LEXICON-GRAMMAR AND THE SYNTACTIC ANALYSIS OF FRENCH

Maurice Gross

Laboratoire d'Automatique Documentaire et Linguistique

University of Paris 7

2 place Jussieu

75251 Paris CEDEX 05

France

ABSTRACT

A lexicon-grammar is constituted of the elementary sentences of a language. Instead of considering words as basic syntactic units to which grammatical information is attached, we use simple sentences (subject-verb-objects) as dictionary entries. Hence, a full dictionary item is a simple sentence with a description of the corresponding distributional and transformational properties.

The systematic study of French has led to an organization of its lexicon-grammar based on three main components:

- the lexicon-grammar of free sentences, that is, of sentences whose verb imposes selectional restrictions on its subject and complements (e.g. *to fall, to eat, to watch*),
- the lexicon-grammar of frozen or idiomatic expressions (e.g. *N takes N into account, N raises a question*,
- the lexicon-grammar of support verbs. These verbs do not have the common selectional restrictions, but more complex dependencies between subject and complement (e.g. *to have, to make* in *N has an impact on N, N makes a certain impression on N*)

These three components interact in specific ways. We present the structure of the lexicon-grammar built for French and we discuss its algorithmic implications for parsing.

The construction of a lexicon-grammar of French has led to an accumulation of linguistic information that should significantly bear on the procedures of automatic analysis of natural languages. We shall present the structure of a lexicon-grammar built for French <2> and will discuss its algorithmic main implications.

1. VERBS

The syntactic properties of French verbs have been limited in terms of the size of sentences, that is, by restricting the type of complements to object complements. We considered 3 main types of objects: direct, and with prepositions *à* and *de*. Verbs have been selected from current dictionaries according to the reproducibility of the syntactic judgments carried out on them by a team of linguists. A set of about 10,000 verbs has thus been studied.

The properties systematically studied for each verb are the standard ones:

- distributional properties, such as human or non human nouns, and their pronominal shapes (definite, relative, interrogative pronouns <3>, clitics), possibility of sentential subjects and complements *que S* (that S), *si S* (whether S, if S) or reduced infinitive forms noted *V Comp*,

- transformational properties, such as passive, extraposition, cliticization, etc.

Altogether, 500 properties have been checked against the 10,000 verbs <4>.

More precisely, each property can be viewed as a sentence form. Consider for example the transitive structure

(1) N_0 *V* N_1

We are using Z.S. Harris' notation for sentence structure: noun phrases are indexed by numerical subscripts, starting with the subject indexed by 0. We can note the property "human subject" in the following equivalent ways:

(2) *Nhum V* N_1 or N_0 (=: *Nhum*) *V* N_1

where the symbol =: is used to specify a structure . A passive structure will be noted

(3) N_1 *be V-ed by* N_0

A transformation is a relation between two structures noted "=":

(1) = (3) corresponds to the Passive rule

The syntactic information attached to simple sentences can thus be represented in a uniform way by means of binary matrix (Table 1). Each row of the matrix corresponds to a verb, each column to a sentence form. When a verb enters into a sentence form, a "+" sign is placed at the intersection of the corresponding row and column, if not a "-" sign. The description of the French verbs does not have the shape of a 10,000x500 matrix. Because of its redundancy (cf. note 4), the matrix has been broken down into about 50 submatrices whose size is 200x40 on the average. It is such a system of submatrices that we call a lexicon-grammar.

---

1 E.R.A. 247 of the C.N.R.S. afiliated to the Universities Paris 7 and Paris VIII.

2 Publication of the lexicon-grammar is under way. The main segments available are: Boons, Guillet, Leclère 1976a, 1976b and Gross 1975 for French verbs, Giry-Schneider 1978, A. Meunier 1981, de Négroni 1978, for nominalizations.

3 Actually, the shape of interrogative pronouns: *qui* (who), *que-quoi* (what) has been used to define a formal notion of object.

4 Not all properties are relevant to each of the 10,000 verbs. For example, the properties of clitics associated to object complements are irrelevant to intransitive verbs.

