1

Organization of the Lexicon-grammar of French Verbs*

Christian Leclère
Institut Gaspard Monge (CNRS), Université de Marne-la-Vallée

Introduction

The lexicon-grammar developed at LADL (Laboratoire d'Automatique Documentaire et Linguistique, CNRS) can be described as an electronic syntactic dictionary. The information it contains is systematically classified, with the goal of reducing as much as possible the arbitrariness inherent in this type of project. The data can be accessed by computer by using the individual words in the index (the lexical intentory) as pointers to entries in the data base.

The lexicon-grammar is organised in a series of tables. Each table groups lexical items with related behaviour (in the structure of basic sentences, in the distribution of arguments, in terms of semantics). Each table includes a set of properties (in columns) and a flag of "+" and "-" indicating whether or not the item has this property. I will say one word about the general principles of this classification, which have been presented on numerous occasions previously (see in particular M. Gross 1975, 1982, J.-P. Boons, A. Guillet and Ch. Leclère 1976a,b, Ch. Leclère 1989, A. Guillet and Ch. Leclère 1992, Vivès 1993a).

With each word in an electronic morphological dictionary, we can associate a tag which refers to the table (or tables) where the word is described. Thus, we find, in a first version of the DELAS, the entry:

achever, V6 [finish]

V6 indicates the type of conjugation. (DELAS is an acronym for: Electronic Dictionary of LADL for Simple Forms (B. Courtois 1990).

In fact, a different gloss for *achever* is appropriate for each of the different syntactic tables of the electronic dictionary in which it appears. So this becomes, in the new version of the DELAS:

achever, V6 + 1, 4, 32H, 38R, 39

to indicate that the verb *achever* belongs to syntactic classes 1, 4, 32H, 38R and 39. My goal, in this short presentation, is to specify what these codes mean, by presenting an overview of the general organization of the lexicon-grammar of French verbs (there are other indications in the DELAS that I won't comment on here).

1. Entries in the tables

Each entry in a table includes:

- the morphological entry (e.g. the verb in the infinitive);

2

- the structure of the "defining sentence", which characterizes the particular use of the word which the entry relates to (with the particular distribution of arguments);
- some constructions associated with this defining structure (particular distributions, substructures, metaphors, etc.).

A verb appears in as many tables as it has uses that are judged to be distinct. A morphological entry is not always a simple item. For verbs, we have distinguished:

-- **simple verbs**: a simple verb is one where all the elements which surround it have variable distributions. This is the case for *mettre* [put, set] in the use:

Luc met le livre sur la table
Luc puts the book on the table

(One can put a lot of things in a lot of places).

-- **compound verbs**: in many cases (the majority in fact), we view the verb (or verbal predicate) as consisting of several unseparable words. It does not make sense, for example, to study *mettre* [set] and *voile* [sail] separately in describing:

Flo [*met les voiles*]
Flo makes herself scarce, *lit.* Flo sets sails

This expression therefore goes into a table of frozen expressions in the form *mettre les voiles* (table C1D in this case). It is also in this form that you will find it in the electronic dictionary of compound words DELAC (B. Courtois & M. Silberztein 1989, 1990). In the dictionary of simple words, in contrast, you will find both *mettre* (put) and *voile (sail)* as access keywords (indispensible for computer analysis). In the same way,

se mettre mal avec (quelqu'un)
get on the wrong side of someone / get into someone's bad books

se mettre à dos (quelqu'un)
make (someone) hostile / get someone's back up

are compound verbs.

Some cases fall somewhere between simple and compound. Table 32R3, for example, groups together simple verbs whose direct object is very constrained, but not unique: there is a small, well-delimited set of words involved. We have chosen in these cases to introduce a canonical representative of the class in the entry. We have, for example, an entry *parler/langue* [speak/language] for describing the use:

Luc parle le francais, le chinois, l'arabe
Luc speaks French, Chinese, Arabic

-- What we have called **support verbs** constitute another special case. We consider that in combinations like:

3

mettre une gifle (*à quelqu'un*) [give a slap (to somebody)]
(= *gifler* [slap])

the noun is the morphological entry that determines the predication. The verb *mettre* [set] and its variants like *donner* [give] only provide "support". That is, they support the tense, person and number of the nominal predicate.

The uses of such support verbs (like *avoir* [have], *faire* [do, make], *donner* [give], etc.) depend on the noun. So they can't be described as verbal entries. It is more interesting to classify such uses starting from the description of the nouns (Giry 1978, 1987, G. Gross 1989, J. Labelle 1974, R. Vivès 1983 and, for an overview, R.Vivès 1993).

