



HAL
open science

Lexicon-Based Algorithms for the Automatic Analysis of Natural Language

Maurice Gross

► **To cite this version:**

Maurice Gross. Lexicon-Based Algorithms for the Automatic Analysis of Natural Language. *Theorie und Praxis des Lexikons. Proceeding of the Conference on Theories and Applications of Lexicology and Lexicography*, Walter de Gruyter, pp.218-236, 1993. halshs-00278305

HAL Id: halshs-00278305

<https://shs.hal.science/halshs-00278305>

Submitted on 11 May 2008

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.

LEXICON BASED ALGORITHMS FOR THE AUTOMATIC ANALYSIS
OF
NATURAL LANGUAGE

Maurice GROSS

Laboratoire d'Automatique Documentaire et Linguistique_

University of Paris 7

Let us examine the following discourse (D) from the point of view of elementary grammatical analysis:

(D) Two men cleaned the offices, then, they waited for the janitor

This discourse is composed of two members: two simple sentences connected by the conjunction then. One of the elements needed for the interpretation of (D) lies in the nature of the antecedent of the pronoun they. In principle, the pronoun they refers to the noun phrase two men of the first member, but it might indicate a group of persons different from these two men, if (D) is attached to an appropriate context or background. Whether the scene_ which constitutes the interpretation of (D) includes 3 persons (2 men and 1 janitor) or more depends entirely on the analysis of they.

Such questions of resolution of pronouns are trivial for a native speaker of English, but they become of paramount importance when one attempts a computer analysis of texts, and also when a reader who does not know well the language in which the discourse (D) is written tries to understand it. In both situations, in order to interpret (D), detailed dictionaries and grammars must be available which account for the relations occurring between the terms of (D). In this article, we are going to simulate the computation of the search for the antecedent(s) of they. We will simplify this procedure by omitting its clerical aspects. In this way, we throw into relief the nature and the amount of information that must be stored in a lexicon-grammar, since, as we will see, we do not draw the usual line of demarcation between these two components of a language.

1. DICTIONARY LOOK UP

Each word of (D) is identified by look up in a dictionary which provides the different meanings of each word and the grammatical properties attached to each of these meaning. Thus, the word form they is described as a plural definite pronoun, either human or non-human on the one hand, and indefinite human on the other_, as in:

They (= the Government) are wasting our money !

with respect to (D), this second interpretation is not essentially different from the interpretation where they refers to people outside of (D); we shall disregard this case.

Two common interpretations of they are then possible:

- (1) they refers to two men,
- (2) they refers to offices.

The reader, however, does not feel that (D) is ambiguous in these two ways; in other words, only interpretation (1) is possible. Therefore, we have to exclude (2), by showing that the interpretation they = the offices is not compatible with the structures and word meanings of (D).

We can accomplish this, even before studying grammatical compatibilities, by consulting the dictionary for the word form office. We have listed in the annex II the entries, that is, the common meanings, of the word office.

We use the following terminology: we say that there exist 6 words office, or 6 entries for the form office. The implication of such a statement is that there is no synchronic relation between the 6 entries, unless otherwise stated_. We will avoid talking about the various "meanings or uses" of a given word or form or morpheme, for the reason that this terminology is too suggestive of implicit relations between the various meanings_. For example, one could contrast the entries 1) and 2) on the basis that they are respectively the proper and the figurative meanings of the word office. In such situations, we can justify setting up two independent entries by the numerous selectional restrictions that make the two meanings differ (Z.S. Harris 1957); we can see this in sentences such as:

- Bob locked the office (proper)
- Bob shut down the office (figurative)

which are not ambiguous since each noun office selects its own set of verbs. Generally speaking, the syntactic properties of the two entries are different.

The proposal that dictionary entries limited to single words be characterized by their contexts is a classical idea that we have approached from a strictly syntactic point of view:

- for verbs, the formalization of the proposal is clear: a verb cannot be isolated from its subjects and some essential complements. The only difficulty is separating the object complements from the adverbial ones, which are roughly the non essential complements. Hence, a verb form will have as many entries as it has syntactic structures that cannot be related by transformational relations. In the case of to wait described in 4.2, we have the following elementary constructions:

- (1) Max waited for the Ambassador
- (2) Max waited on the Ambassador

(3) Max waited upon the Ambassador

These three sentence types correspond to three meanings or uses of the verb form, hence they constitute three independent entries of the lexicon,

- adjectives combined with to be are not essentially different from verbs,

- nouns raise various problems with respect to their syntactic characterization.

