Is the Lexicon-Grammar exploitable for language processing?

Eric Laporte

Is the Lexicon-Grammar exploitable for language processing ?

Éric Laporte
Université Paris-Est

Abstract

The Lexicon-Grammar of French is a dictionary with structured syntactic-semantic information. In order to assess its exploitability in language processing, we survey four criteria: readability, degree of formalisation, degree of validity of information content, and richness in information. We contribute concrete examples to inform this discussion. We compare the significance of the criteria, in order to evaluate the validity of the priorities retained and of the compromises adopted in the course of the construction of the Lexicon-Grammar.

Keywords : lexicon, syntax, lexicon-grammar, parsing, language resource, language processing

1. Introduction

The Lexicon-Grammar (LG) of French is a large lexical, syntactic and semantic database. It was not designed for exploitation in language processing from the beginning of its construction, namely around 1968, in any case not only. However, since then, several authors of the LG claimed such a potential of exploitation, and this potential was one of the main reasons why the elaboration of the LG met a long-lasting support, continued beyond the first years, was extended to other languages than French, and is still alive now. Some researchers have questioned whether the LG is exploitable for language processing, generally in private conversations. Others exploited substantial portions of the LG of French verbs for automatic sentence parsing (Roche, 1999; Paumier, 2001; Blanc, 2006; Tolone, 2010). Still others studied obstacles to the exploitation of the LG in language processing (Hathout and Namer, 1997, 1998; Gardent *et al.*, 2005, 2006). In this paper, we intensify reflection on this issue, and we bring to the debate concrete examples on four aspects : visual readability of the format, degree of formalization, degree of validity and quantity of information content[1].

With sincere admiration, we dedicate this study to Christian Leclère, one of the main authors of the LG of French distributional verbs, and one of the pillars of the Laboratoire d'automatique documentaire de linguistique (LADL)[2].

---

[2] This is the translation of an article in French published in T. Nakamura, É. Laporte, A. Dister, C. Fairon (Eds.), *Les Tables. La grammaire du français par le menu. Mélanges en hommage à*

2. Visual readability

Even a quick look at an extract of the LG (Fig. 1) shows that one of the priorities in its conception was the readability of its format.

N0 =: Nhum	N0 =: N-hum	N0 =: Nnr	N0 =: V-inf W	N0 être V-n	Ppv	Ppv =: se figé	Ppv =: en figé	Ppv =: y figé	Nég	<ENT>	N0 être V-ant	N0 est Vpp	N0 V de N0pc	[extrap]	Nactif V N0	<OPT>
~	~	~	~	~	<E>	-	-	-	-	barboter	~	~	~	~	~	Le gaz barbote dans l'eau
+	+	-	-	-	<E>	-	-	-	-	barboter	-	-	-	+	~	Max barbote dans l'eau
~	~	~	~	~	<E>	-	-	-	-	basculer	~	~	~	~	~	La chaise bascule
-	+	-	-	-	<E>	-	-	-	-	battre	-	-	+	-	~	Son coeur bat
-	+	-	-	-	<E>	-	-	-	-	béer	+	-	-	+	~	Sa bouche bée
-	+	-	-	-	<E>	-	-	-	-	blouser	+	-	-	-	~	Le chemisier blouse
~	~	~	~	~	<E>	-	-	-	-	boiter	~	~	~	~	~	Cette chaise boite
~	~	~	~	~	<E>	-	-	-	-	bomber	~	~	~	~	~	La voiture bombe
~	~	~	~	~	<E>	-	-	-	-	boucler	~	~	~	~	~	Le programme boucle
-	+	-	-	-	<E>	-	-	-	-	bouffer	+	-	-	-	~	Ses manches bouffent
~	~	~	~	~	<E>	-	-	-	-	bouger	~	~	~	~	~	La dent bouge
~	~	~	~	~	<E>	-	-	-	-	bouillir	~	~	~	~	~	L'eau bout à cent degrés

Figure 1. Extract of table 31R (Boons *et al.*, 1976) in version 1.1.

