverse example of *ill*, as opposed to *sick*, which can only be attribute and not modifier :
78) The children are ill.
*78a) * I saw ill children.* [68]
Adjectives can thus produce three kinds of tree structure : sentential trees (initial or auxiliary depending on whether they take nominal or sentential complements) , initial AP-type trees (for their being arguments), auxiliary trees rooted in NP for heir being modifiers.

---

[65]Jean sees a blue truck.
Jean sees a pretty woman.

[66]Jean finds Mary pretty.

[67]This is a departmental decision. * This decision is departmental

[68]Unless it has a completely different meaning, as in 'to have an ill effect', which is more of an idiom.

5   The adjunction of adverbs

Adjuncts are not always easy to distinguish from arguments of a given structure. We have listed some commonly used tests in section 1, but we are aware of the fact that the question is still vivid[69].

It is clear that there are obligatory adverbs, such as :
Jean va bien.
Jean agit intelligemment.
There are also adverbs that have a fixed position :
*Jean arrive à peine. * Jean à peine arrive.*
A peine Jean arrive. [70]

[Pollard and Sag 1988] observe that all selectional constraints about adverbs have to be stated on the modified phrase not on the adverbial, in all formalisms, although it is not altogether clear, linguistically speaking, in what direction the selection takes place. It is not completely true for CG which define the type of the adverbs with the type of the phrase they modify, constraining an adjective to apply to a noun for example : N/N. But only in TAG, it seems, is it possible to define the constraints in two ways : either as constraints on adjunction (on the modified phrase), or as feature equation on the foot and root node of the auxiliary tree which represent the adverb and which has to unify with the equation of the phrase. Minimally, an adverb bears the category of the phrase it modifies (by definition of an auxiliary tree)

Adverbs can be :

- 'lexical' adverbs : *souvent, rarement*
- PP : *à huit heures*
- NP : *ce jour-là*
- subordinate clauses : *pendant que Jean lit le journal.*[71]

Lexical adverbs, PP introduced by prepositions, and subordinate clauses introduced by conjunctions, are represented by the proper auxiliary tree(s) in the lexicon.

Lexical adverbs are auxiliary trees with one node (non counting the foot and root nodes) in the general case (e.g. 'probablement'[72]), and more than one for adverbs taking complements (e.g. 'indépendammnet de NP'[73]) :

---

[69]There have been discussions on extraction properties, among others the controversy on parasitic gaps or that on obligatory adverbs [Gross 1988b] and on argument structure [Kegl,Fellbaum 1988]

[70]One might consider the subject verb inversion in :
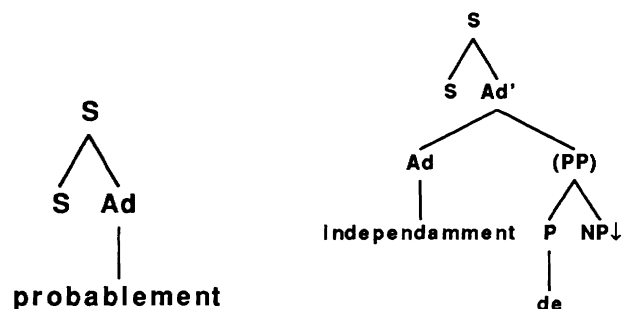*A peine Jean arrive-t-il que tout le monde est en emoi.*
But it is the complex conjunction 'à peine...que' that seems to be at stake here. Similar examples in English are 'once' or 'hardly' :
*John hardly reads French. * John reads hardly French. * John reads French hardly. * Hardly John reads French.*

[71]often, seldom/ at eight o'clock/ that day/while Jean is reading the paper

[72]probably

[73]independently of NP

```
        S                              S
       / \                            / \
      S   Ad                         S   Ad'
           |                            /    \
      probablement                    Ad     (PP)
                                       |      / \
                                Independamment  P  NP↓
                                                |
                                                de
```

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). Prepositons yielding adverbials are thus listed separately from those yielding sentential complements, only the adverbial ones are considered to be the 'head' of the PP. If all prepositions introducing verbal complements(and subcategorized by the verb) yield adverbials also, the opposite is not true. One should also notice that prepositions are listed differently depending on whether they introduce an adverbial phrase attaching to NP, VP, or S. The distinction to make between VP- and S- attachment is not that clear, but that between NP- and VP- attachment opens a new (and promising) field for study. For example, 'excepté NP ' attaches only to NP (and plural NP):

79) Jean aime tout le monde excepté Marie.

80) Tout le monde, excepté Jean, marchait d'un bon pas.[74]

'Sauf' , on the other hand, attaches to either NP or VP , but imposes a different internal structure on its NP argument in each case : the surface structure is (sauf det n) if it attaches to NP (the determiner being that required by the noun ; it can be zero), and is (sauf n) if it attaches to VP:

81) Tout le monde sauf la voisine est venu. [75]

*82) *Jean viendra sauf un accident/la tempête.*

83) Jean viendra sauf contretemps/ accident.[76]

One has the feeling that the phrase attaching to VP is elliptic for ('sauf s'il y a'), but this does not really explain the zero determiner.

