Grammatical signs in the semantics-syntax interface of a unification grammar

Sylvain Kahane

Abstract

This paper presents a unification grammar where grammatical morphemes are handled in the same way as lexical morphemes: Both kinds of morphemes are considered as signs and described by elementary structures which directly unify with each other. We illustrate our purpose with a fragment of the semantics-syntax interface of French for verbs and adjectives: voices, moods, tenses, impersonal, and *tough*-movement.

1 Introduction

Most of the present linguistic models have been developed from syntactic models, beginning with the X-bar Syntax (Jackendoff 1977). When lexical variations shifted into focus in the 70ies, the models that emerged, like LFG, TAG or G/HPSG brought to the fore the need of a structured lexicon. Presently, the architecture of the lexicon in these models is as follows: A first component gives the descriptions of the lexemes (= *lexical signs*); a second component specifies some operations, called *lexical relations*, which allow the derivation of all inflected forms corresponding to a given lexeme (Bresnan 1998, Sag & Wasow 1999). This architecture of the lexicon can be considered as the third step of the formalization of grammatical signs in linguistic models. Consider the well-known example of the passive voice. In Chomsky 1957 (and further Chomskyan models), the passive is treated as a transformation between two syntactic structures. With the development of the lexical component of linguistic models, it has been proposed to treat the passive as a relation in the lexicon (Wasow 1977, Bresnan 1982). It is not yet a relation between a lexeme and a word-form, but rather a relation between two word-forms, the active and the passive form of a verb.

In a model where the relation between lexemes and word-forms are treated inside the lexicon, grammatical signs, henceforth called *grammmemes*, do not have a real status. It can be said that they correspond to the operations that allows to derive the word-forms from the lexemes, but this question is not put up. In HPSG for instance, word-forms and lexemes explicitly receive the status of linguistic signs, but the question whether lexical relations are also signs is never asked.[1]. Moreover the treatment of lexical relations remains generally quite rough: For instance, in Sag & Wasow 1999:189, the transition from the lexeme EAT to the word-form *eats* is performed in one step, without distinguishing the grammemes of voice, mood, tense and agreement in person and number.

In this paper, we limit ourselves to the semantics-syntax interface, that is, to the correspondence between a semantic representation, limited here to a graph of predicate-argument relations between semantic units, and a syntactic representation, encoded here by a non-ordered dependency tree. It is important to note that our syntactic structure is not concerned with the word order, which is encoded in the (morpho-)topological structure, intermediate between the syntactic and the phonological/graphical structure. For a presentation of the syntax-mophotopology interface, see Gerdes & Kahane 2001. Our whole linguistic model is based on the Meaning-Text Theory (Mel'cuk 1988) and is also inspired by HPSG and TAG. As in HPSG, the structures associated to the signs combine by unification. As in TAG and contrary to HPSG, these signs combine directly, provided that the combination of two signs is validated by a schema of combination and without associating a separate representation

[1] In fact, the question is excluded by the formal nature of lexical relations: A sign associates a *signifié* and a *signifiant*, that is, two different levels of representation, while a lexical relation relates two objects of the same level, either two word-forms or a lexeme and a word-form.

to the phrase resulting of the combination of two units. This, more than the use of dependency trees for the syntactic representation, ensures that that the resulting grammar is a dependency grammar (see works on the derivation tree in TAG such as Vijay-Shanker 1987 or Rambow & Joshi 1994).

Our formalism will be presented in Section 2, as well as first examples of structures associated to lexical signs. Section 3 and 4 are respectively dedicated to the description of grammatical signs for verbs and adjectives in French.