N_0 = N hum	N_0 = N pc	N_0 = N-hum	N_0 = N nr	N_0 = V Ω	N_0 = V-n	N_0 = N plur obl		N_0 V	N_0 est V-ant	N_0 est V pp	N° pc lui V	N_0 de N° pc		N_1 = N hum	N_0 V Prép N pc de N_1	N_0 lui V Prép N^1 pc	N_0 V Prép N pc de N_0	N_0 V Prép N° pc	N_1 = N-hum	N_1 = le fait que P	Ppv	N_1 = V-n	N_1 = N plur obl	N_0 V de N° pc Prép N_1	N hum V sur ce point	N_0 est V-ant Prép N_1	il V N_0 Ω
+	-	-	-	-	-	-	dîner	+	-	-	-	-	de	-	+	-	-	-	+	-	-	+	-	-	-	-	+
-	+	+	-	-	-	-	éclater	+	-	+	+	-	en	-	-	-	-	-	+	-	-	-	+	-	-	-	+
+	-	+	-	-	-	-	écoper	+	-	-	-	-	de	+	-	-	-	-	+	-	+	-	-	-	-	-	-
+	+	-	-	-	-	-	s'emparer	-	-	-	-	-	de	+	+	-	+	-	+	-	+	-	-	-	-	-	-
+	-	-	-	-	-	-	s'emporter	+	-	+	-	-	contre	+	+	-	+	-	+	+	-	-	-	-	+	-	-
+	+	+	+	+	-	-	l'emporter	+	-	-	-	-	sur	+	+	-	+	-	+	+	-	-	-	-	+	-	-
+	-	+	+	+	-	-	évoluer	+	-	+	-	-	vers	-	-	-	-	-	+	-	-	-	-	-	-	-	+
-	+	+	-	-	-	-	exploser	+	-	+	+	-	en	-	-	-	-	-	+	-	-	-	+	-	-	-	+
+	-	-	-	-	-	-	se fendre	+	-	-	-	-	de	-	-	-	-	-	+	-	+	-	-	-	-	-	-
+	-	-	-	-	-	-	festoyer	+	-	-	-	-	de	-	+	-	-	-	+	-	+	-	-	-	-	-	-
+	-	-	-	-	-	-	financer	+	-	-	-	-	avec	+	+	-	+	-	+	-	-	-	-	-	+	-	-
+	+	+	+	+	-	-	en finir	+	-	-	-	-	avec	+	+	-	+	-	+	+	-	-	-	-	+	-	-
+	-	-	-	-	-	-	fondre	-	-	-	-	-	en	-	-	-	-	-	+	-	-	-	+	-	-	-	-
+	+	+	-	-	-	-	fondre	-	-	-	-	-	sur	+	+	+	+	-	+	-	-	-	-	-	-	-	+
+	-	-	-	-	-	-	fuir	+	-	-	-	-	devant	+	-	-	-	-	+	+	-	-	-	-	-	-	+

(Column groups: N_0 | | | N_1 | P.A.)

Intransitive Verbs *(From Boons, Guillet, Leclère 1976a)*

Table 1

Although the 3 prepositions "zero", *à* and *de* are felt and described as the basic ones by traditional grammarians, the descriptions have never received any objective basis. The lexicon-grammar we have constructed provides a general picture of the shapes of objects in French. The numerical distribution of object patterns is given in table 2, according to their number in a sentence and to their prepositional shape.

N_0 V	1,800
N_0 V N_1	3,700
N_0 V à N_1	350
N_0 V de N_1	500
N_0 V N_1 N_2	150
N_0 V N_1 à N_2	1,600
N_0 V N_1 de N_2	1,900
N_0 V à N_1 à N_2	3
N_0 V à N_1 de N_2	10
N_0 V de N_1 de N_2	1

DISTRIBUTION OF OBJECTS

Table 2

As can be seen on table 2, direct objects are the most numerous in the lexicon. Also, we have not observed a single example of verbs with 3 objects according to our definition.

In 2. and 3. we will make more precise the lexical nature of the N_i's attached to the verbs.

The signs in a row of the matrix provides the syntactic paradigm of a verb, that is, the sentence forms into which the verb may enter. The lexicon-grammar is in computer form. Thus, by sorting the rows of signs, one can construct equivalence classes for verbs: Two verbs are in the same class if their two rows of signs are identical.

We have obtained the following result: for 10,000 verbs there are about 8,000 classes.

On the average, each class contains 1.25 verb. This statistical result can easily be strengthened. When one studies the classes that contain more than one verb, it is always possible to find syntactic properties not yet in the matrix and that will separate the verbs. Hence, if our description were extended, each verb would have a unique syntactic paradigm.