Important classes of nouns are now being characterized by means of a support verb (J. Giry-Schneider 1978, 1987; M. Gross 1981; G. Gross 1989; G. Gross et R. Vivès 1986, J. Labelle 1974, A. Meunier 1977, D. de Negrone 1978). A support verb is an auxiliary verb that carries no primary information i.e. no selectional restrictions, in an elementary sentence, whereas the noun that it supports is the predicative element. This property is obvious in nominalization or adjectivization relations such as:

Max does not fear such an event
= Max has no fear of such an event
= Max (was, went) in fear
= Max is fearful of such an event

Max looked at the park
= Max took a look at the park

The surgeon operated on Max
= The surgeon performed an operation on Max

Standalone nouns can be supported by the same verbs, as in:

Max has a clear stand on such events
Max has the creeps
Max is in trouble, etc.

Support verbs have syntactic properties that distinguish them from verbs taking the predicative role in the simple sentence. Let us compare the two sentences:

- (1) Max took a firm stand on the incident
- (2) Max recommended a firm stand on the incident

Clefting may be applied to the support construction (1) but not to (2):

It is on the incident that Max took a firm stand
*It is on the incident that Max recommended a firm stand

Passive applies differently to (1) and (2):

A firm stand was taken by Max on the incident
*A firm stand was recommended by Max on the incident

Sentence (1) is considered as elementary. Sentence (2) will be analyzed as containing sentence (1) in nominalized form in its object. This is one of the many

reasons why nominal entries are described in terms of support verbs, rather than by semantically well chosen sentences, as in current dictionaries;

- adverbs can also be described in elementary sentences with specific support verbs (M. Gross 1991; Z.S. Harris 1976), as in:

The accident (occurred took place) (at 3 o'clock, in the office)

Thus, all major parts of speech, that is the productive vocabulary is to be described syntactically in a lexicon-grammar.

Returning to the form office, we have three clear cases of characterization by means of support verbs: entries 4), 5) and 6), whereas entries 1), 2) and 3) are given in examples that are more semantic descriptions than syntactic characterizations. Nonetheless, the semantic descriptions 1)-3) are given in terms of elementary sentences which contain general support verbs (to be, to have).

2. EXPLOITING DICTIONARY INFORMATION

To answer the question we have raised about (D), we have to check the incompatibility of they = offices in the subject position of to wait. We thus have to look up the dictionary entry for wait which is given in 4.2. In 4.2 the description of wait is made independently of the example of parsing under discussion. Hence, from now on, the reader must refer to 4.2.

Since wait in (D) occurs with the tense suffix -ed, it has to be a verb. Hence, the entries of the noun wait given in 4.2 are not relevant to the analysis of (D). Neither are the entries 7) (keep waiting), 10) (NO WAITING) and 11) (Waiting game) relevant, since they obligatorily take the ending -ing. In the same way, the following entries cannot take the suffix -ed:

- entry 6):

*Max (cannot, could not) have waited for the results

- entry 8), - ed is presumably not allowed:

?I waited and saw

?I have waited and (have) seen

- entry 9):

*Just (E, you) waited !

- entry 12):

*While (you waited, U-waited)

We can exclude entries 4) and 5) since the particles that must accompany wait are not observed in (D). The same procedure excludes entries 2) and 3) which contain

prepositions not found in (D). The only possibility left in (D) is entry 1) and the possibility is confirmed by the presence in (D) of the preposition for which is explicitly allowed in entry 1).

However, from the information given in entry 1), we see that wait for can have both human and non human subjects. Thus, the elimination process we have just applied does not quite resolve the ambiguity of they, although it does make the status of wait considerably clearer.

At this point, we want to emphasize that the explicit process that we are presenting here must make use of these mostly negative observations. Negative information is necessary to determine the interpretation of (D), since (D) depends not only on the specification of they, but as well on the determination of the entries compatible with each other in (D).

