



HAL
open science

The Grande Grammaire du français project

Anne Abeillé, Danièle Godard

► **To cite this version:**

Anne Abeillé, Danièle Godard. The Grande Grammaire du français project. LREC, 2010, La Valette, Malta. halshs-00751055

HAL Id: halshs-00751055

<https://shs.hal.science/halshs-00751055>

Submitted on 21 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

The Grande Grammaire du français project

Anne Abeillé, Danièle Godard

LLF, University Paris 7 & CNRS

30 rue du château des rentiers

75013 Paris, France

E-mail: abeille@linguist.jussieu.fr, dgodard@linguist.jussieu.fr

Abstract

We present a new reference Grammar of French (*La Grande Grammaire du français*), which is a collective project (gathering around fifty contributors), producing a book (about 2200 pages, to be published in 2011) and associated databases. Like the recent reference grammars of the other Romance Languages, it takes into account the important results of the linguistic research of the past thirty years, while aiming at a non specialist audience and avoiding formalization. We differ from existing French grammar by being focused on contemporary French from a purely descriptive point of view, and by taking spoken data into account. We include a description of all the syntactic phenomena, as well as lexical, semantic, pragmatic and prosodic insights, specially as they interact with syntax. The analysis concerns the data from contemporary written French, but also includes data from spoken corpora and regional or non standard French (when accessible). Throughout the grammar, a simple phrase structure grammar is used, in order to maintain a common representation. The analyses are modular with a strict division of labor between morphology, syntax and semantics. From the syntactic point of view, POS are also distinguished from grammatical relations (or functions). The databases include a terminological glossary, different lexical databases for certain POS, certain valence frames and certain semantic classes, and a bibliographical database.

1. Presentation of the project

Sizable comprehensive grammars, written by linguists but aiming at a larger audience, have been published for several European languages: for Italian, the *Grande Grammatica italiana di consultazione* (Renzi, Salvi, & Cardinaletti, 1989-1991), for Spanish the *Gramática descriptiva de la lengua española* (Bosque & Demonte, 1999), for Catalan the *Gramàtica del Català contemporani* (Solà *et al.* 2002), for English the *Cambridge Grammar of the English Language* (Huddleston & Pullum, 2002).

Nothing of the sort exists for French. Most grammar books are normative, while our aim, as linguists, is purely descriptive. The best existing descriptive French grammars are smaller, written by 2 or 3 authors, and mainly based on written literary French: M. Grevisse, A. Goosse (1936-2008) *Le Bon Usage*, M. Riegel *et al.* (1994-2009), *La grammaire méthodique du français*. While useful, they are based on a “rules and exceptions” approach, with no clear-cut distinctions between grammatical relations and categories, between syntax and semantics. They are not focused on the modern stage of the language, including references to 17th or 18th century authors such as Corneille or Marivaux.¹

Our project, called the *Grande Grammaire du français* is an international research project, which started in 2001 and was sponsored by the French CNRS. It will produce a book of about 2200 pages, to be published in 2011, and several associated electronic databases. It is a collective project, with 3 editors (Anne Abeillé, Annie Delaveau, Danièle Godard), gathering more than 50 authors, from different countries.² Like other large grammars for other Romance languages, it includes the results of recent linguistic research and makes them accessible to a larger audience. It is systematic, fully explicit, but not formalized; it uses both conventional and modern terminology, with explicit definitions and a rich array of examples.

We aim at a complete description of all syntactic phenomena of French, while also taking into account some morphological, lexical, semantic, pragmatic and prosodic aspects. Our descriptions are based on speakers’

² The authors are: A. Abeillé, P. Amsili, I. Bartning, A-M. Berthonneau, C. Beyssade, O. Bonami, P. Cappeau, M. Charolles, B. Combettes, V. Dahlet, M. Darnat, E. Delais-Roussarie, A. Delaveau, M. Desmets, A. Di Cristo, J. Doetjes, B. Fagard, F. Gadet, D. Godard, B. Kampers-Manhe, K. Jonasson, J. Jayez, J-P. Koenig, L. Kupferman, B. Laca, B. Lamiroy, K. Lambrecht, F. Landragin, J-M. Léard, J-M. Marandin, A. Mari, L. Mélis, P. Miller, C. Molinier, F. Mouret, F. Mourlhon-Dallies, W. de Mulder, M. Noailly, B. Post, C. Rossari, L. Roussarie, F. Sabio, C. Schnedecker, I. Simatos, H. de Swart, D. Van De Velde, M. Van Peteghem, L. Tasmowski, L. Tovenà, J. Tseng, C. Vet, C. Vettters, M-T. Vinet, R. Vivès, A. Zribi-Hertz

¹ For a more detailed comparison with other grammars, including Damourette & Pichon (1911-1946), see Abeillé, Delaveau, Godard 2007.

intuitions as well as on large available corpora, both written and spoken. They also include regional variation (esp. Belgium, Switzerland, Quebec) and non standard data. To ensure maximal coherence and readability, it is based on a surfacist phrase structure grammar (with a fixed set of parts of speech and grammatical relations).