2 The formalism and the lexical signs

We will begin our presentation of the formalism of our semantics-syntax interface with an example: the structure (in XML) associated to the French verb MANGER 'eat'. This rule indicates the correspondence between the predicate 'eat'(x,y) and the lexeme MANGER. The semantic arguments x and y respectively correspond to the subject (*subj*) and the direct object (*dobj*) of MANGER, which must be nouns. The direct object is optional, as indicated by the feature-value opt=+ (and the brackets in the graphical representations). The writing of the rule favors the syntactic structure, which is encoded by the features <tree> for the dependencies (or, equivalently, for the subtrees) and the features <node> for the nodes. The semantic representation is encoded by the features <sem> for the nodes and <arg> for the predicate-argument relations. The correspondences between the syntactic and semantic nodes are indicated by the embedding of the <sem> into the <node>. The features /p_voice/ and /p_mood/ are polar features taking the values + or –: a value – indicates a need, while a value + indicates a resource. Combining p_t=– and p_t=+ entails the suppression of /p_t/: the need has been satisfied. The features p_voice=– and p_mood=– thus indicate the verbal lexeme MANGER must be combined with a voice and a mood. The features /id/ allow us to indicate a feature sharing; for instance, here, the first argument of 'eat' corresponds to the semantic content of the subject of MANGER.

```
<rule><title> MANGER </title>
<tree>
      <node lex="MANGER" cat="V"
      p_voice=- p_mood=->
            <sem cont="eat">
                  <arg n="1" id="[x]"/>
                  <arg n="2" id="[y]"/>
            </sem>
      </node>
      <tree fct="subj">
            <node cat="N">
                  <sem id="[x]"/>
            </node>
      </tree>
      <tree fct="dobj" opt=+>
            <node cat="N">
                  <sem id="[y]">
            </node>
      </tree>
</tree>
</rule>
```

Favoring the syntactic structure gives us a structure very similar to the ones of HPSG: The feature <node> corresponds to the feature HEAD and the features <tree> of a same level to the subcategorization (SUBCAT). Nevertheless, our structure also plays the role of the *head-daughter-phrase* schema allowing the combination of a head and its syntactic arguments: In this case, it is possible to interpret the feature <node> as a feature HD-DTR (head-daughter) and the feature <tree> as a feature NHD-DTR (non-head-daughter).

The structure proposed for MANGER can be represented graphically. In Figure 1 we continue to favor the syntactic structure and in Figure 2 bring to the fore the correspondences between the semantic and syntactic representation. A need p_t=– is represented by →t.

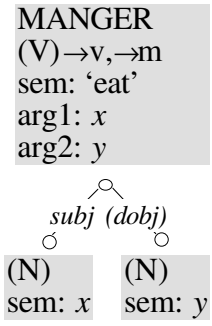


Figure 1 : Rule for MANGER 'eat' (first graphical representation)

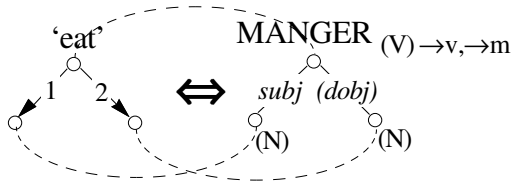


Figure 2 : Rule for MANGER 'eat' (second graphical representation)

The treatment of raising verbs (such as SEMBLER 'seem') and control verbs (such as VOULOIR 'want') forces us to enrich the formalism. We must indicate, in the structure of such a verb, that its subject is also the potential subject of its verbal complement. In order to ensure that, we introduce, in the structure of these verbs, a particular dependency, noted <quasi> for *quasi-dependency* and represented by a dashed arrow. A quasi-dependency unifies with an ordinary dependency (of same function) giving a quasi-dependency.

Below we give the structure of SEMBLER 'seem'. This verb corresponds to a unary predicate, which does not take its syntactic subject as semantic argument. The subject of SEMBLER is entirely controlled by its verbal complement, which is in the infinitive mood and will loose its subject by unification its *subj* dependency with *subj* quasi-dependency of the structure of SEMBLER. On the other hand, the verbal complement of SEMBLER must receive a voice. In the graphical representation, the sequence p_t=+, t="v" is abbreviated by t→v.

```
<rule><title> SEMBLER </title>
<tree>
    <node lex="SEMBLER" cat="V" p_mood=->
          <sem cont="seem">
               <arg n="1" id="[y]"/>
          </sem>
     </node>
     <tree fct="subj">
          <node id="[X]"/>
     </tree>
     <tree fct="pred">
          <node cat="V" p_mood=+
          mood="inf">
               <sem id="[y]"/>
          </node>
          <quasi fct="subj">
               <node id="[X]"/>
          </quasi>
     </tree>
</tree>
</rule>
```