Let us examine more closely the hypothesis where they is non human, that is, they = offices. The entry of the office is described in 4.3. Again, the reader will have to refer to this section to find the definitions of the various entries. The possible occurrence of compound nouns can be eliminated by checking the context). We thus have to check which of the 6 various nouns office can be found in (D). Two sets of constraints can be used to determine which entries are compatible with (D):

- office is the object of to clean,
- office (i.e. they) is the subject of to wait.

Notice that the two operations of checking verb-noun compatibilities are independent, hence they can be performed in any order, or in parallel computations.

If we were to look up the entry for to clean (not given here), we would find that it can only have a concrete object. Hence, we can only select entry 1) of the word form office.

If we look up the entry for wait, we see that no entry of office can be the subject of to wait. Hence, the ambiguity is resolved: they cannot refer to offices, and (D) is not ambiguous: they = two men.

Let us now modify (D) slightly and replace to clean by to paint:

(D1) Two men painted the offices, then, they waited for the janitor

To paint differs from to clean in that it can be followed by a human object. Then the entry 3) of office has to be considered and this entry can be the subject of wait entry 1). Hence, discourse (D1) is ambiguous. Although this is not apparent from reading (D1), a simple paraphrase may make this point more clear. Let us replace to paint someone by to take a picture of someone in (D1). Notice also that one may, in some contexts, orient the interpretation of office towards entry 3) by adjoining whole, as in the whole office. We obtain:

(D2) Two men took pictures of all the whole offices, then, they waited for the janitor

It now seems clear that the interpretation they = the people in the office has to be considered.

Consider the following discourse where to inspect has replaced to clean in (D):

(D3) Two men inspected the offices, then they waited for the janitor

The entries of inspect are such that human or non human objects are authorized:

- (1) Max inspected the walls of the building
- (2) Max inspected the troops

In this case, (D3) is ambiguous in the same way as (D2).

Let us now consider another modification of (D), in which office is in the singular:

(D4) Two men cleaned the office, then they waited of the janitor

The problem becomes trivial, since the phrase two men is the only candidate for the antecedent of they. However, let us put janitor in the plural:

(D5) Two men cleaned the office, then they waited for the janitors

janitors now becomes a possible antecedent for they. The reason for which this interpretation is not possible is a grammatical rule that forbids forward reflexivization in English, in principle, this is not a lexical rule.

The tenses of the verbs in the discourses (D) are all in the preterit. But if (D) contains the progressive form to be waiting for example, then new lexical candidates will have to be checked and new rules will have to be applied.

Remark

We considered that office-2) (the abstract sense) could not be the subject of wait-1), and this is not entirely clear: it depends on the acceptability of such a sentence as:

These offices wait for this inspectors to check their accounting procedures

In fact, the problem is even more difficult. Consider the sentence form where office has the interpretation 1):

The offices are waiting for the painters to whitewash their walls

If this sentence is acceptable, then the representation of entry 1) that we have given for office will have to be made more precise. This, in turn, means that we authorize restricted non human subjects for wait-1). How are we going to distinguish those non human nouns which are acceptable as subjects and those which are not? The problem is open and only a thorough distributional study of wait-1) can lead to an answer. It will be noted that such distributional studies where noun-verb combinations are systematically considered have never been performed yet.

The modifications of (D) we have just considered should make it clear that the process here described is extremely unstable in the sense that if we change a discourse slightly, by introducing synonyms for example, the whole procedure for the identification of pronouns changes. It can suddenly become quite complex or quite simple depending on occurrences of forms whose corresponding number of lexical entries is unpredictable.

3. REQUIREMENTS FOR ANALYSIS

We can see from this type of discussion that a full description of the lexicon and of its grammar is needed. The various discourses (D) are in fact quite simple from a structural point of view: (D) is a conjunction of two elementary active sentences.

The active sentences that compose (D) could include adverbs, they could undergo transformations (Passive, Clefting, etc.), as in the discourses:

There were two men who cleaned the offices, it was they who waited for the janitor

The janitor was waited for by two men. The offices had just been cleaned by them

The information needed to recognize these forms do not appear in the list of elementary forms given in 4.2 and 4.3. A full listing of the syntactic variations allowed for each elementary sentence (i.e. each verb) has to be supplemented as a systematic grammar. This grammar must tell whether a given verb has a Passive form or not, if yes, whether its preposition or particle can be stranded or not, etc. A great deal of this information is lexical and has to be represented in the form of lexical tables (cf. 4.4). We insist on an essential feature of these tables: they contain negative information, that is, the type of information that allowed us to reject undesirable analyses of (D).