Thus, the correspondence between a verb morpheme and the set of sentence forms where it may occur is one-to-one.

Another way of stating this result is by saying that structures depend on individual lexical elements, which leads to the following representation of structures:

N_0 *eat* N_1

N_0 *give* N_1 *to* N_2

We still use class symbols to describe noun phrases, but specific verbs must appear in each structure. Class symbols of verbs are no longer used, since they cannot determine the syntactic behaviour of individual verbs.

The nature of the lexicon-grammar should then become clearer. An entry of the lexicon-grammar of verbs is a simple sentence form with an explicit verb appearing in a row. In general, the declarative sentence is taken as the representative element of the equivalence class of structures corresponding to the "+" signs of a row.

The lexicon-grammar suggests a new component for parsing algorithms. This component is limited to elementary sentences. It includes the following steps:

- (A) Verbs are morphologically recognized in the input string.

- (B) The dictionary is looked up, that is, the space of the lexicon-grammar that contains the verbs is searched for the input verbs.

- (C) A verb being located in the matrix, its rows of signs provide a set of sentence forms. These dictionary forms are matched with the input string.

This algorithm is incomplete in several respects:

- In step (C), matching one of the dictionary shapes with the input string may involve another component of the grammar. The structures represented in the lexicon-grammar are elementary structures, subject only to "unary" transformations, in the sense of Harris' transformations or of early generative grammar (Chomsky 1955). Binary or generalized transformations apply to elementary sentences and may change their appearance in the sentence under analysis (e.g. conjunction reduction). As a consequence, their effect may have to be taken into account in the matching process.

- Looking up the matrix dictionary may result in the finding of several entries with same form (homographs) or of several uses of a given entry. We will see that these situations are quite common. In general, more than one pattern may match the input, multiple paths of analysis are thus generated and require book keeping.

We will come back to these aspects of syntactic computation. We now present two other components of the lexicon-grammar of simple sentences.

2. IDIOMS

The sentences we just described can be called free sentences, for the lexical choices of nouns in each noun phrase N_i has certain degrees of freedom. We use this distributional feature to separate free from frozen sentences, that is, from sentences with an idiomatic part.

The main difference between free and frozen sentences can be stated in terms of the distributions of nouns:

- in a frozen nominal position, a change of noun either changes the meaning of the expression to an unrelated expression as in

to lay down one's arms vs *to lay down one's feet*

or else, the variant noun does not introduce any difference in meaning (up to stylistic differences), as in

to put someone off the (scent, track, trail)

or else, an idiomatic noun appears at the same level as ordinary nouns of the distribution, and the general meaning of the (free) expression is preserved, as in

to miss (an opportunity, the bus)

- in a free position, a change of noun introduces a change of meaning that does not affect the general meaning of the whole sentence. For example, the two sentences

The boy ate the apple

My sister ate the pie

that differ by distributional changes in subject and object positions have same general meaning: changes can be considered to be localized to the arguments of the predicate or function with constant meaning EAT.

We have systematically described the idiomatic sentences of French, making use of the framework developed for the free sentences. Sentential idioms have been classified according to the nature (frozen or not) of their arguments (subject and complements). With respect to the structures of Table 2, a new classificatory feature has been introduced: the possibility for a frozen noun or noun phrase to accept a free noun complement. Thus, for example, we built two classes CP1 and CPN corresponding to the two types of constructions

N_0 *V Prep* C_1 =: *Jo plays on words*

N_0 *V Prep Nhum's* C_1 =: *Jo got on Bob's nerves*

The symbol C refers to a frozen nominal position and *Prep* stands for preposition.

Although frozen structures tend to undergo less transformations than the free forms, we found that every transformation that applies to a free structure also applies to some frozen structures. There is no qualitative difference between free and frozen structures from the syntactic point of view. As a consequence, we can use the same type of representation: a matrix where each idiomatic combination of words appears in a row and each sentence shape in a column (cf. Tables 3 and 4).