The subordinating conjunctions are represented as auxiliary trees rooted in S, in which sentential trees (derived or initial) are substituted :
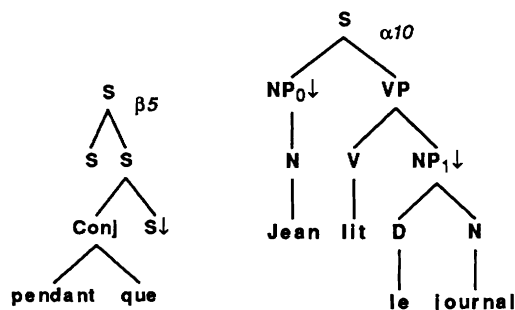
---

[74]Jean likes everyone except Mary.
Everyone except Jean was striding along.

[75]with also the order : ' Tout le monde est venu sauf la voisine'

[76]* Jean will come unless an accident/a tempest.
Jean will come unless accident.

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:

*84) Marie regarde la télé pendant que Jean lit le journal.*
85) * Qu$_i$'est ce que Marie regarde la télé pendant que Jean lit e$_i$?[77]

We also have a case of adjunct clause without subordinating conjiction :

86) Le nez de Cléopatre eût-il été moins long, la face du monde en eût été changée. [78] (Pascal 1666)

We consider the first clause (with the inversion) as an auxiliary tree rooted in S, adjoined to the second one, It is part of the tree family of the verb (être), or more precisely of the tree family projected by its subjunctive form.

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 :
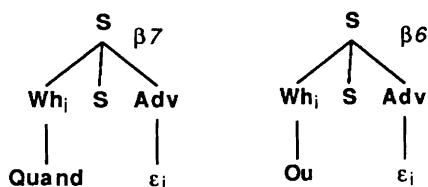
Although the formalism rules out extraction out of adjuncts [Kroch 1986], it does not rule out wh-movement of the adverbial as a whole. It further predicts that only S-rooted adverbials give rise to wh-question:

87) Jean a déploré la destruction de Beirouth Est le 4 Juin [79]

87 is analyzed as being ambiguous, between an S- and an NP- attachment of the

---

[77]Marie is watching TV while Jean is reading the paper.
⋆ What is Mary watching TV while Jean is reading e$_i$?
[78]Has Cleopatra's nose been shorter, the world's fate would have been changed
[79]Jean deplored the destruction of East Beirut on June, 4th

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 1987].

Let us now consider the case of adverbs which are obligatory in a sentence, such as :

88) Jean va bien.[80],

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.

The situation is also interesting when one considers light verb constructions. John made the claim that Mary came yesterday exhibit the same ambiguity as (33), and receive the same treatment. But when we have the same surface structure in a construction whithout a light verb, e.g. 'John beleived/rejected the claim that Mary came yesterday', the only interpretation is 'yesterday' modifying the complement clause. This is represented the following way :

S-node is not available in the upper structure (John believed the claim) to which the that-clause is adjoined. So, it seems natural to say that the upper clause is not accessible to the scope of 'yesterday', if 'yesterday' is first adjoined to the that-clause.

We leave the question of word order variation for adverbs to further research. There is a way of stating their position orthogonally to their attachment described in [Joshi 1987], that is for TAG what LD-IP rules are to CF rules. But in languages like French, like English, word-order variation depends on more constraints, among them the argument structure of the verb and the fact that the insertion occurs in a matrix or an embedded clause. Furthermore, linear precedence statements, as the ones associated with structures such as N_0 V prep N1 prep N_2, or N_0 V S prep N_2, that allow scrambling (cf section 1), might be modified by adjunction (heavy-NP shift). For example, subject-inversion which is not free in french, becomes obligatory if *peut-être* is the first element of the sentence :

*89) * Peut-être Marie a mangé tout le chocolat*
89a) Peut-être Marie a-t-elle mangé tout le chocolat.[81]

6   The representation of idioms

Because in a TAG the linguistic unit is the sentence or the phrase, and not the word, entries comprising of several words can easily be defined. Compound phrases, which

---

[80]Jean is doing fine.
[81]* Maybe Marie ate all the chocolate. Maybe did Mary eat all the chocolate

can be discontinuous constituents, are assigned a head with the same criteria as 'free' structures. There might be items which do not have a clear category because they occur only in frozen phrase, or none of the items involved are of the category of the whole , e.g. a prepositon followed by 'que' yields in french a compound subordinating conjunction. We might then choose the most 'significant' item (lowest frequency). This point will be developed in another paper.

The head produces the subtree corresponding 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*[82] : the adjective *grande* is adjoined to the noun *majorité* as to any N.

---
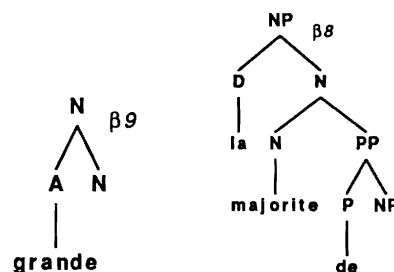
[82] A majority of NP
A large majority of NP

Conclusion

We have discussed both the computational and linguistic advantages of using the TAG formalism for the gramar of French. 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 match 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 1988].

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 light verb constructions are handled in a natural way. Our current approach is being used to build a TAG grammar for English along the same lines [Bishop, Cote, Abeillé 1988].

We also have arguments for thinking that it will be possible to have a complete grammar with only lexicalized rules. Furthermore the choice of the head for each structure can always be linguistically motivated (it would not be so using only CF rules), and a structure can have several arguments at the same level of depth (contrary to a categorial grammar for example). Coordination is still a crucial issue, and we leave the explanation for it to a paper on its own.