Let us reverse our point of view and instead of focusing on the lexicon let us consider the nature of texts submitted to automatic parsing. Practically nothing can be said a priori about the lexical nature of texts. Statistical analysis applied to occurrence of words in texts are performed on word forms. To be meaningful for parsing, statistics should be carried out on dictionaries entries separated the way they are in the examples of 4.2, 4.3 and 4.4. In order to perform these improved statistics, entries have to be recognized beforehand in texts, either by hand or by computer. In order words, the problem of parsing must be solved before statistics can be performed.

Practically nothing is known about the syntactic nature of a text, since to a large extent syntactic shapes are lexically determined.

Under such conditions of indetermination, the complexity of the analysis of a text is totally unpredictable. To face this situation, there is no other solution than to build the extensive lexical and grammatical data base that we will now outline. Such a lexicon-grammar of a language is the only rational point of departure for text analysis.

4. DICTIONARIES AND LEXICON-GRAMMARS

4.1. Notations

We note elementary sentential structures as follows:

$$N0 \ V \ W$$

where $N0$ is the subject, V the verb and W a variable ranging over possible complements. To specify a structure or a class, we use the symbol "=", as in the formulas:

$$\begin{aligned} W &=: N1 \text{ Prep } N2 \\ N0 &=: N_{\text{hum}}, N\text{-hum} \end{aligned}$$

which mean:

- W is a string of 2 complements, the first, $N1$, is direct, the second, $N2$, is prepositional
- the subject $N0$ is either a human or a non human noun.

In the formula:

$$N0 \ V \ N1 \text{ to } V1 \ W1 =:$$

the superscript 1 attached to $V1$ means that $N1$ is the subject of $V1$; W is the complement sequence of $V1$.

The symbol "=" is used between two sentences to indicate that a transformational relation holds.

We give here an outline of the description of the two word forms: wait and office. This outline is intended to show how a systematic separation of entries can be achieved on SYNTACTIC grounds_. The list of entries will have to be refined further. For example, a sentence such as:

Bob is waiting in line

appears to be difficult to relate to any of the entries of wait, because of the quasi idiomatic combination with the adverb in line. The form:

$$N0 \ \text{wait} \ \text{in line}$$

may then have to be added to the lexicon-grammar as a new entry, with the implication that it is unrelated to the other entries.

A few examples of compound nouns have been given for various word forms. The lists should be completed and each compound should be represented with its support

verb(s).

The formal entries are those which include variables for their attached noun phrases (i.e. Ni). We provide actual examples for these abstract forms. Comments about the motivations for separating entries are also given.

4.2. The entry wait

VERBS

- 1) N0 wait for N1 to V1 W =:
Max is waiting for Bob to bring the results
The car is waiting for Bob to pick it up
Max will wait (till, until) Bob leaves
The trip must wait (till, until) the car is ready

With non human subjects, there seems to be an obligatory modality:

*waits, *waited
The trip will have to wait (till, until) the car is ready
has to wait

The sentential complements for N1 to V1 W and (till, until) P are in complementary distributions, which justifies a single entry.

Max is waiting for (Bob, the results)

The nominal complements for N1 =: for (Bob, the results) are considered a reduction of the for--to construction.