Lexical items are easy to identify visually and to compare : they are the rows in the table. Similarly, the syntactic-semantic features are materialized by the vertical alignment of their values: they are the columns. This tabular format crosses on a single screen dozens of lexical items with dozens of features. Thus, when a lexicologist encodes an item, he or she views the description of comparable ones, provided that each table collects a homogeneous enough class of items. This visualization facilitates the encoding. It requires that feature labels (or headings) are repeated on each screen, and therefore very brief: headings hardly exceed thirty characters.

If syntactic-semantic lexicons large and satisfactory enough for language processing were presently available, if the scientific community had reached a consensus about the features to be provided or the lexical items to be distinguished, or if it were proved that the perspective of constructing such lexicons in a completely automated way is not largely utopian, then the tabular format of Fig. 1 might be out of purpose. But it is far from being the case. This format facilitates, in particular, scientific discussion about manual construction of language-processing lexicons, at a moment when such discussion is needed: in the LG, for example, some lexical items have not been encoded, such as those of *basculer* "totter" and *boiter* "limp" in Fig. 1; some constructions, in particular pronominal ones, are not represented; adjectival items are under construction...

Each syntactic-semantic feature contributes to describing a construction. The '*N0 est Vpp*' heading in Fig. 1 stands for a sentence with stative interpretation, consisting of

the subject N_0 of the basic construction, of the copula *être* "be", and of the past participle of the verb. In the case of *s'évanouir* "faint", an item with a basic construction exemplified by *Max s'évanouit* "Max is fainting", the stative construction in question is that of *Max est évanoui* "Max has fainted". The '*N0 =: N-hum*' feature specifies that the subject position in the basic construction can be occupied by a noun phrase denoting a non-human entity. For the *blouser* "blouse" item described in this figure, this corresponds to sentences such as *Le chemisier blouse* "The blouse blouses".

As feature labels are required to be mnemonic, they were made of symbols for values of word features or phrase features: *N* for noun, *est* "is" for the verb *être* "be", *pp* for past participle; but, in order to keep labels brief, the names of the word features or phrase features themselves have generally not been specified, respectively here: part of speech, support verb, tense/mood.

It would be aventurous to exploit the LG in a language processing system without making sure first that it can be completed and updated, and therefore that it exists in a readable and editable format. Yet, few authors pay attention to this criterion in their assessment of the LG. Hathout and Namer (1997, 1998), Gardent *et al.* (2005, 2006), for example, do not allude to it. They adopt a consumer attitude and do not care about the beginning of the production chain. Most actors involved in language processing, incidentally, consider manual elaboration of language resources as one of the nightmares of the field, a pitfall to be avoided, and a source of errors. This view is probably a consequence of intellectual inclinations and tastes of computer scientists, but it is objectively irrational, and we think the domain should collectively question it.

For Gardent *et al.* (2005, 2006), the LG format is not standard, because constructions do not take the form of feature structures, with explicit feature names and value names, such as those used by present systems. With such conventions, the two syntactic-semantic features above become formulae such as those of Fig. 2, or equivalent formulae in XML, even more verbose.

```
construction:[predicate:[part-of-speech="verb", mode="participle",
            tense="past"],
        support-verb:[part-of-speech="verb", lemma-list:[value="être"]],
        arguments:(constituent:[position="0",
            distribution:[component:[category="NP"]]])]

constituent:[position="0",
        distribution:[component:[category="NP", human="false"]]]
```

Figure 2. Two features in the form of feature structures.

These conventions are obviously incompatible with the brevity requirements of manual edition on a tabular format. The feature-structure standard is meant for other uses than readable display. For lexicon edition and updating, the Comlex (Grishman *et al.*, 1994) and FrameNet (Fillmore et Atkins, 1994) projects did not adopt feature-structure formats either. In research, conformity to standards must go with more open-

minded questioning than in engineering. On the contrary, experience gained in the construction of LG — and in other projects of production of large-coverage language-processing lexicons, but there are few — has direct bearing on the construction of standards and norms. Here lies the relevance of the reflection on the LG format conducted by the Genelex project (Alcouffe *et al.*, 1993), which was one of the sources of the standardization project Eagles. The Lexsynt project, similarly, took the opportunity of taking account of the LG when contributing to the LMF standard (Francopoulo *et al.*, 2006).

3. Degree of formalization

One of the obstacles to the use of the LG in language processing is its degree of formalization. It is more formalized than the TLF (Dendien and Pierrel, 2003), in which syntactic-semantic features are described by text or suggested by examples, and not specified in normalized labels; but it is less so than a syntactic parser.