SEMBLER
(V)→m
sem: 'seem'
arg1: *y*

subj *pred*
subj
(V)m→inf
sem: *y*

Adjectives are by default noun modifiers and the noun they modify is their first semantic argument. The French adjective FACILE 'easy' is thus associated to the following structure (its other arguments are omitted):

```
<rule><title> FACILE </title>
<tree>
     <node cat="N">
          <sem id="[x]"/>
     </node>
     <tree fct="mod">
          <node topcat="A" cat="A"/>
               <sem cont="easy">
                    <arg n="1" id="[x]"/>
               </sem>
          </node>
     </tree>
</tree>
</rule>
```

(N)
sem: *x*

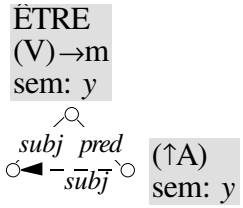
mod

FACILE
(↑A,A)
sem: 'easy'
arg1: *x*

The previous structure can unify with a noun in order to describe a phrase such as *une question facile* 'an easy question'. But the same structure can be used in the other uses of the adjective, as in the predicative construction *La question est facile* 'the question is easy'. The copula ÊTRE 'be' receives the following description:

```
<rule><title> ÊTRE </title>
<tree>
     <node lex="ÊTRE" cat="V" p_mood=->
          <sem id="[y]"/>
     </node>
     <tree fct="subj">
          <node id="[X]"/>
     </tree>
```

```
      <tree fct="pred">
            <node topcat="A">
                  <sem id="[y]">
            </node>
            <quasi fct="subj">
                  <node id="[X]"/>
            </quasi>
      </tree>
</tree>
</rule>
```



Some remarks are in order. First, the copula does not have its own semantic contribution. It receives the semantic content of its adjectival complement because this semantic unit becomes the argument of a predicate that will take the copula as an argument, such as CROIRE 'believe' in *Je crois que la question est facile* 'I believe that the question is easy'. Second, the complement of the copula is not necessary an adjective: It suffices that this unit is transferred into an adjective, what will be indicated by the feature topcat="A" (graphically encoded by (↑A)).[2] Third, we consider that the *mod* dependency acts as an inverted *subj* dependency and can thus unify with a *subj* quasi-dependency; this ensures that the first semantic argument of the adjective will correspond to the subject of the copula.[3] At last, this dependency inversion also inverts the value of the features /cat/ and /topcat/ (because /cat/ controls the category vis-à-vis the dependents and /topcat/ vis-à-vis the governor). Thus in the predicative construction, the adjective FACILE 'easy' can have a verb as argument provided that this verb is transferred into a noun (and have topcat="N") as in *Lire ce livre est facile* 'to read this book is easy'.

We will now present the grammatical signs and the way they unify with lexical signs.

3 Grammatical signs for the verb

As seen before, a verb must combine with a mood and eventually a voice. We consider the following moods for French: indicative, subjunctive, imperative, infinitive, past participle and present participle. Ou category of mood mixes two inflectional categories: the finiteness (finite, infinitive, participle) and the mood proper when the verb is finite. Finite moods require the combination with a tense, using a polar feature /p_tense/. The trivial rule for indicative follows:

```
<rule><title> indicative </title>
<tree>
      <node topcat="V" cat="V" mood="ind"
      p_mood=+ p_tense=->
            <sem mood="ind">
      </node>
</tree>
</rule>
```

We can now give an example of a combination of structures for a whole sentence (Figure 3). The resulting structure is well formed if every polar features has been consumed and if each syntactic node as an instanciated /lex/ feature.

The same grammar could be formalized in HPSG provided that we introduce a specific schema for the combination a lexical sign and a grammatical sign (which will be the head). It results that we will have as many phrasal projections for a verbal construction as kinds of grammatical signs that will combine with the verb. The resulting phrase structure will be closer to the syntactic structures of generative grammars or RRG (Van Valin & La Polla 1997) than usual phrase structure in HPSG (Pollard & Sag 1994).