- 2) N0 wait on N1 =:
Max waited on the guests
= Maw was on waiting on the guests
- 3) N0 wait upon N1 =:
Bob waited upon the Ambassador

Entries 2) and 3) appear to be synonymous:

Bob waited on the Ambassador
= Bob waited upon the Ambassador

Max waited upon the guests
= Max waited on the guests

If this relation proves to be general:

Maw was upon waiting on the guests
= Maw was on waiting on the guests

ADJECTIVES

- 11) N0 play a waiting game

ADVERBS

- 12) While you wait
While U wait

NOUNS

- 13) There is a two hour wait
= The wait is two hours long
= The wait lasts two hours
- 14) There is a wait between the two acts
- 15) N0 (be, lie) in wait for N1 =:
The hunter (is, lies) in wait for his prey
The reporter (is, lies) in wait for the President

These two examples may in fact correspond to two different entries. We face here a problem which arises frequently: what is the relation between the proper and the figurative meaning of a given word form? This question cannot be answered a priori, that is, before a rather complete study of the lexicon has been carried out. Such a study of French has shown that in certain cases there are rules (e.g. the rule of Fusion, cf. entry 5)) which can establish a link between the two meanings, but in many cases similar to the pair under discussion, one has to set up two independent entries. The explanation of the similarity between the meanings and the syntactic structures then has to be etymological.}

COMPOUND NOUNS

- 16) N0 is on a waiting list =:
Bob is on a waiting list
- 17) N0 is in the waiting room =:
Bob is in the waiting room

4.3. The entry office

NOUNS

- 1) N0 have an office =:
Bob has an office

A physical description of an office can be given in terms of elementary sentences such as the following ones:

There is a desk in Bob's office, etc.

Bob's office has two windows, etc.
The office is 10 feet wide
The office is whitewashed

- 2) N0 (be, work) in an office =:
Bob (is, works) in an office
Bob's office deals with cash sales

Here the meaning of office is abstract as opposed to the concrete meaning of entry
1) The difference is easily perceived in the sentences:

Bob's locked his office
Bob's shut down his office

Semantic specifications of this entry correspond to such sentences as:

Bob's office hired ten clerks
Bob's office makes money

Since the specifications given above for the concrete meaning are incompatible with these specifications for the abstract entry 2) of office, interdictions such as the following one can be used to reduce potential ambiguities:

*Bob's whitewashed office deals with cash sales

- 3) The (office, whole, office) celebrated Bob's promotion

This entry may be derived from entries 1) or 2) by means of a general rule which states that a non human container noun of humans can be used (metonymically) for the contained humans.

- 4) N0 (hold, take) an office =:
Bob holds the office of treasurer

N0 (be in, come into, go out of) office =:
Max (is in, went out of) office

These sentences contain support verbs that are synonymous to within aspectual differences, which is one of their regular features. The sentence:

Max will suppress Bob's office of treasurer

is a complex sentence in which the verb has for its object a reduction of a sentence with support verb.

Abstract offices can also be described by means of analytical or definitory sentences such as:

Treasurer is an office
Dealing with protests is an office

- 5) N0 propose his offices to N1 =:
 The lawyer proposed his good offices to Bob
 Max used the lawyer's offices

This entry may, in fact, be a compound noun with obligatory plural: good offices.

- 6) N0 performed an office =:
 The priest performed an office for the dead

COMPOUND NOUNS

lawyer's office, doctor's office, etc.
 box office, post office, etc.
 divine office, etc.
 office boy, office building, office hours, office job, etc.

4.4. Lexicon-Grammars

The following tables represent the syntactic variations that are allowed for certain verb entries. Each row corresponds to an entry in the sense of 4.1 and each column to a structure in which the entry is acceptable or not. The sign "+" or "-" is placed at each intersection according to the acceptability.

We first give an example of an English syntactic table taken from a study by M. Salkoff 1983. The construction of this table was motivated by a lexical study of the transformational relation:

N0 V Prep N1 = N1 V Prep N0 =:
 Bees are swarming in the garden
 = The garden is swarming with bees

The other table is a sample of the French lexicon-grammar (from M. Gross 1975), it is the table numbered 6, which is an arbitrary code standing for the structure:

N0 V N1 with N1 =: Qu S
 =: Max attend que Bob arrive
 (Max is waiting for Bob to arrive)

namely, the structure subject-verb-object with sentential object introduced by the subordinating conjunction que.