3.1. Representation of features

Syntactic-semantic features are represented by brief labels (cf. Section 2), less precise than formalisms handled by syntactic parsers or by grammars to represent syntactic constructions. For example, in '*N0 V vers N*', which labels a construction exemplified by *Des animaux divaguent vers le fleuve* "Animals are wandering towards the river", the *N* symbol stands for a noun phrase, determiner included, such as *le fleuve* "the river". In '*N0 V N1 Dnum N*', which labels the construction in *Max loue son studio 400 euros* "Max rents out his studio for 400 euros", the same *N* symbol now stands for a noun, whereas the determiner, here *400*, is separately symbolized by *Dnum*. Relevantly, Hathout and Namer (1997) notice that some information is implicit, not entirely specified or represented in a non-uniform manner.

During the 2000s, the Lexsynt and LMF projects sparked among specialists of syntactic parsing a revival of interest in LG. This motivated investigation into solutions to this insufficiency of formalization, in particular through the use of recursive transition networks (Paumier, 2001; Blanc, 2006) or of formulae more precise than the traditional feature labels. However, such formulae cannot be as brief as the latter (cf. Section 2): thus, the solution is not to simply substitute the formulae for the labels, since these retain their function. Gardent *et al.* (2005) suggest rather than the LG information should be made usable in language processing systems through a preprocessing that would shift them to a level of formalization equivalent to that of the LMF standard, and possibly be encoded in XML[3]. Thus, Constant and Tolone (2010) transcode the LG information in the form of sets of feature structures comparable to those in Figure 2. This process combines features that contribute to

---

[3] This gap between LG and usual lexicon formalisms for language processing explains probably why some specialists of the domain do not consider the LG to be a lexicon. It has however the major characters of a lexicon, notably its structure in lexical items and its information content.

describing a same construction, for example the two features mentioned in Section 2: the LG itself does not explicitly connects them (Gardent *et al.* 2005), except through the N_0 symbol in the two headings. The formulae produced by Constant and Tolone (2010) are more appropriate to language processing, but do not definitively resolve the problem, because syntactic-semantic metalanguage varies from system to system and from theory to theory, and LMF does not attempt to standardize it. It is difficult to represent syntactic-semantic features by formulae both complete and satisfactory for all systems and all theories. Other similar processes can therefore be considered in parallel.

Furthermore, systematic work on feature labels has been undertook at LIGM, to slightly raise their degree of formalization, without however substantially changing their conventions, their brevity and their readability. Thus, the horizontal headings which materialize a classification of the features in printed versions of the tables have been deleted in 2003-2004: they certainly helped readability, and provided information, but made the column headings complex objects consisting of several labels. In connection with the remotion of the horizontal headings, the information in them has been incorporated into the labels. Another example, in 2009, the '*(N1)(de V1 W)*' feature, encoded in class 12 of verbs, has been relabelled as '*Qu Psubj =: Qu Ni Vsubj W = (Ni) (de Vi-inf W)*'. This feature links the construction exemplified by *Le ressort empêche la bague de glisser* "The spring prevents the ring from sliding" to that of *Le ressort empêche que la bague glisse* "The spring avoids that the ring slides". The use of the *N1* symbol to refer to the raised subject, here *la bague* "the ring", was questionable because this symbol designates already the whole sentential object, here *que la bague glisse* "that the ring slides"; this is why *Ni* was substituted for it.

3.2. Documentation of features

Syntactic and semantic features not being defined precisely by their labels, they are documented in scientific publications (for distributional verbs, Gross, 1975; Boons *et al.*, 1976a, 1976b; Guillet and Leclère, 1992). But this documentation is not sufficient:

- none of these four books has been published in English translation to date;

- Boons *et al.* (1976b) has never been commercially disseminated;

- definitions are not always accurate enough for experts in parsing, which are not always experts in syntax;

- and a same label may be used for different features according to the classes; for example '*N0 =: N-hum*' indicates that the subject position in the basic construction (N_0) can be occupied by a noun phrase denoting a non-human entity, while the verb keeps its canonical sense (cf. above *Le chemisier blouse* "The blouse blouses"), except in class 31H, where this same label indicates that the sentence then takes a metaphorical sense, as in *Le paysage sommeille* "The landscape dozes" vs. *Max sommeille* "Max is dozing".