We need a rule for the infinitive when it is not imposed by the subcategorization as in *Voler est un délit* 'to steal is a crime' or *un livre à lire* 'a book to read'. This infinitive morpheme must block the realization of the subject of the verb and attribute a semantic content to the argument potentially realized

---

[2] The theory of transfer (Fr. *translation*) has been proposed and developed by Tesnière 1959. One consider that a unit is transferred into an adjective if it can fill a position prototypically filled by an adjective. Thus the preposition *of* assumes a role of transferer of the noun *Peter* in *the book of Peter.*

[3] HPSG grammars make the same hypothesis, sharing the values of the features MOD and SUBJ of an adjective.

as a subject; this value is called GEN (as generic). We are not concerned here by the

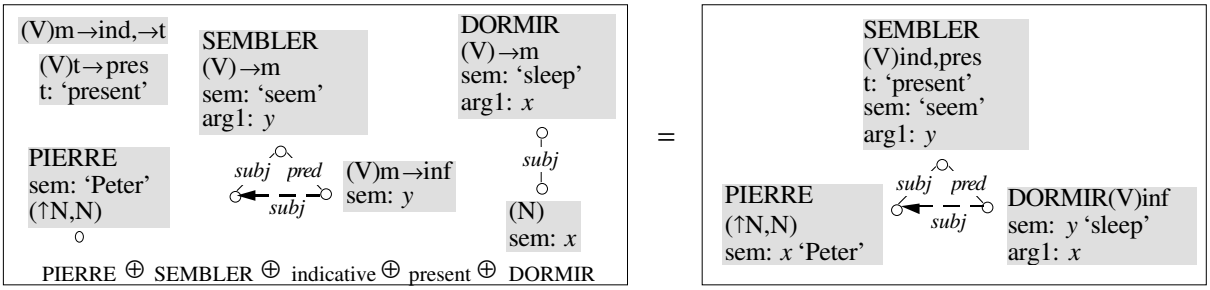


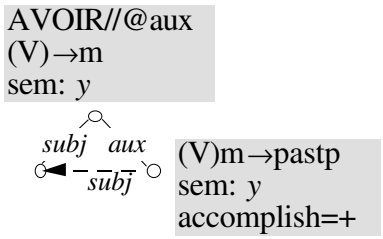
Figure 3: Combination of signs for *Pierre semble dormir* 'Peter seems to sleep'

interpretation of GEN; the semantics-syntax interface must only ensure that in such construction the subject receives a particular interpretation. Note also that the infinitive verb receives a nominal topcat: Such a verbal form can only fill positions where a noun is waited, for example in a subject position or after a preposition. Only finite verbal forms receive a verbal topcat.

```
<rule><title> infinitif </title>
<tree>
      <node topcat="N" cat="V" mood="inf"
      p_mood=+/>
      <quasi fct="subj">
            <node lex=->
                  <sem cont="GEN"/>
            </node>
      </quasi>
</tree>
```

The infinitive does not combine with a tense, but it can combine with the accomplish realized by AVOIR 'have' (or eventually ÊTRE 'be' if the feature /aux/ is instanciated) + past participle:

```
<rule><title> accomplish </title>
<tree>
      <node lex="AVOIR//@aux" cat="V"
      p_mood=-><sem id="[y]"/>
      </node>
      <tree fct="subj">
            <node id="[X]" topcat="N"/>
      </tree>
      <tree fct="aux">
            <node cat="V" mood="pastp"
            p_mood=+>
              <sem id="[y]" accomplish=+/>
            </node>
      </tree>
      <quasi fct="subj"><node id="[X]"/>
      </quasi>
</tree>
</rule>
```