Since the publication of these tables, the French lexicon-grammar has been substantially revised and expanded. It now contains near 12 000 simple verb entries. In earlier versions of the lexicon-grammar, we had included a sample of complex entries such as avoir en tête:

Max a en tête que Bob va arriver
 (Max has in mind that Bob will arrive)

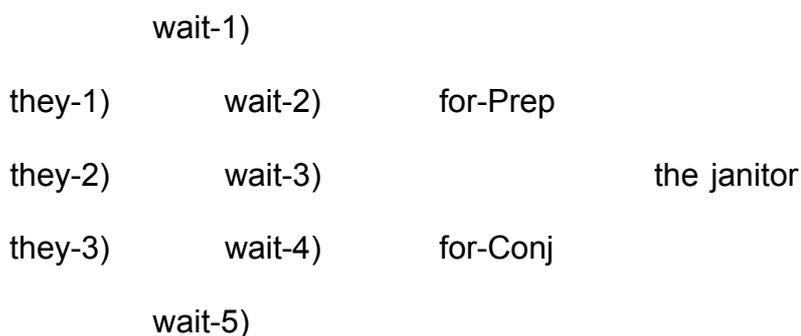
Such complex verbs or idiomatic sentences are now described in separate sections of the lexicon-grammar. It includes about 50 000 such complex or idiomatic entries and has not yet reached the level of completeness of the lexicon-grammar of the 12 000 simple verbs (M. Gross 1982). These two figures should give pause to theoreticians who have a tendency to think about idiomatic forms as exceptions. It is true that the birth of a compound or of an idiom is irregular: etymological dictionaries of idioms contain as many anecdotes as they provide idioms_. But the structure and variability of idioms appear to be more general than the properties of the so-called normal sentences that are usually studied. This observation has implications both for the representation of linguistic phenomena and for the way they are put to use, for example in a parsing procedure, as the one we have outlined here.

5 CONCLUSION

It is now interesting to compare the analysis we outlined here with the common approach to parsing.

Currently, a model of grammar as general as possible is used to represent syntactic phenomena (e.g. String Grammar, TAG, GPSG). Such a grammar is construed as a function that maps a sentence of a text, that is a string of concatenated words, to a syntactic structure that will allow semantic interpretation. Our main point here is that the function is global in that it does not separate phenomena such as the elimination of a lexical ambiguity and the characterization of the right boundary of a relative clause.

Our approach distinguishes an elaborate lexical level. Hence, the result of the look-up procedure for the words of a given sentence can be represented in the form of an automaton (Directed Acyclic Graph) which is a natural generalization of the notion of string (E. Roche 1992, M. Silberztein 1990). For example, for the sentence member of (D), we have the following graph:



where the lexical ambiguities previously discussed are represented by parallel edges; the figures refer to the numbering of the separate entries (cf. note 3 and 4.2).

A general parser has to cope with $3 \times 5 \times 2 = 30$ paths of the graph, attempting to eliminate 29 of them.

Our attitude is different, we consider that we have various lexical ambiguities to suppress:

2 for they, 4 for wait and 1 for for

namely, that we have $2 + 4 + 1 = 7$ possibilities to remove. This view is not merely an equivalent, although additive, way of counting but it implies a clear separation between the many heterogeneous functions performed by a general parser.

As a consequence, we will for example apply local rules, sometimes ad hoc rules, to solve ambiguities (i.e. to remove edges from the automaton of the text). Applying local grammars does not directly provide syntactic structures which will have to be determined separately, possibly using the same information that is used for removing ambiguities.

We think that this presentation of the parsing process clarifies the role of the lexicon and could be the basis for the construction of new parsers, more specific, hence more efficient. In to-day's parsers, there are numerous situations where it is not clear whether a parser applies a grammar rule or a computation rule. Grammar rules are supposed to be neutral between analysis and synthesis of sentences but most parsers so far have merged the grammatical and the recognition processes to a point where they become undistinguishable. We think that progress in parsing requires clarification of such issues and more urgently, examination of the role of lexical information.

REFERENCES

Boons, Jean-Paul, Alain Guillet, Christian Leclère. 1976a. La structure des phrases simples en français. Constructions intransitives, Geneva : Droz, 377 p.

Boons, Jean-Paul, Alain Guillet, Christian Leclère. 1976b La structure des phrases simples en français. III Classes de constructions transitives, Rapport de recherches du LADL. University of Paris 7, 143 p.

Courtois, Blandine 1990. Un système de dictionnaires électroniques pour les mots simples du français, Langue française No 87, Paris : Larousse.