Our objectives are as follows:

- describe the variety of uses in modern French, both written and spoken, in France and abroad;
- synthesize existing linguistic knowledge, complete it, and make it accessible also to non linguists;
- clarify and stabilise grammatical terminology (often obscure, not modular, and variable from one French grammar to another).

As our focus is on « ordinary » French (to use Gadet’s 1989 term), the data we analyse come from a variety of sources:

- excerpts from available written corpora (Frantext, newspapers such as Le Monde, le Soir...)
- excerpts from spoken available corpora (Elicop, Ester, Corpaix, C-oral-rom, CRFP, Sankoff-Cedegren corpus...)
- constructed examples for simple cases and ungrammatical forms.

For language variation, we also rely on a team of “regional” experts coordinated by Françoise Gadet.

2. The book

The book will comprise 7 millions characters (about 2200 pages), with two parts: *La phrase et ses constituants* (the clause and its constituents), *De la phrase au discours* (from sentence to discourse).

It is divided into 20 chapters concerned with all major problems of French syntax, and their interaction with the lexicon, morphology, semantics, discourse and prosody. Five chapters are each devoted to a part of speech (POS): Verb, Noun, Adjective, Preposition, Adverb, while three are devoted to a certain type of phrase (Sentence, Relative clause) or construction (Complex predicates, Coordination). Some chapters are more semantically oriented: those on determination and quantification, on proforms, on negation, on tense, aspect and mode, on comparison or adverbial adjuncts (« circumstances »). We also have a chapter on clause types, one on word order, one on discourse, one on phonology and prosody, one on punctuation and the new writing codes (such as those found in SMS).

We also include a glossary and a comprehensive bibliography.

3. The data

We describe a variety of data from modern French (after 1950), based on common usage, both written (newspapers, novels) and spoken (radio, interviews).

We define an appropriate marking to distinguish whether these data are attested or not, grammatical or not,

standard or not, spoken or not, regional or not (see Table 1). We call «non standard» those existing forms which are sociologically stigmatized. We call « regional» forms found in several regions and not stigmatized there. For exemple, the comparative correlatives are introduced by a comparative adverb (or adjective) in standard French (1a), but in Belgium and Provence, they may be introduced by *au* preceding the comparative form :

- (1) a Plus il pleut plus il se plaint
(the more it rains the more he complains)
b R Au plus il pleut, au plus il se plaint

Similarly, causative constructions with *faire* where the causee precedes the infinitival verb are marked: usually considered as typical of Quebec, they are in fact also found in Britany and in the French West Indies, so we mark them as regional:

- (2) a Il fait rire son fils
(he makes his son laugh)
b R Il fait son fils rire

| Mark | Status | Examples |
|------|--|---|
| | unmarked = grammatical, non regional, and standard | <i>Je n'ai vu personne</i> (I have seen noone) <i>Je le leur ferai lire</i> (I'll make them read it) <i>Je n'ai pas de vélo</i> (I don't have a bike) |
| * | ungrammatical | * <i>Je ferai le leur lire</i> * <i>J'ai de vélo</i> |
| ? | dubious | ? <i>Ils se sont restés fidèles</i> (they remained faithful to one another) |
| % | grammatical with speaker's variation | % <i>Je les ferai le lire</i> (I'll make them read it) % <i>Je me demande où est-ce qu'il est allé</i> (I wonder where did he go) |
| ! | non standard (stigmatized) | ! <i>C'est le livre que j'ai besoin</i> (that's the book I need) ! <i>Où que tu vas comme ça ?</i> (Where you go like this ?) |
| R | regional | R <i>J'ai personne vu</i> (I saw noone) R <i>Au plus il pleut, au plus il se plaint</i> (1b) R <i>J'ai fait mon fils rire</i> (2b) |
| B | Belgium | B <i>On est le quantième du mois ?</i> (we are which of the month?) B <i>J'ai difficile de venir</i> (it's tough for me to come) |
| Q | Quebec | Q <i>J'ai pas vu personne</i> (I saw noone) Q <i>J'ai pas benben de chance</i> (I got not much luck) |
| † | Archaic | † <i>Il vaut mieux n'en rien faire</i> (it's better not to do it) |

Table 1: Data marking

4. The terminology

We propose a consistent and workable terminology, which takes linguistic results and generalizations into account. We try to keep existing terms, but redefine them with explicit linguistic properties.