Finally, we note that uses that contain a simple preverbal pronoun, such as:

s'évanouir [faint]
l'emporter (sur) [have an advantage over]

or verbs with negation as:

ne laisser pas [never stop]
Paul ne laisse pas de m'étonner [Paul never stops surprising me])

although they relate several elements, are currently classified as simple verbs.

2. Verb properties

Properties of various types are listed in the tables, in describing the different uses of verbs. I only report now on those which play a crucial role in the asignment of verbs to different tables (the "defining properties").

2.1. Structural properties

These are the formal basis of our work. Ignoring complements viewed as adjuncts (though there would be a lot to say on this fundamental aspect of the work), verbs are classified according to the structures in which they appear which have the most fully-specified meaning. Such constructions are usually the most structurally complete ones: those which contain the most relevant complements.

Simpler constructions are seen as derived, provided that the derivation processes are familiar and well-described. If they are not, we consider that we have autonomous uses which deserve a separate entry. Thus, for the verb *recouvrir* [cover], we have the sentences:

(1) *On a recouvert la table d'une nappe* (N_0 *V* N_1 *de* N_2)
We covered the table with a tablecloth

(2) *On a recouvert la table* (N_0 *V* N_1)
We covered the table

(3) *Une nappe recouvre la table* (N_2 *V* N_1)

4

A tablecloth covers the table

It is the structure of (1) which will provide a basis for the classification (in table 37M). Sentence (2) is derived from (1) by a well-known operation (*Prép* N_2 has been erased). The possibility of a derived sentence like (3) is indicated by property N_2 *V* N_1 in table 37M, which specifies the extent of the phenomenon to several other verbs of the same class.

The presence of a *Qu P* [that S] complement in the construction is one determining factor in the choice of the class to which the verb belongs and thus of the table in which it appears. The verb *confier* [confide, entrust], for example, as in:

(4) *Paul confie son problème à Marie*
Paul entrusts his problem to Mary

(5) *Paul confie à Marie qu'il doit partir*
Paul confides to Mary that he must go

will be classified as a sentence complement verb (table 9) because of (5). Sentence (4) is considered as being derived from (5) (that he must go is his problem), and inventoried as such in table 9. In contrast, the sentence

(6) *Luc confie sa valise a Max*
Luc entrusts his suitcase to Max

cannot be derived from a sentence complement, and so it appears in a table for constructions with nominal complements (table 36DT in this case). It is interesting to notice that, in many cases, the uses we distinguish have different translations, but not always in the same constructions, as here for *confide* and *entrust*.

Hence, we arrive at a system of priorities in the placement of verbs into tables (priorities which elsewhere do not only concern structural properties).

The structures which come into play in the classification, and which we will encounter again in the final schemas, are the following (The 'x' after certain codes indicates that there are several tables of this type.):

STRUCTURES	TABLES
N_0 *V*	C0, CADV, 31x
N_0 *V* N_1	C1x, CAN, CDN, 4, 6, 32x
N_0 *V Prép* N_1	CP1, CPN, C5, CV, 1, 2, 5, 7, 8, 34L0, 35x (except 35RR)
N_0 *V* N_1 *Prép* N_2	C7, C8, C1PNx, CNP2, C1P2, C6, CPQ, 3, 9, 10, 11, 12, 13, 16, 36x, 37x, 38x (except 38L and 38RR), 39
N_0 *V Prép* N_1 *Prép* N_2	CPP, 14, 15, 17, 18, 35RR

5

N_0 V N_1 Prép N_2 Prép N_3 CPPN, CPPQ, 38L, 38RR

Each of the *N* positions in these structures can be occupied by:

- a noun (tables 31 to 39)
- a sentence complement (tables 1 to 18)
- a noun phrase constrained by a frozen expression (tables C)

The order of separation is in fact the opposite of the above: the frozen expressions with compound verbs are classified first, then sentences with sentence complements, and finally nominal constructions with simple verbs.

2.2. Distributional properties

The structure alone, though it is at the basis of the classification, is not a sufficient criterion: the classes defined are too large and heterogeneous. We have added specifications on the nature of possible arguments at each of the positions.

In the first place, as we have seen, we distinguish verbs whose construction type contains a *Qu P* [that S] complement. Here we must add infinitive phrases (*V-inf*), for which we can specify the subject when its coreference is constrained (e.g. *V^1-inf W* for an infinitive phrase whose subject is the *N_1* of the main clause; *W* represents any set of complements).