Thus, Hathout and Namer (1997) consider the interpretation of tables difficult. To remedy this problem, the most comprehensive documentation, which is that of Guillet and Leclère (1992: 409-430), has been completely revised, extended to all features and translated into English, in a collective work in which Christian Leclère took part.

3.3. Delimitation of classes

The LG distributes lexical items in classes. An item's membership to a class implies certain characteristic syntactic-semantic features that define the class. The authors have selected for this the major features: the number of objects in the basic construction, the presence or absence of prepositions introducing these objects, the possibility that they consist of a clause, etc. Thus, class 9 (Gross, 1975) collects those verbs with the basic construction N_0 V N_1 $à$ N_2, where the direct object N_1 can consist of a clause, but where the indirect object N_2 with preposition $à$ "to" cannot, as in *Max dit qu'il pleut à tout le monde* "Max tells it is raining to everyone". However, the features defining the classes are documented imprecisely, either in text or by formulae of the same type as the labels. They do not appear in the tables: for example, table 9 does not have a '*N0 V N1 à N2*' column; but this feature serves as a reference for the representation of other constructions such as '*N0 V à N2*' (*Max téléphone à tout le monde* "Max phones to everyone"), and for distributional features, such as '*N0 =: N-hum*': the numbering of arguments, here N_0 and N_2, is the link between these features[4]. These conventions seem to have complicated the understanding of the features by users. Gardent *et al*. (2005), for example, wonder whether subscripts refer to the position of the constituent in the basic construction or in another.

To formalize the definition of classes, it was decided to fill in a table of classes, in which features are not assigned to lexical items, but to whole classes (Constant and Tolone, 2010). This work is in progress.

3.4. Delimitation of lexical items

As with any dictionary in the linguistic sense, the basic objects of LG are the lexical items. In case of polysemy, items are separated: different items for *foncer* ("rush" in *Max fonce au port* "Max rushes to the port" and "darken" in *Le pigment fonce les couleurs* "The pigment darkens colors") are distinguished from each other in the same way as *foncer* is from *fonder* "found" (in *Max fonde une agence* "Max founds an agency"). Several constructions can fall within a single item. For instance, the LG does not register a separate item for *Les couleurs foncent* "The colors darken": it ascribes this construction to the same as *Le pigment fonce les couleurs* "The pigment darkens colors", through a feature labelled '*N1 V*' . Some classes are exceptions to this principle. Thus, *Max saupoudre du sel sur les frites* "Max sprinkles salt on the fries" is

---

[4] In the printed version of table 34L0 (Boons *et al.*, 1976a), the numbering of the arguments of certain constructions is independent of the basic construction, and horizontal headings resolve the ambiguity informally. During the remotion of the horizontal headings, the feature labels in this table have been adapted to the conventions of the other tables.

described in class 38LS as being the canonical construction N_0 V N_1 Loc N_2, where Loc denotes a locative preposition. The construction with the non-prepositional location, *Luc saupoudre les frites de sel* "Luc sprinkles the fries with salt", is specified in the same item through the '*N0 V N2 (de N1)*' feature, but it is also described separately and in more detail in class 37M4, with an independent numbering of the arguments. In the future, we want to make these tables homogeneous with others in this regard.

Sometimes, the specification of several constructions within a single item has been badly accepted by users. Thus, Gardent *et al.* (2005), about class 1, generate new lexical items in order to represent a feature corresponding to the replacement of an infinitive complement (*Max commence par examiner Eric* "Max starts by examining Eric") by a human complement (*Max commence par Eric* "Max starts with Eric").

4. Degree of validity

Gardent *et al.* (2006) note that some information contained in the LG may be incorrect. As a matter of fact, several sources of error explain the presence of invalid information.

First, there are material errors. Bugs in table management programs have reversed all the + and - signs in some items, for example[5] *traîner là* "drag along" in class 1. I have myself introduced during the revision of labels (see section 3.1) several errors that I have corrected in 2009.