We define 10 morphosyntactic tags for POS (table 2).

| POS | Examples of subtypes |
|--------------------------|---|
| Adjective | cardinal (<i>trois</i>), ordinal (<i>premier</i>), qualifying (<i>rouge</i>), relational (<i>national</i>), intensional (<i>supposed</i>), |
| Adverb | associative (<i>aussi</i>), degree (<i>très</i>), frequency (<i>souvent</i>), manner (<i>gentiment</i>), modal (<i>forcément</i>), |
| Complementizer | <i>comme, quand, que, si, parce que,</i> |
| Coordinating conjunction | <i>car, et, mais, ni, ou, soit</i> |
| Determiner | definite (<i>le</i>), indefinite (<i>un</i>), free choice (<i>n'importe quel</i>), quantifier (<i>chaque</i>) |
| Interjection | <i>bravo, bof, bon</i> |
| Noun | non predicative (<i>table</i>), and predicative (<i>avoir faim</i>) common noun, proper (<i>Jean</i>) |
| Preposition | weak (<i>à, de</i>), locative (<i>dans</i>), temporal (<i>pendant</i>) |
| Pronoun | demonstrative (<i>ceci</i>), indefinite (<i>quelqu'un</i>), interrogative (<i>lequel</i>), personal (<i>lui</i>), quantifier (<i>tout, personne</i>), relative (<i>lequel</i>) |
| Verb | auxiliaries (<i>avoir, être</i>), agentive (<i>parler</i>) and non agentive (<i>arriver</i>) intransitive, transitive (<i>manger</i>), raising (<i>sembler</i>), control (<i>vouloir</i>) |

Table 2: Morphosyntactic tags

We define six phrasal categories:

- NP (noun phrase) : *un enfant* (a child), *celui de Paul* (that of Paul)
- AdvP (adverbial phrase) : *très bien* (very well), *pas du tout* (not at all), *contrairement à vous* (contrary to you)
- AP (adjectival phrase) : *très content* (very happy), *trois ou quatre* (three or four)
- PP (prepositional phrase) : *chez lui, à Paris* (in Paris), *depuis mars* (since March)
- VP (verb phrase) : *va à Paris* (goes to Paris), *tout lire* (read everything), *ne pas comprendre* (to not understand)

- Sentence : *Paul viendra* (Paul will come), *Quel plaisir ce cours* (What a pleasure this class) !
Tous dehors ! (Everyone outside)

| Function | Examples |
|------------------------|---|
| Adjunct | <i>Si tu viens, tu verras</i> (if you come you'll see). <i>Très bien !</i> (very well) |
| Coordinate | <i>Paul et Marie</i> (Paul and Marie) <i>livres et dossiers</i> (books and files) |
| Extracted | <i>Un ami avec qui parler</i> (a man with whom to talk) <i>Combien ça vaut ?</i> (how much does it cost) |
| Head | <i>Le livre</i> , (the book) <i>Paul viendra</i> , (Paul will-come) <i>très bien</i> (very well) |
| Marker | <i>commencer à lire</i> (begin to read) <i>Paul et Marie</i> (Paul and Marie) <i>Je crois que Paul vient</i> (I think that Paul comes) |
| Object | <i>Paul regarde le ciel</i> . (Paul watches the sky) <i>Il est venu des visiteurs</i> . (there came visitors) |
| Oblique complement | <i>Paul va bien</i> , (Paul is doing fine) <i>Je te retrouve rue Lepic</i> (I meet you rue Lepic) |
| Peripheral | <i>Paul, tu le connais ?</i> (Paul you know him) <i>La mienne, de voiture, marche</i> (Mine of cars works) |
| Predicative complement | <i>Il est médecin/ intelligent</i> (he's a physician/ smart) |
| Specifier | <i>trois questions</i> , (three questions) <i>semblables aventures</i> (similar adventures) <i>quantité de livres</i> (quantity of books) |
| Subject | <i>Les enfants sont là</i> . (the kids are there) <i>Que tu sois venu me fait plaisir</i> . (that you came makes me happy) |

Table 3: Grammatical functions

We also define 11 grammatical relations (or functions) associated with words and phrases (see Table 3).