A second distinguishing criterion involves the specification of a preposition (*Prép*):

Prép =: à	C7, 7, 9, 11, 14,15, 33, 36DT, 36R
Prép =: de	C8, 8, 12, 13, 37E, 37Mx
Prép =: Loc (locative)	2, 3, 34L0, 35L, 35ST, 36SL, 38LH, 38LS, 38LD, 38L0, 38LR, 38L
Prép =: avec (symmetrical)	35S, 36S
Prép miscellaneous	CP1, CPN, C5, CV, C1PNx, CNP2, C1P2, C6, CPQ, C0, C0Q, CPP, CPPN, CPPQ, 1, 5, 10, 16, 17, 18, 35R, 38PL, 38R, 39, 35RR, 38RR.

The combination of these criteria allows us for example to distinguish between *crier* [shout], *employer* [use], and *tendre* [offer] in the following uses:

Paul crie que tout va bien à Lea *N_0 V Qu P à N_2* (table 9)
Paul shouts to Lea that everything's OK

Paul emploie Max à laver le sol *N_0 V N_1 à V^1-inf W* (table 11)
Paul employs Max to wash the floor

Luc tend un stylo à Marie *N_0 V N_1 à N_2* (table 36DT)

6

Luc offers Mary a pen

A third type of distributional criterion has been used, which involves the characterization of nouns. These simple tests, although sometimes difficult to define formally, have been retained to the extent that they permit the isolation of a particular "natural class" of verbs. This last notion, initially intuitive, can often be made more precise by reliable syntactic tests. These properties are basically the following (we have placed with each of them the code of one table in which they play a crucial role):

N =: Nhum Human
32H: *Paul raped Ida* [Paul a violé Ida]

N =: Nplur Plural noun
32PL: *collectionner les timbres* [to collect stamps]

N =: Npc Part of the body (inalienable)
32CL: *Paul caresses le dos de Marie* [Paul caresses Mary's back]

N =: Nconcret Concrete noun
32C: *Luc a abimé le livre* [Luc damaged the book]

N =: V-n Noun morphologically related to a verb. It could be a root noun or a derived one (cf. to cover/ cover, coverage, covering)
37M: *Paul couvre le gateau d'un linge* [Paul covers the cake with a cloth] (the cloth is a cover)

N =: Nloc Locative noun.
38L1: *Paul quitte la pièce* [Paul leaves the room]

N =: Ncontraint Constrained noun belonging to a small well-defined set. A canonical example of the class is recorded in the table
32R3: *parler une* <u>*langue*</u> (*français, anglais*...) [speak/<u>language</u>]

N =: C Constrained noun, constituting a compound verb with the primitive verb.
C1: *Paul met les voiles* [Paul makes himself scarce]

In the last case, there is an ambiguity in what we regard as being a verb. For practical reasons, we have, in the C tables, favoured the decomposed structure where the *V* means the primitive verb of a frozen expression. Thus, *Flo met les voiles* is accounted for by the $N_0\ V\ N_1$ class, where $N_1 = C$. But, as I indicated, a better analysis would be to say that [*V C*] (*mettre les voiles*) is a compound verb. We can therefore simultaneously speak of a structure $N_0\ V$ where $V = V\ C$.

2.3. Semantic properties

Some of the properties which I have called 'distributional' appeal to semantic intuition, at least in uncertain cases. Those listed here explicitly concern semantic interpretation :

7

V =: Vmvt	Verb of movement: 2, 3;
Active/Stative	Specifies the interpretation of spatial constructions: 35L, 35ST;
V = transformer en *V-n*	The verb can be paraphrased with *transformer en* [transform into] *V-n* : (*gazéifier* = transform into gas): 32CV;
V = rendre Adj	The verb can be paraphrased with *rendre* [make] *Adjective*, or, more often, with *rendre plus* [make more] *Adjective*, with an adjective morphologically related to the verb (*éclaircir* [brighten] = make or make more bright [clair]: 32RA;
N1 = appearance	The direct object is created in the course of the event, as in *Luc a sculpté un bas-relief* [Luc sculpted a bas-relief]: 32A;
V = mettre/enlever	The verb can be paraphrased with *mettre* [put, set] (or *enlever* [take away]) in constructions where the locative is the direct object, thereby placing this object in a *PP* complement (*munir la porte d'un verrou* [furnish the door with a deadbolt] = *mettre un verrou à la porte* [put a deadbolt on the door]. *Débarrasser le grenier des caisses* [clear the attic of boxes] = *enlever les caisses du grenier* [take the boxes out of the attic]: 37E, 37M;
Double event	This property does not appear explicitly; however, it is used to distinguish, in constructions with *avec* [with], between those which are symmetrical: (*flirter* [flirt]) and those which aren't (*manger* [eat], for example): 35S, 36S, 36SL;
Dative	This notion makes it possible to distinguish, among the set of possible dative complements, between those which characterize "real" dative verbs involving an exchange between two humans (*offrir qqch à qqn* [offer somebody something] and those which don't, such as *bricoler qqch à qqn* [fix/mend something for somebody]: 36DT.