Giry-Schneider, Jacqueline. 1978. Les nominalisations en français. L'opérateur FAIRE dans le lexique, Geneva : Droz. 353 p., tables : 61p.

Giry-Schneider, Jacqueline 1987. Etude de prédicats nominaux en français. Les constructions faire N, Genève : Droz, 399 p.

Gross, Gaston, Robert Vivès eds. 1986. Les verbes supports en français. Langue Française, Paris : Larousse.

Gross, Gaston 1989. Les constructions converses du français, Genève : Droz, 513 p.

Gross, Maurice, 1975. Méthodes en syntaxe, Paris : Hermann, 414 p.

Gross, Maurice, 1981. Les bases empiriques de la notion de prédicat sémantique, *Langages* 63, Paris, : Larousse, pp. 7-52.

Gross, Maurice, 1982. Une classification des phrases "figées" du français. *Revue québécoise de linguistique*, vol. 11, No 2, pp. 151-185.

Gross, Maurice, 1991. *Grammaire transformationnelle du français. 3. Syntaxe de l'adverbe*, Paris : ASSTRIL, 670 p.

Guillet, Alain ; Christian Leclère 1992. *La structure des phrases simples en français. Verbes à complément direct et complément locatif*, Genève : Droz.

Harris, Zellig 1976. *Notes du cours de syntaxe*, Paris : Le Seuil.

Harris, Zellig 1957. Cooccurrence and Transformations in Linguistic Structure, *Language* 33, Baltimore : The Waverly Press, pp 283-340.

Labelle, Jacques 1974. *Etude de constructions avec opérateur AVOIR (nominalisations et extensions)*, Université Paris VIII : LADL, thèse de doctorat, 315 p.

Meunier, Annie 1977. Sur les bases syntaxiques de la morphologie dérivationnelle, *Linguisticae Investigationes*, Vol. I, No 2, Amsterdam : J. Benjamins B.V., pp. 287-332.

Négroni-Peyre, Dominique de 1978. Nominalisation par être en et réflexivation (admiration, opposition, révolte et rage), *Linguisticae Investigationes*, Vol. II, No 1, Amsterdam : J. Benjamins B.V., pp. 127-164.

Roche, Emmanuel 1992. *Automates des phrases simples du français*, *Linguisticae Investigationes*, Amsterdam : J. Benjamins B.V.

Salkoff, Morris 1979. *Analyse syntaxique du français : Grammaire en chaîne*, Amsterdam-Philadelphie : J. Benjamins Pub. Co., 334 p.

Salkoff, Morris 1983. Bees are swarming in the garden: a synchronic study of productivity, *Language*, Vol.59, No 2, Baltimore : The Waverly Press, pp. 288-346.

Silberztein, Max 1989. *Dictionnaires électroniques et analyse lexicale automatique*, thèse de doctorat, Université Paris 7, Paris : LADL.

Silberztein, Max 1990. *Le dictionnaire électronique des mots composés*, *Langue française* No 87, Paris : Larousse, pp. 71-83.

—
_. CNRS, Institut Blaise Pascal. I am indebted to Morris Salkoff for substantial improvements of my initial text.

_. Or "frame" or "scenario" in the language of Artificial Intelligence.

_. One then counts three meanings for they.

_. The reader will find in 4.3, a relation that might reduce entry 6) of office to entry 1) and/or 2). There may exist a similar relation between entries 1) and 2), but it is not clear whether it is synchronic or diachronic (i.e. etymological).

_. The display of entries in commercial dictionaries often has the effect of introducing implicit relations between subentries. Such implicit structures reflect in part a mixed bag of the author's intuitions: that is, intuitions of form, of meaning and of etymology, without separations among these based on theoretical principles.

_. Morphological descriptions of simple words are given in a specialized electronic dictionary (DELA, B. Courtois 1990), compound words are also described in a special component (DELAC, M. Silberztein 1989, 1990).

_. Other sections of the French lexicon-grammar have been published: cf. J-P. Boons, A. Guillet, C. Leclère 1976a, 1976b, A. Guillet and C. Leclère 1992 for verbs, J. Giry-Schneider 1978, 1987 and G. Gross 1989 for nouns.

_. Notice that there is no essential difference between the ways non productive metaphors, technical terms and idioms are created.