Our linguistic choices are based on modularity. POS are based on morphological and distributional criteria and are carefully distinguished from valence frames and from grammatical relations. In most French grammars and lexicons, these notions are not kept apart.

For example, in most lexicons, the form *avant* (before) receives at least three different POS: Preposition, when followed by an NP as in (3a), Adverb when used alone

(as in 3b), and Subordinating Conjunction when used with ‘que’, as in (3c). When it comes to verbs, a verb like *attendre* (wait) allows for the same continuations and does not change its POS (3d-f). We thus propose to have one word *avant* with one POS Preposition and three different valence frames: an NP complement, no complement or a sentential complement.

- (3) a Je viendrai avant le match (I’ll come before the game)
 b C’était mieux avant (it was better before)
 c Je viendrai avant que tu partes (I’ll come before you leave)
 d J’attends (I wait)
 e J’attends le match (I wait for the game)
 f J’attends que tu partes (I wait for you to leave)

We also keep POS separated from grammatical relations. Nouns for example can modify other nouns ; this does not turn them into adjectives, as shown by the fact that they keep their gender, as *maison* in (4a), and possibly their number (4b). Similarly, adjectives may modify verbs ; they are not analysed as adverbs because they lack a characteristic property of adverbs, which is to occur between auxiliaries and past participles without an intonation break ; thus, the adjective *net* in (4d) contrasts with the adverb *nettement* in (4e).

- (4) a un sandwich maison (a home-made sandwich)
 b un fichier clients (a file (for) customers)
 c Paul a refusé net (Paul has refused flat)
 d * Paul a net refusé
 e Paul a nettement refusé (Paul has flatly refused)

Grammatical relations are based on syntactic criteria rather than semantic ones. The subject relation for example is based on verb agreement, canonical preverbal position, the possibility of realisation by weak forms such as *il* and *ce*, and the impossibility of realisation by other weak forms like *en*. Objects, on the other hand, are defined by the possibility of realisation by different possible weak proforms (*le, la, les* for definite objects, *en* for indefinite objects) and canonical postverbal position. With this definition, NP subjects can play different semantic roles; for example: agent for *Paul* in (5a), location for *ce tiroir* in (5b), goal for *Paul* in (5c). Postverbal NPs in (6a,c) on the other hand are analysed as objects, even though they correspond to the first semantic argument of the verb *arrive* in (6a) and of the infinitive in (6c), because it can be realised by *en* in the first case (indefinite), and by *le* in the second one (definite).

- (5) a Que lave Paul ? (What is Paul washing ?)
 b Ce tiroir contient de l’argent (This drawer contains money)
 c Paul a reçu des livres (P has received books)

- (6) a Il arrive souvent des accidents
 (there often happen accidents)
 b Il en arrive souvent
 c Paul fait travailler son fils
 (Paul makes his son work)
 d Paul le fait travailler

5. The structures

Throughout the book, we use a simple phrase structure representation for all the constructions presented. We do not force binary branching structures, and use “flatter” structures than usually assumed, for example for tense auxiliaries (see Abeillé & Godard 2002 for discussion). Figure 1 shows the structures assumed for tense auxiliary constructions : *a lu Proust* (has read Proust) and *est allé à Rome* (is gone to Roma).

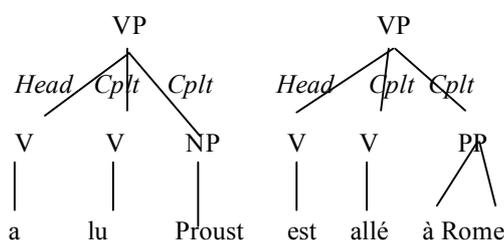


Figure 1: Tense auxiliaries

We do not define Sentences by the usual NP-VP schema, which is too restricted to account for the data. We define sentences (the non elliptical ones) as saturated phrases with a predicative head. Accordingly, they may not have a syntactically realized subject nor a verbal head.

We follow Miller (1992) in analysing weak personal proforms as verbal affixes, that leads us to simpler structures than usual for certain sentences. An imperative clause with a proform complement (such as *regarde-le* ‘watch him’) is a sentence reduced to a verbal head. An interrogative clause with a postverbal proform subject (such as *comment va-t-il ?* ‘how is he going’) can be reduced to an extracted adverb and a verb (Figure 2).