2.4. Properties of related constructions

To define certain classes which seemed interesting to us, we availed ourselves of two base constructions instead of one.

In one case, these two constructions are linked in a relation of paraphrase which defines the class (the "=" sign indicates this paraphrase, which is not always exact):

8

Qu P V (Prép) N1 *Que Paul vienne plaît à Marie*
[That Paul is coming pleases Mary]
= *Il V Qu P (Prép) N1* *Il plaît à Marie que Paul vienne* 5
[It pleases Mary that Paul is coming]

N_0 V Qu P *Paul admire que Marie fasse cela*
[Paul admires the fact that Mary does this]
= *N_0 V N_1 de V-inf W* *Paul admire Marie de faire cela* 12
[Paul admires Mary for doing this]

N_0 V Loc N_1 *Les fautes pullulent dans ce texte*
[Mistakes abound in this document]
= *N_1 V de N_0* *Ce texte pullule de fautes* 34L0
[This document abounds with mistakes]

N_0 V Na de Nb *Luc imite les attitudes de Max*
[Luc imitates Max's postures]
= *N_0 V Nb Prép Na* *Luc imite Max dans ses attitudes* 32R1
[Luc imitates Max in his postures]

N_0 V avec N_1 *Luc flirte avec Léa*
[Luc flirts with Léa]
= *N_0 et N_1 V* *Luc et Léa flirtent* 35S
[Luc and Léa flirt]

N_0 V N_1 avec N_2 *Luc a marié Max avec Léa*
[Luc married Max to Léa]
= *N_0 V N_1 et N_2* *Luc a marié Max et Léa* 36S, 36SL
[Luc married Max and Léa]

Another case is where one of these related constructions is grammatically unacceptable. Here we find a well-known productive transformation which is blocked for certain uses of the verb (for exemple, the passive is blocked for constructions *N_0 V N_1* in table 32NM):

N_0 V (Prép) Qu P **Paul commence à ce qu'il travaille*
[Paul begins that he work]
= *N_0 V (Prép) V-inf W* *Paul commence à travailler* 1, 2
[Paul begins to work]

N_0 V N_1 (Prép) Qu P **Paul envoie Léa qu'elle cherche du pain*
[Paul sends Léa that she look for bread]
= *N_0 V N_1 V-inf W* *Paul envoie Léa chercher du pain* 3
[Paul sends Léa to look for bread]

N_0 V N_1 *Le sac pèse 3 livres*
[The sack weighs 3 kg]
= *N_1 être V-pp par N_0* **3 livres sont pesées par le sac* 32NM

9

[Three pounds is weighed by the sack]

N_0 *V* N_1	*Luc caresse le dos de Léa*		
	[Luc caresses Léa's back]		
= N_1 *est V-pp* (stative)	**Le dos de Léa est caressé*		32CL
	[Léa's back is caressed]		

Apart from these cases of defining relations, the transformational operations that we have retained are coded in columns in the tables. This is the case for example for the relation sometimes called "moyenne" ('middle') which links N_0 *V* N_1 and N_1 *V*: a column N_1 *V* appears in the tables of transitive verbs. As we saw for :

Paul covers the bed with a blanket
A blanket covers the bed

In conclusion, each entry of a verb in a table supposes that:

- the verb can be used in the defining construction of the table;
- it cannot be used with the same meaning in any more complex defining construction with higher priority in any other table;
- the construction in question is not a derived sentence. If this is the case, it is the source sentence that must be considered.