The overall size of the French grammar amounts to 70 basic structures (tree-families), which correspond to simple verbs (12), verbs with sentential complements (28), light verb-noun combinations (20), and light verb-adjective combinations (10). An average tree-family is comprised of 15 trees, and the whole size of the grammar will be roughly 1000 trees.

One should notice that what crucially matters is the number of tree-families, which is closed, if our initial collection is exhaustive. We have not incorporated yet pronominalization and coordination, two major remaining phenomena[83]. 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 1987]. 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 70 in the worst case.

Our grammar has been implemented on two parsers : an Earley-type parser as defined in [Schabes and Joshi 1988] and a DCG type parser also written by Yves Schabes [84]. The grammar uses a dictionary which is comprised of more than 3000 lexical items, that are the most common for French.

---

[83]Recent work in TAG has suggested several approaches but we have not pursued them here.

[84]We are experimenting to see which parser is the most efficient one.

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), 'Parsing French with Tree Adjoining Grammar : Some linguistic Accounts', COLING 1988.

K. Bishop, S.Cote, A. Abeillé, 'A lexicalized Tree Adjoining Grammar for English', forthcoming 1988.

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.

J. Bresnan, 'Polyadicity', in it The mental representation of Grammatical Relations, Cambridge University Press, 1982.

A.Daladier,*Problèmes d'Analyse d'un Type de Nominalisations en Français et de Certains Groupes Nominaux Complexes*, Thèse de 3ème cycle, LADL, Paris VII,1978.

L. Danlos, *Representation d'Informations Linguistiques : Les constructions ETRE Prep*,Thèse de 3ème cycle, LADL, Paris VII, 1980.

G.Gazdar, P.Pullum, 'Subcategorization, constituent order and the notion of 'head' ', in *The scope of Lexical rules*, M.Moortgat et al. ed, Foris 1981.

J.Giry-Schneider, *Les Prédicats nominaux en Français : les phrases à verbe-support*, Droz, Genève, 1987.

J.Giry-Schneider, *Les Nominalisations en Français : L'opérateur FAIRE dans le lexique*, Droz, Genève, 1978.

M.Gross, "Lexicon-Grammar : The representation of compound words", *Proceedings of the 11th International Conference on Computational Linguistics*, Bonn, 1986, pp.1-7.

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).

M.Gross, *Grammaire Transformationelle du Français : l'Adverbe*, Cantilene, Paris, 1988(b), forthcoming.

A. Gunnarson, 'Predicative Structures and projection of lexical dependencies', *Linguistic Inquiry*, 1986, vol 18, n0 1.

A. Joshi, "Properties of formal grammars with mixed type of rules and their linguistic relevance", in *Proceedings of International Conference on Computational*

Linguistics, Sanga Saby, 1969.

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.

L. Karttunen, 'Radical Lexicalism', CSLI Report, 1986; to appear in *New Conceptions of Phrase Structure*, M.Baltin, A.Kroch (eds), New York, 1988.

J.Kegl, C.Fellbaum, 'Some non canonical argument identifications', in *Proceedings of WCCFL*, Irvine, 1988.

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.

A. Kroch, "Asymmetries in long distance extraction in a TAG grammar", in M. Baltin and A. Kroch (eds), *New Conceptions of Phrase-Structure*, New-York, 1988.

N.La Fauci, "Aspects du mouvement de wh-, verbes-supports, double analyse. complétives au subjonctif en italien : pour une description compacte", *Linguisticae Investigationes* IV, 2, J.Benjamins, Amsterdam, 1980.

J.Labelle, *Etudes de constructions avec opérateur AVOIR; extensions et nominalisations*, Thèse de 3ème cycle, LADL, Paris VII, 1974.

L.Picabia, *Les Constructions Adjectivales en Français : Systématique Transformationnelle*, Thèse de 3ème cycle, LADL, Paris VII, 1978.

C. Pollard, I.Sag, *Information-based Syntax and Semantics*, vol.1, Univ. of Chicago Press, 1988.

J.Ross, *Constraints on Variables in Syntax*, Unpublished M.I.T. Dissertation, 1967.

M.Salkoff, *Analyse syntaxique du français : Grammaire en chaîne*, J.Benjamins.B.V., Amsterdam, 1979.

M.Salkoff, *Une grammaire en chaîne du Français*, Paris, Dunod, 1973.

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.

K.Vijay-Shanker, A.Joshi, 'Tree-Adjoining-Grammars with Feature Structures', in Proceedings of the *12th International Conference on Computational Linguistics*,

Budapest, 1988.

APPENDIX I : The Tree Families

For each lexical item, we note the name of the structure it is the head of. In the case of predicates (selecting sentences) this structure is a family of trees. These tree-families correspond to surface-argument structures.[85] The overall size of the French grammar amounts to 70 basic structures (tree-families), which correspond to simple verbs (12), verbs with sentential complements (28), light verb-noun combinations (20), and light verb-adjective combinations (10). An average tree-family is comprised of 15 trees, and the whole size of the grammar is roughly 1000 trees.

Below are the conventions adopted for naming the tree structures. We try to make them reflect the composition of the structure.

Notations for basic categories :

P : preposition

Sc : subordinating conjunction

N : noun

D : determiner

C : complementizer

V : verb

A : adjective

Ad : adverb

Cc : coordinating conjunction

Notations for trees :

The name of a tree shows its internal structure.

α means initial tree. The capital letter following α is the category of the root, except for sentential initial trees that are noted T (for Tree-family).