Then, some support verbs have been encoded in the LG of distributional verbs, as *faire* "do" in *Max fait du tennis* "Max does tennis" or *subir* "undergo" in *Le pétrole subit une hausse* "Oil prices undergo an increase". Since there are tables for those nouns that can play the role of predicate in sentences, and since these tables describe *tennis* or *hausse* "increase", it is necessary to collate these items and exclude items for support verbs from classes of distributional verbs.

Lastly, the authors sought to slightly inflect their judgements of acceptability in the direction of tolerance. Gardent *et al.* (2005) find thus certain constructions rather unlikely to occur. The description of *deviner* "guess", for example, marks as acceptable the construction N_0 *deviner N être Adj* (*Luc devine cette question être cruciale* "Max guesses this question to be crucial", class 6). It should however be noted that *je te devine être capable d'autant de répartie* "I guess you to be able of as much retort" is attested[6] on a blog in a posting of September 2008. The authors of the LG of French distributional verbs did not seek support for their decisions in corpus attestations. It was unrealizable at the time (Boons *et al.*, 1976:37). A more objective validity check would have been heavy and would have compromised the feasibility of the project. Coverage in information was preferred over objectivity. Today, cross-checking facts between the LG and a corpus would be interesting, but it would be

---

[5] Gardent *et al*. (2005) take this item as an example, but do not mention the error.
[6] http://capmetz57.over-blog.com/article-22749572-6.html, December 2, 2009.

unrealistic to try to link up to attestations all the information contained in the LG. This dictionary is a review of the vocabulary (13 000 verbal items, but not all have been coded) crossed with a review of 460 syntactic-semantic features, in which one tests the same constructions on rare items like *godailler* "ruck" as on frequent ones like *bouillir* "boil". A corpus is also a crossed review, but a partial one, with no guarantee that all combinations are tested; it does not attest unacceptabilities. The choice to cover a large amount of information in the LG also justifies in part a moderate degree of formalization (see Section 3): a more complex formal apparatus might have hampered the implementation of so many tests.

The presence of errors in the LG should not overshadow its strengths with regard to validity.

The fact that errors can be detected is in itself a sign of falsifiability of the LG in the epistemological sense: it takes an explicit position on verifiable points.

It is also fairly neutral with respect to various syntactic theories. The authors have focused on relatively verifiable phenomena, that is, those for which observation is more reproducible. They have thus marked the processive vs. stative aspect of some constructions, such as $N_2 \ V \ N_1$, illustrated by *Le rideau cache le sac*, "The curtain hides the bag", which is stative, vs. the basic construction of the same item, *Max cache le sac derrière le rideau* "Max hides the bag behind the curtain", which is processive; but in the case of the construction $N_1 \ V \ Loc \ N_2$, the marking of that semantic feature was not considered sufficiently reproducible to be worth doing systematically: if the aspect is definitely processive in *Le volet claque contre le mur* "The shutter slams against the wall", and definitely stative in *Le carton tient contre la caisse* "The cardboard stays in place against the box", semantic intuition is less clear in *Le frein frotte sur la jante* "The brake rubs the rim". In general, the authors of the LG surround themselves with methodological provisions to ensure the reproducibility of their observations (Laporte, 2009), and resort to intuition is more severely circumscribed than, for example, in Levin (1993), resulting in a more solid empirical basis[7].

It sometimes happens that a theoretical framework has a difficulty in accounting for an observable fact recorded in the LG: this is probably what Hathout and Namer (1997: 5) mean by "certain transformations are linguistically incorrect, within the theoretical framework considered" (HPSG), and exemplify by the construction $N_1 \ se \ V \ auprès \ de \ Nhum \ de \ ce \ Qu \ P$ (*Max se réjouit auprès d'Alice de ce que le film sorte* "Max expresses delight to Alice that the movie comes out")[8]. However, what is incorrect is rather the theoretical framework than the phenomenon. Neutrality with respect to syntactic theories, incidentally, is an additional explanation of the choice of a

---

[7] Beth Levin nevertheless knew the work of Boons *et al*. (1976) (personal communication).

[8] This construction with three arguments, which denotes a speech act, coexists with a two-argument construction (*Que le film sorte réjouit Max* "That the movie comes out delights Max"); in the latter, it cannot always be assumed that the third argument, which is formally absent, is in fact semantically present. Such a situation is an anomaly for most current theories.

moderate degree of formalization. A more complex formalism, necessarily more dependent on a theory, might have obstructed the possible observation of facts to which this theory would not have been adapted.