So, finally, the five entries of the verb *achever* that I took as an example in the introduction of this article are the following:

achever Table 1	*Max achève de peindre le mur* Max is finishing painting the wall
achever Table 4	*Qu'il ait fait cela a achevé ses parents* That he did it exasperated his parents
achever Table 32H	*On a achevé les blessés* We have shot the wounded
achever Table 38R	*Max a achevé son discours par un plaisanterie* Max rounded off his speech with a joke
achever Table 39	*Max achève sa carrière (E + comme) colonel* Max is coming to the end of his career as colonel

Conclusion

About 6,000 verbs (spelt differently) have been examined. The separation of these uses has led us to classify about 31,000 entries in 81 tables which can be partitioned as follows:

- **25,000 frozen expressions** (with compound verbs) divided into 20 tables;
- **3,400 simple verbs with sentence complements** divided into 18 tables;

10

- **10,300 simple verbs with nominal complements** divided into 43 tables.

The counts given here (and in the tables) are approximate: the tables are constantly being revised. The verbs (such as technical ones) whose use presents too many problems do not appear in the tables.

As said before, we have only listed here the defining properties of the tables. Within these tables, numerous other properties figure in the columns, selected as a function of their value in describing the syntactic class under consideration (see Annex 2). In all, some 500 properties have been examined.

References

Boons, Jean-Paul; Alain Guillet; Christian Leclère. 1976a. *La structure des phrases simples en français: constructions intransitives*. Genève: Droz, 378 p.

Boons, Jean-Paul; Alain Guillet; Christian Leclère. 1976b. *La structure des phrases simples en français: classes de constructions transitives*. Rapport de Recherches du LADL n° 6, Paris: Université Paris 7.

Courtois, Blandine; Max Silberztein. 1989. Les dictionnaires électroniques DELAS et DELAC. In *RELAI*: *Recherches en Linguistique Appliquée à l'Informatique*, Actes du colloque "La description des langues naturelles en vue d'applications informatiques" (Québec, 1988), Québec: Université Laval.

Courtois, Blandine; Max Silberztein (eds.). 1990. *Les dictionnaires électroniques du français. Langue Française* 87, Paris: Larousse.

Courtois, Blandine. 1990. Un système de dictionnaires électroniques pour les mots simples du français, *Langue Française* 87, Paris: Larousse.

Giry-Schneider, Jacqueline. 1978. *Les nominalisations en français. L'opérateur* faire *dans le lexique*. Genève: Droz.

Giry-Schneider, Jacqueline. 1987. *Les prédicats nominaux en français. Les phrases simples à verbe support*. Genève: Droz.

Gross, Gaston. 1989. *Les constructions converses du français*, Droz: Genève.

Gross, Maurice. 1975. *Méthodes en syntaxe*. Paris: Hermann.

Gross, Maurice. 1982. Une classification des phrases "figées" du français. *Revue Québécoise de Linguistique* 11:2, Montréal: UQAM.

Guillet, Alain; Christian Leclère. 1992. *La structure des phrases simples en français 2*:. les constructions transitives locatives. Genève: Droz.

Labelle, Jacques. 1974. *Etude de constructions avec opérateur avoir (nominalisations et extensions)*. Thèse de troisième cycle, Laboratoire d'Automatique Documentaire et Linguistique, Paris: Université Paris 7.

Leclère, Christian. 1989. Les mots ont-il une syntaxe? *Le Français dans le Monde*, numéro spécial: *Et la grammaire?*, Paris: Edicef.

Vivès, Robert. 1983. *Avoir, prendre, perdre: constructions à verbe support et extensions aspectuelles*. Thèse de troisième cycle, LADL, Université Paris 7.

Vivès, Robert (ed.), 1993a. Le lexique-grammaire du français. *L'Information grammaticale* 59, *Le lexique-grammaire du français*, Paris, L'Information grammaticale.

Vivès, Robert. 1993b. La prédication nominale et l'analyse par verbes supports. *L'Information grammaticale* 59, *Le lexique-grammaire du français*, Paris: L'Information grammaticale.

11

Notes

*) A first version of this article has been published in French in *Langue française* 87, 1990, Paris: Larousse. It has been translated by Matthew Stone, University of Pennsylvania, and revised by Antoinette Renouf, University of Liverpool, to both of whom I am grateful. Any remaining errors and clumsy turns of phrase are of course mine.

The translations provided are as close as possible to the French examples, rendering some of them rather unnatural.

Summary

The "Lexicon-grammar" of LADL describes about 15,000 simple verbs and 25,000 complex verbs, according to the syntactic, distributional or semantic properties of their main constructions. I present the types of properties that have been selected as the basis for the classification of these verbs.

Received december 25, 2001.