β means auxiliary tree. The capital letter following β is the category of the root and the foot.

T is the symbol for sentence-structures (with verbal head). T' is the symbol for sentence structures headed by nouns or adjectives (copular or light verb constructions). The arguments of a given sentence are numbered from 0 (for the subject) to 3.

n means nominal complement

p means preposition (for a prepositional complement).

s means sentential complement (it is only an S-node, and the whole structure is an S-auxiliary tree).

a means adjectival argument.

---

[85]For predicative nouns we don't want to duplicate entries, for their appearing as sentential head and their appearing as any NP (as subjects for example). We will then consider that an initial tree αNP, with the NP and its (optional) complements, is part of the family of the light verb construction.

Optional complements are parenthesized.
Auxiliary-trees corresponding to modifiers take their name from the category they modify (which is both their foot-node and their root-node). The capital letter following β is the label of the root (and of the foot-node).

Examples[86] :

$Tn_1 =$ this family is comprised of the following trees:
αtn_1, $\alpha W_0 tn_1$, $\alpha W_1 tn_1$, $\beta R_0 tn_1$, $\beta R_1 tn_1$, αCtn_1, $\alpha W_0 Ctn_1$, $\alpha W_1 Ctn_1$, $\beta R_0 Ctn_1$, $\beta R_1 Ctn_1$.

$T'[n_1\ pn_2] =$
$\alpha t'[n_1\ pn_2]$, $\alpha W_0 t'[n_1\ pn_2]$, $\alpha W1+2t'[n_1\ pn_2]$, $\beta R_0 t'[n_1\ pn_2]$, $\beta R_1 t'[n_1\ pn_2]$, $\alpha Ct'[n_1\ pn_2]$, $\alpha W_0 Ct'[n_1\ pn_2]$, $\alpha W[1+2]Ct[n_1\ pn_2]$, $\beta CR_0 tn_1$, $\beta R_1 Ct'[n_1\ pn_2]$, $\alpha In_0 t'[n_1\ pn_2]$, $\alpha In_1 t'[n_1\ pn_2]$ $\alpha In(1+2)t'[n_1\ pn_2]$, $\alpha In_2 t'[n_1\ pn_2]$ $\alpha Wn_0 t'[n_1\ pn_2]$, $\alpha Wn_1 t'[n_1\ pn_2]$ $\alpha Wn_2 t'[n_1\ pn_2]$,$\alpha Wn(1+2)t'[n_1\ pn_2]$ $\alpha Wn_0 Ct'[n_1\ pn_2]$, $\alpha Wn_1 Ct'[n_1\ pn_2]$, $\alpha NP(pn_2)(pn_0$.

Some of the trees of Tn1 :
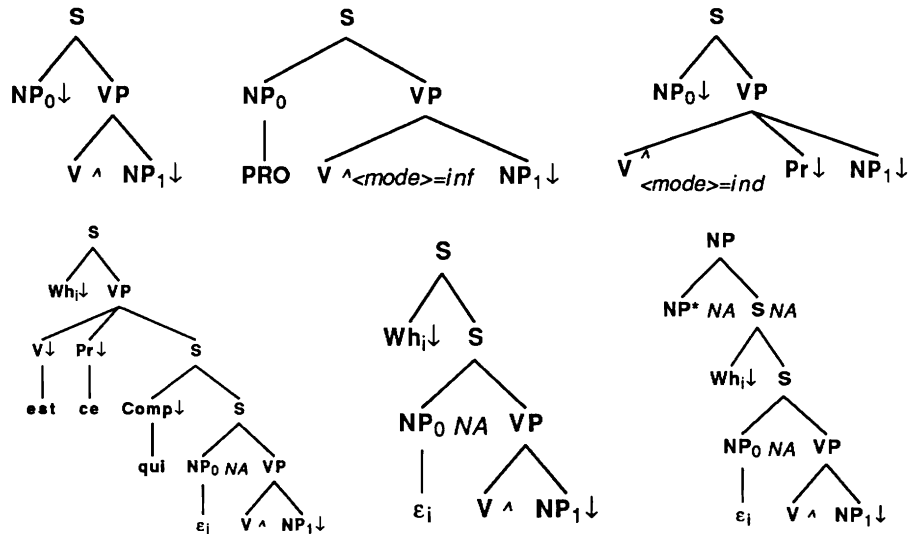
[86]W_i stands for Wh-question on argument i. R_i stands for relativization on argument i. I_i stands for relativization on argument i.
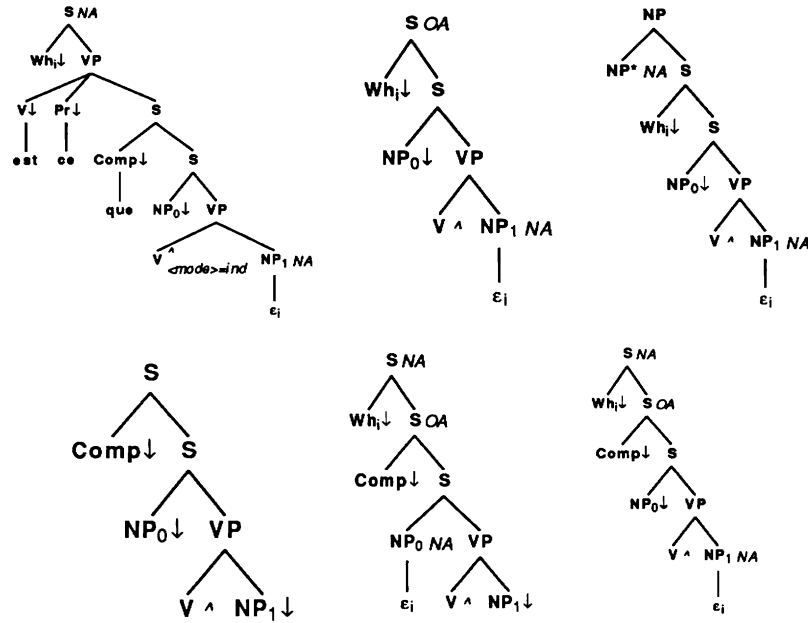
List of the structures handled by the grammar