SUJETS N_0 = Nhum N_0 = ? Nhum	VERBES	ADVERBES FIGES		
+ -	VENIR	DANS	LA	"PERIODE"
+ -	PARTIR	SUR	CE	-
+ -	DEMONTRER N A N	PAR	L'	ABSURDE
+ -	PARTIR	DANS	L'	AFFIRMATIVE
+ -	DIRE NA N	EN	L'	AIR
+ -	TRICHER	A	POSS-O	AISE
+ -	ARRETER S	A	L'	ALLER
+ -	VENIR	A	TOUTE	ALLURE
+ -	ESPERER N	DE	TOUTE POSS-O	AME
+ -	ARRANGER N	A	L'	AMIABLE
+ -	GAGNER N	A	L'	ARRACHE
+ -	VENIR	CONTRE	TOUTE	ATTENTE
+ -	PARTIR	A	L'	AUBE
+ +	VENIR	PAR	L'	AUTOBUS
+ -	PAYER N	A	L'	AVANCE
+ -	CONSULTER N	A	L'	AVENIR
+ -	CONSULTER N	DANS	L'	AVENIR
+ -	CHOISIR N	A	L'	AVEUGLETTE
+ -	DISCUTER	-	TOUT	AZIMUT
+ -	BOIRE N	AVANT	LA	BAGARRE
+ -	SPECULER	A	LA	BAISSE
+ -	PARLER	-	TOUT	BAS
+ -	TRICHER	DE	PLUS	BELLE
+ -	FONCER	A	TOUTE	BERZINGUE
+ -	AGIR	A	LE	BESOIN
+ -	CUIRE N	A	LE	BEURRE
+ -	FONCER	A	TOUTE	BITURE
+ -	CUIRE N	A	LE	BOIS
+ -	ACCEPTER N	EN	TOUTE	BONNE FOI
+ -	RIRE	DE	TOUTE POSS-O	BOUCHE
+ -	LUTTER	JUSQU'A	LE	BOUT
+ -	CUIRE N	SUR	LA	BRAISE
+ -	FONCER	A	TOUTE	BRIDE
+ -	CUIRE N	A	LA	BROCHE
+ +	VENIR	PAR	LE	BUS
+ -	CUIRE N	A	LE	BUTAGAZ
+ -	CUIRE N	A	LE	BUTANE
+ -	DORMIR	EN	TOUT	CAS
+ -	CUIRE N	SOUS	LA	CENDRE
+ -	REMBOURSER N	A	LE	CENTUPLE

Frozen adverbs

Table 3

We have systematically classified 15.000 idiomatic sentences. When one compares this figure with those of table 2, one must conclude that frozen sentences constitute one of the most important components of the lexicon-grammar.

An important lexical feature of frozen sentences should be stressed. There are examples such as

They went astray

where words such as *astray* cannot be found in any other syntactically unrelated sentence; notice that the causative sentence

This led them astray

is considered as syntactically related. In this case, the expression can be direcly recognized by dictionary look-up. But such examples are rare. In general, a frozen expression is a compound of words that are also used in free expressions with unrelated meanings. Hence, frozen sentences are in general ambiguous, having an idiomatic meaning and a literal meaning.

However, the literal meanings are almost always incongruous in the context where the idiomatic meaning is intended (unless of course the author of the utterance played on words). Thus, when a word combination that constitutes an idiom is encountered in a text, one is practically ensured that the corresponding meaning is the idiomatic one.