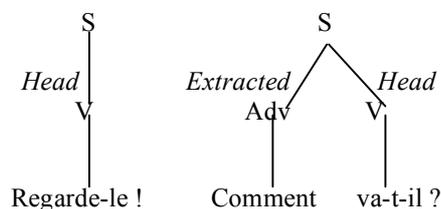


Figure 2: Weak proforms as affixes

We also analyse as sentences (clauses) certain verbless utterances, with a propositional meaning, and a non verbal head. Figure 3 shows the structure associated with an interrogative clause with an adverbial head *Combien le café ?* (How much for the coffee ?) and a clause with a prepositional head, with the pragmatics of an order *Les petits en scène !* (the small ones on stage !):

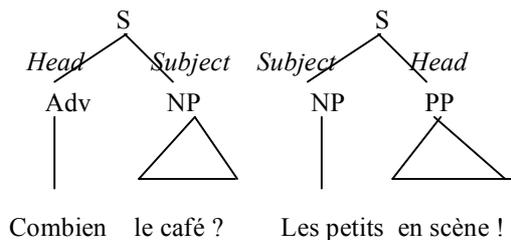


Figure 3: Verbless sentences

For the sake of simplicity, we do not use feature structures. The only complications we allow ourselves is the annotation of the branches of the trees with the grammatical relations, (see figures 1, 2, 3) and referential indices, to indicate similarity or difference of reference. There are few places where this choice raises a difficulty. One is the case of antecedentless relative clauses (7). The first phrase is a *wh*-phrase which functions both as the antecedent of the relative clause and the extracted phrase in the relative clause. Thus, if the extracted phrase corresponds to a PP in the relative clause, it is only acceptable if the same PP is required in the matrix clause (hence, the contrast between (7b) and (7c)).

- (7) a J'ai rencontré [qui m'avait invité]
(I met who had invited me)
- b J'ai discuté [avec qui tu avais discuté]
(I discussed with whom you had discussed)
- c *J'ai rencontré avec qui tu avais discuté
(I met with whom you had discussed)

We can represent this structure as a fragment : a structure without a head (as answers often are, as in *When will you be here ? – Tomorrow*). Thus, the matrix PP directly dominates a relative clause whose extracted constituent is also a PP. But, to express the fact that one phrase plays two roles, it is necessary to use either feature structures, or to allow a representation which is not strictly a tree, since an expression can be dominated by several (two) branches (as do Huddleston and Pullum 2002).

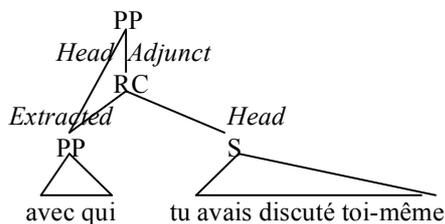


Figure 4: Antecedentless relative clause (6b)

6. The databases

We define several lexical databases, for closed classes of grammatical words (like complementizers, determiners), for certain valence frames (subjectless verbs, verbs with a predicative complement), and for certain semantic classes (degree and associative adverbs, locative

prepositions). An example is Table 4 which gives an exhaustive list of French determiners. Note that we include some originally complex expressions which have evolved into lexemes.

| | Determiners |
|----------------------------|---|
| definite | <i>Ce, le, lequel (relative, interrogative), son</i> |
| Indefinite and quantifiers | <i>aucun, un, un de ces (+ Nsg), un certain, certains, chacun, chaque, de, de ce, de son, du, nul, plusieurs, quel, quelque (sg), zéro, je ne sais quel / lequel, on ne sait quel / lequel, Dieu sait quel / lequel</i> |
| Free choice items | <i>nimporte quel, tout</i> |

Table 4: French determiners

The other items which can combine with a Noun to form a Noun Phrase are adjectives (cardinals, indefinites such as *quelques* and adverbs (degree quantifiers such as *beaucoup*). Adjectives may combine with determiners (8a,b) whereas determiners are defined as non combining with one another (8c)). Adverbs, contrary to Determiners, may not combine with unmarked Nouns, and must be followed by *de* (8d,e) (a preposition used with the marker function):

- (8) a Mes trois chats (my three cats)
- b Ces quelques fleurs (these few flowers)
- c *Ces mes chats (these my cats)
- d *Beaucoup chats
- e Beaucoup de chats (many cats)