TREE-FAMILIES (headed by V)[87]

    a- with nominal arguments

T : Jean marche

Tn_1 : Jean embrasse Marie.

$T(n_1)$: Jean mange (une pomme)

Tpn_1 : Jean compte sur Marie.

$T(pn_1)$: Jean part (de Paris)

Tn_1pn_2 : Jean met un livre sur la table.

$Tn_1(pn_2)$: Jean vend une maison (à Marie)

$Tpn_1 (pn_2)$: Cette histoire aboutit à un désastre (pour Marie).

$T(pn_1)(pn_2)$: Jean parle (de son travail)(à Marie)

$Tn_1 \, n_2$: Marie appelle son enfant Jean.

$Tn_1 (n_2)$: On a élu Jean (président)

Ta : Jean est heureux

$Ta(pn_1)$: Ce départ semble fâcheux (à Marie)

Tn_1a : Marie rend Jean heureux

Tad : Jean va mieux

---

[87]The default category of the subject is NP, and only sentential subjects are explicitly mentioned in the structure name.

Tn_1ad : Jean met les choses au pire

$T(n_1)ad$: Jean vole (des fruits) à la tire

$Tn_1(ad)$: Jean trouve des amis (à la pelle)

b- with sentential arguments[88]:

s_0T : Qu'il neige en Août [arrive].

s_0Tn_1 : Qu'il neige [réjouit Jean].

$s_0T(n_1)$: Jouer aux billes [amuse (les enfants)].

Ts_1 : [Jean trouve] qu'il fait froid.

$T(s_1)$: Jean surveille [(que tout aille bien)]

s_0Ts_1 : [Qu'il neige prouve] qu'il fait froid.

Tps_1 : [Jean cherche à ce] que Marie reste.

s_0Tpn_1 : Que Marie reste [compte pourJean]. $s_0T(pn_1)$: Prendre une semaine de vacances par an [suffit (à Jean)].

s_0Tps_1 : [Que Jean parte équivaut à ce] qu'il reste.

Tn_1s_2 : [Jean voit Marie] partir.

$Ts_1 (pn_2)$: [Jean dit] qu'il fait froid (à Marie).

$Ts_1 (ps_2)$: [Jean déduit] qu'il fait froid [(de ce qu'il neige)].

$Ts_1 ps_2$: [Jean tire] qu'il fait froid [de ce qu'il neige].

$Tn_1(ps_2)$: [Jean convaint Marie (de ce] qu'il fait froid).

$s_0Ts_1 (pn_2)$: [Que Marie parte montre] qu'elle est furieuse [(à Jean)].

s_0Tn_1ps : [Qu'il neige habitue Jean à ce] qu'il fasse froid.

$s_0Tn_1 pn_2$: Avoir perdu son portefeuille [a mis Jean dans une situation difficile].

$s_0Tn_1(pn_2)$: Manger de vieilles huitres [donne des boutons (à Marie)].

$s_0Tn_1n_2$: Avoir juré [a consacré Jean membre du clan des damnés].

$Ts_1 n_2$: [Jean considère] aller à la pêche [une perte de temps].

$Ts_1 s_2$: [Jean appelle] aller à la pêche [perdre son temps].

Tpn_1ps_2 : [Jean compte sur Marie] pour que ça aille mieux.

$T(pn_1)ps_2$: [Cet instrument sert (à Jean) à] se curer les ongles.

$s_0Tpn_1 (pn_2)$: Que Jean soit venu [équivaut à une tuile (pour Marie)].

$s_0Tpn_1ps_2$: [Qu'il neige suffit à Jean pour] qu'il achète des skis.

$s_0Tps_1 pn_2$: [Aller à New-York a abouti à] devenir journaliste [pour Marie].

$s_0Tps_1 (pn_2)$: [Travailler chez Jean revient à] ne rien faire [(pour Marie)].

$Tps_1(ps_2)$: [Jean s'aperçoit de ce] qu'il fait froid [(à ce qu'il neige)].

$s_0Ts_1(ps_2]$: [Qu'il neige ajoute] qu'il y ait des morts [(à ce qu'il y a des sans abri)].

$s_0Tps_1 (ps_2)$: [Aller loin diffère d'] aller près [(en ce qu'il faut une voiture)].