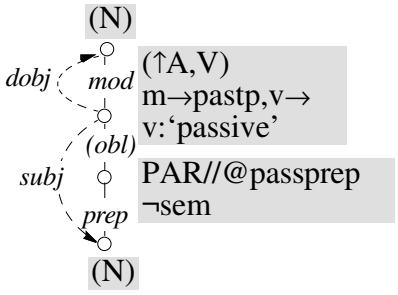
The structure of the accomplish ensures that the potential subject of the verb will be realized on the auxiliary. The auxiliary does not have its own content and receives the semantic content of the participle. The content of the accomplish is encoded by a feature-value accomplish=+.[4] Finally, as the auxiliary requires a mood but not a voice, the accomplish will take place between the mood and the voice (the voice is the most closer morpheme from the verb). The accomplish is compatible with all the moods. It tends to forms "idioms" with the simple tenses of the indicative, like the *passé composé*, syntactically equivalent to a present perfect but semantically equivalent to English simple past.

The structure for the active voice is trivial: It simply provides a feature-value p_voice=+. The treatment of the passive voice is more complex. We consider that in *Le poulet est mangé par Pierre* 'the chicken is eaten by Peter', the verb ÊTRE is the copula as in *La question est facile* 'the question is easy' and that the past participle *mangé* 'eaten' is the

[4] Another possibility will be to introduce a real semantic vertex for the accomplish that would be associated the syntactic vertex of the auxiliary. Nevertheless, we need to ensure that the syntactic governor of the auxiliary takes the content of the verb as semantic argument and not the accomplish.

same element as in *le poulet mangé par Pierre* 'the chicken eaten by Peter' or *Le poulet semble mangé par Pierre* 'the chicken seems (to be) eaten by Peter' (cf. Abeillé & Godard 2002 for a more detailed argumentation). Consequently, we introduce a structure directly for the passive past participle; this structure can combine with the copula to form the passive. In the passive past participle structure, the direct object becomes the nominal governor and the subject becomes an optional oblique complement introduced by the preposition PAR 'by' (unless the feature /passprep/ specifies another value, that is, DE 'of' as in *Pierre est aimé de Marie* 'Peter is loved by Mary').

```
<rule><title> passive past participle
</title>
<tree>
      <node cat="N" id="[Y]"/>
      <tree fct="mod">
          <node topcat="A" cat="V"
          mood="pastp" p_mood=+ p_voice=+>
              <sem voice="passive"/>
          </node>
          <quasi fct="subj">
              <node id="[X]"/>
          </quasi>
          <quasi fct="dobj">
              <node id="[Y]"/>
          </quasi>
          <tree fct="obl" opt=+>
              <node cat="P"
              lex="PAR/@passprep">
                  <sem cont=-/>
              </node>
              <tree fct="prep">
                  <node id="[X]"/>
              </tree>
          </tree>
      </tree>
</tree>
</rule>
```



We finish our presentation of grammatical signs for the verb by the impersonal. The potential subject is demoted in a special position we call quasi-subject (*qsubj*); a subject realized by the impersonal pronoun IL 'it' (semantically empty) replaces it. The impersonal combines freely with a verb, provided that it has no realized subject and that its potential subject is indefinite or verbal (*Il a été invité plus de cent personnes* 'it has been invited more than hundred persons'; *Il me déplaît de venir* 'it dislikes me to come'). We block the presence of a direct object by introducing an optional *dobj* dependency without possible lexical realization (lex=–); this works well if we have well-formedness rules that controls that there is only one direct object and obliges us to unify all the *dobj* dependencies. The application of the impersonal is also blocked if the verb has a *aux* or a *pred* dependent, that is, if it is an auxiliary or a copula; therefore a verb must combine with the impersonal before combining with an auxiliary or a copula. Yet, this rule allows us to construct impersonal passives (*Il a été proposé une nouvelle solution/de changer la direction* 'it has been proposed a new solution/to change the direction'). In this case, the impersonal must combine with the past participle before combining with the copula (and the *subj* quasi-dependency will combine the *mod* dependency between the past participle and its governor; see Section 2 for the combination of the copula with an adjective). The following rule concerns infinitives potential subject.

```
<rule><title> impersonal </title>
<tree>
      <node cat="V"/>
      <quasi fct="subj">
            <node id="[X]" cat="V"
            mood="inf"/>
      </quasi>
      <tree fct="qsubj">
            <node lex="DE" cat="P">
                  <sem cont=-/>
            </node>
            <tree fct="prep">
                  <node id="[X]"/>
            </tree>
      </tree>
      <tree fct="subj">
            <node><sem cont=-/></node>
      </tree>
      <tree fct="dobj|aux|pred" opt=+>
            <node lex=-/>
      </tree>
</tree>
```