5. Information content

Gardent *et al.* (2006 : 145-146) note that some information is missing from LG or incomplete, such as grammatical functions and thematic roles, whereas other features, which are generally not used by parsers and generators, are present, e.g. the temporal interpretation of infinitival clauses. Is the information content of the LG sufficient for the needs of a parser? How does it compare to other structured lexicons?

Not all grammatical functions are coded, because they cover syntactic and semantic features, which are generally more factual, and which they are partly redundant with. Thus, the notion of direct object is based on various features which do not always coincide: position after the verb, no preposition, pronominalization, passivation (Gross, 1969)... The LG encodes these features rather than grammatical functions; this choice makes the content more accurate, and the main grammatical functions can be automatically deduced from such features (Tolone and Sagot, 2009). In particular, the authors of the LG of verbs have played a pioneering role in defining the actual distinction between objects (essential complements) and adjuncts (modifiers, circumstantial complements). Thus, they described as an argument the direct object of verbs of class 32NM (*Max chausse une grande taille* "Max fits a big size (for shoes)", *La pièce sent le jasmin* "The room smells of jasmine"), often regarded as adjuncts. The same occurs for the indirect complement of many locative verbs (*Max place sa voiture contre le mur* "Max parks his car against the wall"). They also inventoried many types of complements with an intermediate behaviour between those of an argument and of an adjunct, for example *sur ce point* "on this" in *Max se ravise sur ce point* "Max changes his mind on this".

With regard to thematic roles and more generally the formalization of meaning, the authors of the LG limited themselves to phenomena for which they could circumscribe observation with syntactic tests (see Section 4).

One could cite other gaps in the information content of the LG: some items have not been encoded yet, e.g. *basculer* "totter" and *boiter* "limp" (Fig. 1); constructions with a regular formation, e.g. negation or relative clauses, have been neglected unless they vary depending on lexical items; some constructions, including pronominal ones, are not encoded; tables of adjectives are under construction (Giry-Schneider and Laporte, 2011)... All this information is certainly essential for the proper operation of parsers based on lexicons and grammars.

In spite of these limitations, it is difficult to dispute the richness of the information in the LG, as compared with other structured lexicons. The traversal of the lexicon and the inventory of constructions are impressive. The systematic demarcation between

frozen vs. free constructions is difficult to find elsewhere, except in Lexicon-Grammars of other languages.

Would such advances have been possible with other methodological options?

Conclusion

The idea that the Lexicon-Grammar is difficult to use for language processing stems in part from the presence of errors and shortcomings, which can be corrected, but also from a feeling of strangeness on the part of parsing experts at choices that are uncommon in most other projects they are aware of. Taking into account the different aspects of the problem leads to refine and moderate this view, and to justify most of these choices by the original features of the Lexicon-Grammar: a comprehensive inventory of the vocabulary and constructions; primacy of factual data over constraints related to specific theories; a requirement of reproducibility of observations. Now these very features are precisely those that open up opportunities for exploitation of the Lexicon-Grammar in language processing systems.

References

ALCOUFFE PH., REVELLIN-FALCOZ B., ZAYSSER L. (1993), « Azote : des tables du LADL au format Genelex », *Actes du Colloque Informatiques et Langues naturelles*, IRIN, Université de Nantes.

BLANC O. (2006), *Algorithmes d'analyse syntaxique par grammaires lexicalisées : optimisation et traitement de l'ambiguité*, Thèse de doctorat, I.G.M., Université de Marne-la-Vallée.

BOONS J.P., GUILLET A., LECLERE CH. (1976a), *La structure des phrases simples en français. 1 : Constructions intransitives*. Genève : Droz.

BOONS J.P., GUILLET A., LECLERE CH. (1976b), *La structure des phrases simples en français. Classes de constructions transitives*. Rapport de recherche du LADL n° 6, Université Paris 7.

CONSTANT M., TOLONE E. (2010), "A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables", M. De Gioia (Ed.), *Actes du 27e Colloque international sur le lexique et la grammaire (L'Aquila, 10-13 septembre 2008). Seconde partie,* Rome: Aracne, Lingue d'Europa e del Mediterraneo, Grammatica comparata, 1, pp. 79–93.