Tas_2 : [Jean trouve heureux] que Marie reste

---

[88]The square brackets are put around the S structure of the matrix clause; as above, parentheses mark optionality

Some of the families are selected by compound verbs [89]:

Vprv , T: s *evanouir*
Vprv , Tpn1{p1=a}: s *attaquer*
Vprv , Tps1{N0=il, p1=de}: s *agir*
Vprv , Ts1{s1.b:⟨comp⟩=que}: s *apercevoir*

TREE FAMILIES (headed by nouns)[90]

a- with nominal arguments

$T'[n_1p'n_2]$: Jean a l'air d'un idiot.
$T'[n_1(p'n_2)]$: Jean commet une agression (contre Luc).
$T'p[n_1p'n_2]$: Jean procède au nettoyage de la maison.
$T'p[n_1(p'n_2)]$: Jean est en colère (contre Luc).
$T'[n_1(p'n_2)]$ (pn_3) : Cette lettre donne de l'amour (pour Marie)(à Jean).
$T'[n_1 \ p'n_2]$ (pn_3) : Cette boisson a fait l'effet d'une bombe (sur Jean).
$T'[n_1 \ [pn_2p'n_3]$: Cet exploit met Jean en admiration devant Marie.

b- with sentential arguments:

$s_0T'n_1p'n_2]$: Qu'il neige donne envie de vacances.
$s_0T'p[n_1p'n_2]$: Qu'il aide Marie plaide en faveur de Jean.
$s_0T'[n_1p'n_2]$ pn_3 : Rencontrer un ami donne le courage de ses opinions à Jean.
$s_0T'p[n_1p'n_2]$ (pn_3) : Qu'il aide Jean parle en faveur de Luc (à Marie).
$s_0T'n_1p[n_2n_3]$: Aimer Max met Marie en contradiction avec elle-même
$T'[n_1 \ s_2]$: Jean a la preuve que Bob ment.
$s_0T'[n_1 \ s_2]$: Qu'il rougisse est une preuve que Pierre ment. $T'[n_1p's_2]$: Jean a tendance à exagérer. $s_0T[n_1p's]$: Qu'il ait une place libre est signe de ce que Pierre est parti.
$T'p[n_1 \ s_2]$: Max est sur le point de partir.
$T'pn_1p's_2$: Max est de taille à faire cela.
$s_0T'p[n_1 \ p's_2]$: Quitter Paris va avec le projet de quitter Marie.
$T'[n_1s_2]$ pn_3 : Bob donne l'impression qu'il dort à Luc.
$s_0T'[n_1 \ s_2]$ (pn_3) : Que Bob ferme les yeux donne l'impression qu'il dort (à Luc).
$T'[n_1p's_2]$ pn_3 : Marie donne l'idée de partir à Jean.
$s_0T'[n_1p's_2]$ pn_3 : Que Marie parte donne l'idée de partir à Jean.

some of these families are also selected by compound nouns :

---

[89]the italics mark the 'head' of the compound itself
[90]Square brackets are here just for the sake of readability. They are put around the NP formed by the first complement and its modifier. Parentheses are the mark of optionality

Nn{n=rendu}, T'[n1(pn2)]{V=faire, p2=de} : Jean fait le *compte* rendu (de la séance)

Npn{p=a,n=part}, T'pn1pn2{V=est, p1=en, p2=avec} : Jean est en *porte* à faux avec Marie.

Na{a=inverse}, T'pn1pn2{V=est, p1=en, p2=de} : L' évolution du chomage était en *raison* inverse de l'inflation

aN{a=bons}, T'pn1pn2{V=est, p1=en, p2=avec}: Jean est en bons *termes* avec ses collègues

Npn{p=a, n=part}, T'[n1(pn2)]{V=faire, p2=avec} : Luc fait *bande* à part (avec Zozo)

Npn{p=d,n=ordre}, T'[n1(ps2)]pn3{V=donne,p2=de,p3=a} : [Jean donne le *mot* d'ordre de] rester ici [à Luc].

TREE FAMILIES (headed by adjectives)

s_0Ta : Que Marie reste est heureux.

sTas_2 : [Que Marie reste rend possible] que Jean soit heureux.

sTn_1a : Que Marie reste rend Jean heureux.

T'[ap'N_1] : Max est fier de ce resultat.

T'[a (p'N_1)] : Max est content (de ce resultat.)

s_0T'[apN$_1$] : Tuer son prochain est passible d'emprisonnement à vie.

T'[ap'n_1] (pn$_3$) : Jean est content de ses résultats (pour Marie).

T'n_1 [ap'n_2] : Cette lettre rend Jean fou de Marie.

b- with sentential arguments:

T'[a (s_1)] : Jean est fier (que tout le monde le regarde)

T'[a s_1] : Jean est capable de faire des miracles.

s_0T'[a s_1] : Qu'il fasse beau est susceptible de faire des heureux.

T'[a(s_1)] (pn$_2$) : Jean est content (que Bob aille mieux) (pour Marie).

s_0T'n_1 [apn$_2$] : Qu'elle l'ait embrassé sur le front a rendu Jean fou de Marie.

s_0T[a (ps$_1$)] : Qu'il neige est nécessaire (pour pouvoir skier).

s_0T'[a ps$_1$] (pn$_2$) : Aller à N-Y est nécessaire pour pouvoir skier (pour Max).

T'n_1 [a (s_2)] : Jean rend Marie heureuse (de vivre).

T'n_1 [a ps$_2$] : Cette nouvelle loi rend cette mesure susceptible d'avoir des conséquences fâcheuses.

s_0T'n_1 [a (s_2)] : Qu'il fasse beau rend Marie contente (d'aller en vacances).

s_0T'n_1 [a s_2] : Que Jean lui ai donne de l'argent rend Marie capable d'aller en vacances.