4 Grammatical signs for the adjective

Contrarily to the verb, the adjective does not combine with any inflectional sign in the semantics-syntax interface.[5] Nevertheless, the adjective can combine with grammatical signs such as the impersonal (*Lire ce livre est facile* 'to read this book is easy' → *Il est facile de lire ce livre* 'it is easy to read this book') or the so-called *tough*-movement construction (*un livre facile à lire* 'a book easy to read'). The copula ÊTRE, which ensures the transfer of the adjective into a verb, could also be considered as a grammatical sign (see Section 2).

The rule of impersonal for adjectives is the same as the rule for verbs, except that the *subj* quasi-dependency will combine with the *mod* dependency and will be replaced by another *mod* dependency). Note that the impersonal combine with the adjective before combining with the copula. In particular the impersonal is possible without copula: *Impossible de dormir* 'impossible to sleep'; *Facile de lire ce livre* 'easy to read this book'.

We will finish this paper by a description of the *tough*-movement in French, illustrated by *un livre facile à lire* 'a book easy to read' (Figure 4). This construction expresses the meaning 'a book such that to read it is easy'. Therefore the predicate-argument relations remain the same: 'book' is second argument of 'read' and 'read' is the unique argument of 'easy'. The difficulty of this construction comes from the mismatches between the predicate-argument relations and the syntactic dependencies, because dependency between LIRE and LIVRE and the dependency between LIVRE and FACILE does not correspond to a predicate-argument relation. Nevertheless, in the description we propose, the structures of FACILE, LIRE and LIVRE are the usual structures. All the difficulties are assumed by a structure associated to the construction, which we call Adj À Vinf and which resembles to an adjectival voice (in the sense that, as verbal voice, it entails a redistribution of the arguments).[6] Our approach differs with certain descriptions (such as for instance Kayne 1974-75) that propose to derive *ce livre est facile à lire* 'this book is easy to read' from the impersonal *il est facile de lire ce livre* 'it is easy to read this book' by an object raising (with an alternation of the preposition in French difficult to explain). In our approach the two constructions are obtained parallelly by the combination of the adjective with two different grammatical signs.

We can note that the syntactic roles on the dependencies play an important role in our grammar. For instance the *dobj* relation validates the combination of a verb with the passive voice or the construction Adj À Vinf. It will also validate the combination with the accusative clitic pronoun or the agreement with past participle.

5 Conclusion

Beyond the treatment of grammatical morphemes as real signs, the formalism we propose for the semantics-syntax interface has several advantages: 1) it is powerful enough for allowing us to use a same structure for the very different uses of a lexeme (while a formalism such as TAG requires a different structure for each use of a lexeme); 2) it avoids us to recourse to phrasal descriptions or schemata and is thus lighter than a formalism such as HPSG; moreover, this makes the formalism nearly associative, in the sense that the signs can be combined in whatever order (except for the grammatical signs) and the combination is not constrained by the need to saturate the phrases before a combination as in HPSG.

One of the qualities of the formalism comes from the use of quasi-dependencies, which allows the derivation structure not to be a tree (the derivation structure is the structure that describes which sign combines which other; see Vijay-Shanker for TAG). In our

---

[5] In French the adjective agrees with the noun, but this only concerns the syntax-morphotopology interface.