DENDIEN J., PIERREL J.M. (2003), « Le Trésor de la Langue Française informatisé : un exemple d'informatisation d'un dictionnaire de langue de référence », dans *TAL* 44(2) : 11-37.

FILLMORE CH., ATKINS S. (1994), "Starting where the dictionaries stop: The challenge for computational lexicography", S. Atkins, A. Zampolli (Éds.), *Computational Approaches to the Lexicon*. Oxford University Press : 349-393.

FRANCOPOULO G., GEORGE M., CALZOLARI N., MONACHINI M., BEL N., PET M., SORIA C. (2006), "Lexical Markup Framework (LMF)", *Proceedings of LREC, Genoa* : 233-236.

GARDENT C., GUILLAUME B., PERRIER G., FALK I. (2005), "Extracting subcategorisation information from Maurice Gross' grammar lexicon", dans *Archives of Control Sciences* 2005 : 289-300.

GARDENT C., GUILLAUME B., PERRIER G., FALK I. (2005), « Le lexique-grammaire de M. Gross et le traitement automatique des langues », ATALA Workshop : Interface lexique-grammaire et lexiques syntaxiques et sémantiques.

GARDENT C., GUILLAUME B., PERRIER G., FALK I. (2006), « Extraction d'information de sous-catégorisation à partir des tables du LADL », *Verbum ex machina*, Actes de TALN, Collection Cahiers du Cental numéro 2, volume 1 : 139-148, Presses universitaires de Louvain.

GIRY-SCHNEIDER J., LAPORTE É. (2011), « Classer et décrire les adjectifs du français », *Cahiers de lexicologie* 98, Mélanges en hommage à Hong Chai-song, Paris : Gallimard.

GRISHMAN R., MACLEOD C., MEYERS A. (1994), "COMLEX Syntax: Building a Computational Lexicon", *Proceedings of Coling* : 268-272.

GROSS M. (1969),. « Remarques sur la notion d'objet direct en français ». Dans *Langue Française* 1, : 63-73.

GROSS M. (1975), *Méthodes en syntaxe*, Paris, Hermann.

GUILLET A., LECLERE CH. (1992), *La structure des phrases simples en français. 2 : Constructions transitives locatives*. Genève, Droz.

HATHOUT N., NAMER F., (1997), "(Semi-)Automatic Generation of the ALEP Analysis Lexicon", *Proceedings of the 3rd ALEP User Group Workshop*, Saarbrücken.

HATHOUT N., NAMER F., (1997), « Génération (semi)-automatique de ressources lexicales réutilisables à grande échelle. Conversion des tables du LADL », *Actes des 1ères JST FRANCIL*, AUPELF-UREF, Avignon.

HATHOUT N., NAMER F. (1998), "Automatic Construction and Validation of French Large Lexical Resources: Reuse of Verb Theoretical Linguistic Descriptions", *Proceedings of LREC* : 627-636.

LAPORTE É. (2009), "Lexicons and grammars for language processing: industrial or handcrafted products?", L.M. Rezende, B.C. Dias da Silva, J.B. Barbosa (Eds.), *Léxico e gramática: dos sentidos à construção da significação (Lexicon and Grammar: from Meanings to the Construction of Signification)*, São Paulo: Cultura Acadêmica, Trilhas Lingüísticas, 16, pp. 51–84.

LEVIN B. (1993), *English Verb Classes and Alternations. A Preliminary Investigation*, The University of Chicago Press.

PAUMIER, S. (2001), "Some remarks on the application of a lexicon-grammar", *Lingvisticae Investigationes* 24(2) : 245-256.

ROCHE E. (1999), "Finite state transducers: Parsing free and frozen sentences", A. Kornai (Éd.), *Extended finite state models of language*, Cambridge University Press : 108-120.

TOLONE E., SAGOT B. (2009), "Using Lexicon-Grammar tables for French verbs in a large-coverage parser", Z. Vetulani (Ed.), *4th Language & Technology Conference* (LTC'09), Poznań, pp. 200–204.

TOLONE E. (2011), *Analyse syntaxique à l'aide des tables du Lexique-Grammaire du français*. PhD dissertation, Université Paris-Est, 332 p.