INITIAL TREES FOR ARGUMENTS(simple and compounds)[91]

1) Determiners (and predeterminers)

D, αD : un

Ddnd$_1$, βNPdnp : la *plupart* des

D,αdD : [tous] les

2) Nouns

N, αNPdn : chat

N, αNPn : Paris

compounds:

Nn{n=chat}, αNPdn : *poisson* chat

Nn{n=Merogis}, αNPn : *Fleury* Merogis

aN{a=grand}, αNP(d)n : grand *mère*

Na{a=bleu}, αNPdn : *cordon* bleu

Npn{p=de,n=campagne}, αNPdn : *maison* de campagne

Npdn{p=a,n=perche}, αNPdn : *saut* à la perche

adN{ad=avant}, αNPdn : avant *goût*

Apa{p=de,a=gris}, αNPdn : *vert* de gris

aA{a=petit}, αNPdn : petit *gris*

Vcv{v=vient,c=et}, αNPdn : *va* et vient

Vn{n=mouche}, αNPdn : *tue*-mouches

3) Adjectives

A , αAP: bleu

Aan{n=roi} ,αAP: *bleu* roi

Aa{a=vert}, αAP : *bleu* vert

Apn{p=d,n=eau} , αAP: *vert* d'eau

4) Prepositions

P, αP : pour

Pp1{p=a}, αP : *de* NP à NP

AUXILIARY TREES FOR MODIFIERS[92]

1) N modifiers:

N, βnN : C'est un [chef-]comptable

N, βNn : C'est un ingénieur [conseil]

pN{p=en}, βNpn : C'est un ingénieur [en chef]

---

[91]The head of the compound is marked with italics ; it is usually the element which imposes the category of the compound as a whole, or the most significant element.

[92]In the examples, the adjoined strings are marked with square brackets.

A, βNa : Jean voit une chaise [rouge]
A, βaN : Jean voit une [belle] femme

compounds:
Ppdn{p=a,d=la}, βNpdn : C'est un ingénieur [à la manque]
An{n=roi} ,αAP: Jean voit un lit [*bleu* roi]
Aa{a=vert}, αAP : Jean voit des yeux [*bleu* verts]
Apn{p=d,n=eau} , αAP: Jean voit un tapis [*vert* d'eau]

2) A modifiers:
Ad, βadA : Jean voit une [[très] belle] femme
Pnp{p=de,n=part}, βApn : Ce comportement est habile [de la part de Paul]
P, βApn : Jean est blanc [comme un linge]
P, βApv : Cette histoire est triste [à pleurer.]

3) NP modifiers[93]:
Ad, βNPad : La destruction [hier] des quartiers Est provoque de violents émois dans
la presse.
N, βNPdn : La destruction [ce *matin*] des quartiers Est provoque de violents émois.
A, βaNP : J'ai connu [feu] le roi d'Angleterre.
Ppdnp$_1$, βNPpdn: à l' *intention* de

4) V modifiers:
V, βVv : Jean [peut] marcher
adAd{ad=ne} , βadV : Jean [ne] connait [*pas*] son bonheur

5) S (or VP) modifiers:
Ad, βSad : Jean viendra [probablement]
Ad, βadS : [certainement] Jean viendra
ADpn , βVPad: *pour* une fois
ADnp , βVPad: trois jours *durant*
ADdn , βVPad: ce *matin*
Sc, βSscs : quand [94]
    compounds:
dAD{d= un} , βadVP : un *peu*
ADapa{a=petit,p=a} , βVPad : *petit* à petit
ADadpad{ad=peu,p=a} , βVPad: *peu* à peu
ADnpn{m=mot,p=a} , βVPad: *mot* à mot
ADpad{ad=peu} , βVPad: *avant* peu

---

[93]relative clauses excepted

[94]In Jean vient [quand il veut], [il veut] is substituted at an S-node in the auxiliary tree corresponding to [quand]

50

pNp₁{p=en,p1=de}, βVPpdn : en *dépit* de

Np{p=a}, βVPpdn : *grâce* à

Adp{p=de} , βVPpdn: *hors* de

SCpnc{p=sous, c=que} , βSscs : sous *prétexte* que

SCpdnc{p=pour,c=que}, βSscs : pour la *raison* que

SCpc{c=que} , βSscs: *dès* que

SCdnc{d=chaque,c=que} , βSscs: chaque *fois* que

SCadc{c=que} , βSscs: *aussitôt* que

APPENDIX II : The TAG Lexicon

The lexicon is currently comprised of 1000 verbs, 1500 nouns, 400 adjectives and 100 prepositions, adverbs and determiners. They amount to more than 3000 words that are the most common for French, as established by [Catach 85] and [Juilland, Brodin, Davidovitch 1970]. All entries have the same format :

word, category : tree-structure(s).

The category is the address of the head in the tree structure selected by the word. In the case of a compound, the category itself is viewed as a tree structure and we put both the internal structure and the address of the 'head' of the compound in the category slot.