[6] The fact that some information correspond to a construction, rather than to particular lexical units, is used in some frameworks for justifying phrasal rules (Ginzburg & Sag 2000). We see with this example that it is possible to consider signs associates to constructions without considering phrasal description.

description of *tough*-movement, for instance,

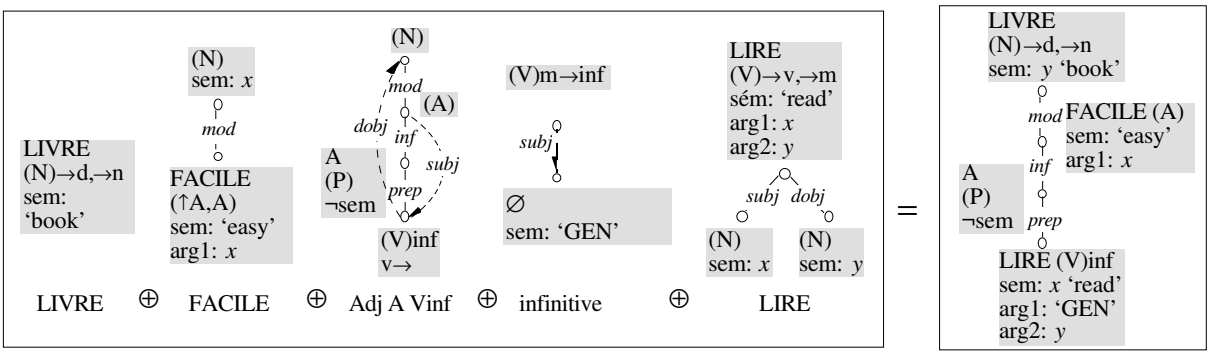


Figure 4 : The construction Adj À Vinf (*tough*-movement)

the sign Adj À Vinf combines with three other signs (a noun, a verb and an adjective) which combine with each other.

The development and the implementation of the French grammar presented here is still in progress. The syntax of nouns and determiners is in the process of realization, as well as clitic pronouns, extractions and the causative construction FAIRE Vinf. A syntax-morphotopology interface has also been developed.

References

Abeillé Anne, Godard Danièle, 2002, "Course on the Syntax of Romance Languages", Bucarest.

Bresnan Joan, 1982, "The passive in lexical theory", in J. Bresnan (éd.), *The Mental Representation of Grammatical Relations*, MIT Press.

Bresnan Joan, 1998, *Lexical-Functional Syntax*, Blackwell, Oxford.

Chomsky Noam, 1957, *Syntactic Structures*, MIT Press, Cambridge.

Gerdes Kim, Kahane Sylvain, 2001, "Word Order in German: A Formal Dependency Grammar Using a Topological Hierarchy", *ACL 2001*, Toulouse, 8 p.

Ginzburg Jonatahn, Sag Ivan, 2001, *Interrogative Investigations: The Form, Meaning, Use of English Interrogatives*, CSLI Publications/University of Chicago Press.

Jackendoff Ray S., 1977, *X' Syntax: A Study of Phrase Structure*, Linguistic Inquiry Monograph, MIT Press.

Kayne Richard, 1974-75, "French relative *que*", *Recherches Linguistiques II*, 40-61, et *III*, 27-92.

Mel' cukIgor, 1988, *Dependency Syntax: Theory and Practice*, State Univ. of New York Press, Albany.

Pollard Carl, Sag Ivan, 1994, *Head-Driven Phrase Structure Grammar*, Stanford CSLI.

Rambow Owen, Joshi Aravind, 1992, "A formal look at dependency grammars and phrase-structure grammars, with special consideration of word-order phenomena", in L. Wanner (éd.), *Current Issue in Meaning-Text Theory*, Pinter, London.

Sag Ivan, Wasow Thomas, 1999, *Syntactic Theory: A Formal Introduction*, CSLI Publications, Stanford.

Tesnière Lucien, 1959, *Eléments de syntaxe structurale*, Klincksieck, Paris.

Van Valin Robert Jr, LaPolla Randy, 1997, *Syntax: structure, meaning and function*, Cambridge University Press.

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

Wasow Thomas, 1977, "Transformations and the lexicon", in P. Culicover, T. Wasow, A. Akmajian (éds), *Formal Syntax*, Academic Press, 327-60.