Another example is Table 5 which gives an exhaustive list of non reflexive verbs taking tense auxiliary *être* in standard modern French.

| | |
|---|---|
| Only être | <i>advenir, aller, (ré)apparaître, arriver, décider, (re)devenir, intervenir, naître, (re)partir, parvenir, provenir, rester, (re)venir, survenir, mourir</i> |
| Être in intransitive use | <i>(re)monter, (re)rentrer, (re)descendre, (re)sortir, retourner, rentrer, tomber</i> |
| Free alternation between être and avoir | <i>accourir, demeurer, (dis)convenir (de), éclore, échapper, échoir, expirer, passer, (dis)paraître, résulter (de)</i> |

Table 5: Non reflexive verbs taking auxiliary *être* in standard French

In French, the default tense auxiliary is *avoir*, but all reflexive verbs take *être* (9a,b). A small subset of intransitive verbs also take *être*, in at least one of their uses, and they can be analysed as (a subset of) non

agentive intransitive verbs (Marandin 1999):

- (9) a Jean a lavé le chien (Jean has washed the dog)
- b Jean s'est lavé (Jean has washed himself)
- c Jean est parti / mort (Jean has left/died)
- d Un train a/est passé (a train has passed)

Another type of database is devoted to certain polymorphic grammatical words, which are notoriously difficult to analyse. Table 6 shows our analysis for the form *que*, which can correspond to a Proform, to a Complementizer or to an Adverb.

| POS | Examples |
|---|--|
| Proform (wh-interrogative word) | <i>Que voulez-vous ? (what do you want?)</i> |
| Degree adverb (wh- exclamative word) | <i>Que de bruit ! (what a noise)</i> |
| Associative Adverb | <i>Paul ne voit que toi (Paul sees only you)</i> |
| Complementizer | <i>Je pense qu'il viendra (I think that he'll come) (complement sentence)</i> <i>Le livre que je lis (the book that I read) (relative clause)</i> <i>Plus grand que toi (bigger than you) (comparative clause)</i> |

Table 6: French *que*

7. Site of the project

The project has a web site with a closed part, for the authors, and an open part.

<http://grammaire.univ-paris-diderot.fr/>

8. References

- Abeillé A., D. Godard. (2002). The Syntactic Structure of French auxiliaries. *Language* 78, pp.404--452.
- Abeillé A., A. Delaveau, D. Godard (2007), La grande grammaire du français : principes de construction, *Revue roumaine de linguistique*, 52-4, pp. 419--436
- Blanche-Benveniste, C., C. Rouget, F. Sabio. (2002). *Choix de textes de français parlé, 36 extraits*. Paris : Champion.
- Bosque I., V. Demonte (Eds). (1999). *Gramatica Descriptiva de la Lengua Española*, 3 vols. Madrid : Real Academia Española.
- Corblin, F., H. de Swart (Eds). (2004). *Handbook of French Semantics*. Stanford : CSLI.
- Cresti, E., M. Moneglia (Eds).(2005.) *C-ORAL-ROM, Integrated Reference Corpora for Spoken Romance Languages*. Amsterdam: J. Benjamins.
- Damourette, J., E. Pichon. (1911-1946). *Des Mots à la Pensée ; essai de grammaire de la langue française*. Paris: Editions Vrin.
- Gadet F. (1989). *Le français ordinaire*, Paris :Armand Colin.
- Grevisse, M., A. Goosse. (2008). *Le Bon Usage*. Louvain-la-Neuve : DeBoeck-Duculot (14th edition).
- Huddleston, R., G. Pullum (Eds). (2002). *The Cambridge Grammar of the English Language*, 1 vol. Cambridge, Cambridge University Press.
- Marandin J-M. (1999). Unaccusative Inversion in French, in Y. d'Hulst et al. (Eds), *Romance Languages and Linguistic Theory*, Amsterdam: John Benjamins.
- Miller P. (1992). *Clitics and constituents in Phrase Structure Grammar*, New York: Garland.
- Renzi, L., G-P. Salvi, A. Cardinaletti (Eds).(1989-1991). *Grande Grammatica italiana di consultazione*, 3 vols. Bologna : il Mulino.
- Riegel, M., J-C. Pellat, R. Rioul. 2009. *Grammaire Méthodique du français*. Paris : PUF (4th edition).
- Solà, J. et al. 2002. *Gramàtica del Català contemporani*, 2 vols. Barcelone : Editorial Empúries.