Sample of the Lexicon :

a, P : βNpn{D=la}, βNpn{D=ϵ}, βVPpn{D=ϵ}.

a,P: αP.

a,V:αV.

a, V {V.b:⟨mode⟩=ind}: Tn1, Tn1p2{p2=de/par}.

avoir,V{V.b:⟨mode⟩=inf} : Tn1, Tn1pn2{p2=de/par}.

abaisser,V{V.b:⟨mode⟩=inf} : Tn1, s0Tn1,Tn1ps2{p2=a}.

abaisse,V{V.b:⟨mode⟩=ind/subj} :Tn1, s0Tn1, Tn1ps2{p2=a}.

abaissement,N{N.b:⟨genre⟩=m} : T'[n1pn2]{D1=l,V=fait/cause/provoque, p2=de}, s0T'[n1pn2]{D1=l,V=cause/provoque, p2=de}, T'p[n1pn2]{D1=l,V=procede,p1=a, p2=de}.

abandon,N {N.b:⟨genre⟩=m}: T'pn1{V=est,p1=a, D1=l}, T'n1pn2{V=met, p2=a,D1=l}.

abandonner,V{V.b:⟨mode⟩=inf} : T(n1){N1=-hum}, Tn1{N1=+hum}.

abandonne,V{V.b:⟨mode⟩=ind/subj} : T(n1){N1=-hum}, Tn1{N1=+hum}.

abattre,V{V.b:⟨mode⟩=inf} : T(n1), s0Tn1.

abat,V{V.b:⟨mode⟩=ind} : T(n1), s0Tn1.

abatte,V{V.b:⟨mode⟩=subj}: s0Tn1.

abbé,N{N.b:⟨genre⟩=m}: αNPdn{D1=l}.

abîme ,N{N.b:⟨genre⟩=m} : αNPdn, αNPdnpn{p=de}.

abimer,V{V.b:⟨mode⟩=inf} : T(n1), s0Tn1.
abime,V{V.b:⟨mode⟩=ind/subj} : T(n1), s0Tn1.
abjurer,V{V.b:⟨mode⟩=inf}: T(n1).
abjure,V{V.b:⟨mode⟩=ind/subj} : T(n1).
aboyer,V{V.b:⟨mode⟩=inf} : T(pn1){p1=contre/apres/sur}.
aboie,V{V.b:⟨mode⟩=ind/subj} : T(pn1){p1=contre/apres/sur}.
abondance,N{N.b:⟨genre⟩=f} : T'[n1pn2]{V=a, p2=de,D1=une}, T'pn1{V=est, p1=en},
T'pn1pn2{V=est, p1=en,p2=Loc}.
abord,N{N.b:⟨genre⟩=m}: αNPdn{D1=l}.
aboutir,V{V.b:⟨mode⟩=inf} : T, s0Tpn1(pn2){p1=a, p2=pour},Tpn1(pn2){p1=a,p2=pour}.

aboutit,V{V.b:⟨mode⟩=ind} : T, s0Tpn1(pn2){p1=a, p2=pour},Tpn1(pn2){p1=a,p2=pour}.

aboutisse,V{V.b:⟨mode⟩=subj}: T, Tpn1(pn2){p1=a,p2=pour}, s0Tpn1(pn2){p1=a,p2=pour}.

aboutissement,N{N.b:⟨genre⟩=m} : T'[n1(pn2)]{D1=l,V=est,p2=de}.
abri,N {N.b:⟨genre⟩=m}: T'pn1{V=est,p1=a,D1=l}, T'[n1(pn2)]{V=est, p2=de/pour}.

absence,N{N.b:⟨genre⟩=f}: αNPdn, αNPdnpn{p=de}.
abus,N {N.b⟨genre⟩=m}: αNPdn, T'[n1pn2]{V=est,p=de}.
abuser,V {V.b:⟨mode⟩=inf} : Tpn1{p1=de}.
abuse,V {V.b:⟨mode⟩=ind/subj}: Tpn1{p1=de}.
accabler,V {V.b:⟨mode⟩=inf} : s0T(n1), Tn1, Tn1p2{p2=de}.
accable,V {V.b:⟨mode⟩=ind/subj}: s0T(n1), Tn1, Tn1pn2{p2=de}.
accepter,V {V.b:⟨mode⟩=inf} : T(n1), Ts1.
accepte,V : T(n1), Ts1.
accompagner,V{V.b:⟨mode⟩=inf} : Tn1, Tn1p2{p2=a/chez}.
accompagne,V{V.b:⟨mode⟩=ind/subj} : Tn1, Tn1p2{p2=a/chez}.
acccompagnement,N {N.b:⟨genre⟩=m}: T'[n1(pn2)]{V=est,p2=pour/de}, T'n1(pn2){p2=a,
V=fait}.
accomplir,V {V.b:⟨mode⟩=inf} : Tn1.
accomplit,V
V.b:⟨mode⟩=ind} : Tn1.
accomplissement,N {N.b:⟨genre⟩=m}:T'[n1pn2]{V=est,p2=de},T'p[n1pn2]{V=procede,p1=a,
p2=de}
accorder,V {V.b:⟨mode⟩=inf} : Tn1, Tn1pn2{p2=a}, Ts1pn2{p2=a}.
accorde,V {V.b:⟨mode⟩=ind/subj}: Tn1, Tn1pn2{p2=a}, TSpn2{p2=a}.
accord,N{N.b:⟨genre⟩=m} : T'n1(pn2){V=donne, p2=pour}, T'n1(ps2){V=donne,p2=pour}.

accoucher,V{V.b:⟨mode⟩=inf} : Tn1(pn2){p2=de, ⟨erg⟩=+}.
accouche,V{V.b:⟨mode⟩=ind/subj} : T(pn1){p1=de, ⟨erg⟩=+}.
accouchement,N {N.b:⟨genre⟩=m}: T'n1(pn2){V=fait,p2=de}.