N0 =: Nhum	N0 =: N-hum		N0 V	Passif		Npc	Autres déterminants	
+	-	CONNAITRE	-	+	LE	-	-	COUP
+	-	CONNAITRE	-	-	POSS-Ø	-	-	DOULEUR
+	-	CONNAITRE	-	+	LE	-	-	TRUC
+	-	NE CONNAITRE PAS	-	-	POSS-Ø	-	-	BONHEUR
+	-	NE CONNAITRE QUE	-	-	-	-	-	ÇA
+	-	CONSERVER	-	-	POSS-Ø	-	-	CHEMISE
+	-	SE CONTEMPLER	-	-	LE	+	-	NOMBRIL
+	-	COUPER	-	+	DET	-	-	CORDON OMBILICAL
+	+	DEBLOQUER	-	+	det	-	-	SITUATION
+	-	DETENIR	+	-	LA	-	-	VERITE
-	+	DISTILLER	-	+	LE	-	+	VENIN
+	+	DOMINER	-	+	LE	-	-	LOT
+	-	DRESSER	-	+	POSS-Ø	-	+	BATTERIES
+	-	ENDOSSER	-	+	LE	-	-	HARNOIS
+	+	ENFONCER	-	+	LE	-	-	CLOU
+	-	ETRE . N PAS	-	-	UNE	-	-	LUMIERE
+	-	ETRE . N PAS	-	-	-	-	-	MANCHOT
-	+	ETRE . N PAS	-	-	LA	-	-	MORT
+	-	ETRE . S DIT	-	-	-	-	-	TOUT
+	-	FAIRE	-	-	UN	-	-	BRIN DE TOILETTE
+	-	FAIRE	-	-	-	-	-	GRISE MINE
+	-	FAIRE	-	-	-	-	-	HARA-KIRI
-	+	FAIRE	-	-	-	-	-	JURISPRUDENCE
+	-	FAIRE	-	+	UNE	-	+	MINUTE DE SILENCE
-	+	FAIRE	-	-	-	-	+	NOMBRE
+	-	FAIRE	-	+	DET	-	-	OPERATION PORTE OUVERTE
+	-	FAIRE	-	-	DU	-	-	QUARANTE CINQ FILLETTE
+	-	FAIRE	-	-	-	-	-	TAPIS
+	-	FAIRE	-	-	-	-	-	TINTIN
+	-	FAIRE ENTENDRE	-	-	POSS-Ø	-	-	VOIX
+	-	FAIRE PASSER	-	-	DET	-	-	ENFANT
+	-	FAIRE SAUTER	-	-	DET	-	-	ENFANT
-	+	FERMER	-	+	POSS-Ø	-	-	PORTES
+	-	FLETRIR	-	+	DET	-	-	CRIME
+	-	FORCER	-	-	LA	-	+	CHANCE
+	-	FORMER	-	+	LE	-	-	CARRE
+	-	FORMER	-	+	DET	-	-	NUMERO
+	-	FORMER	-	+	DET	-	-	NUMERO DE TELEPHONE
+	-	FORMER	-	+	LES	-	-	RANGS
+	+	FRANCHIR	-	+	DET	-	-	CAP

Frozen sentences

Table 4

Returning to the algorithm sketched in 1., we see that we have to modify steps (A) and (B) in order to recognize frozen expressions:

- Not only verbs, but nouns have to be immediately located in the input string.

- The verbs and the nouns columns of the lexicon-grammar of frozen expressions have to be looked up for combinations of words.

It is interesting to note that there is no ground for stating a priority such as look up verbs before nouns or the reverse. Rather, the nature of frozen forms suggests simultaneous searches for the composing words.

About the difference between free and frozen sentences, we have observed that many free sentences (if not all) have highly restricted nominal positions. Consider for example the entry N_0 *smoke* N_1 in

Jo smokes the finest tobacco

In the direct object complement, one will find few other nouns: nouns of other smoking material, objects made of smoking material such as *cigarette, cigar, pipe* and brand names for these objects. This is a common situation with technical verbs. Such examples suggest that, semantically at least, the nominal arguments are limited to one noun, which comes close to having the status of frozen expression. Thus, *to smoke* would have here one complement, perhaps *tobacco*, and all other nouns occurring in its place would be brought in by syntactic operations. We consider that this situation is quite general although not always transparent. Our analysis of free elementary sentences has shown that when subjects and objects allow wide variations for their nouns, then well defined syntactic operations account for the variation:

- separation of entries: For example, there is another verb N_0 *smoke* N_1, as in *They smoke meat*, and a third one: N_0 *smoke* N_1 *out* in *They smoked the room out*; or consider the verb *to eat* in

Rust ate both rear wings of my car

This verb will constitute an entry different of the one in *to eat lamb*;

- various zeroings: The following sentence pairs will be related by different deletions:

Bob ate a nice preparation

= *Bob ate a nice preparation of lamb*

Bob ate a whole bakery

= *Bob ate a whole bakery of apple pies*

Other operations introduce nouns in syntactic positions where they are foreign to the semantic distributions, among them are

- raising operations, which induce distributional differences such as

I imagined the situation

I imagined the bridge destroyed

situation is the "natural" direct object of *to imagine*, while *bridge* is derived;

- other restructuration operations (Guillet, Leclère 1981), as between the two sentences

This confirmed Bob's opinion of Jo

This confirmed Bob in his opinion of Jo

Although the full lexicon of French has not yet been analyzed from this point of view, we can plausibly assert that a large class of nominal distributions could be made semantically regular by using Z.S. Harris' account of elementary distributions, namely, by

- determining a basic form for each meaning, for example

A person eats food

with undetermined human subject and characteristic object, and by

- introducing classificatory sentences that describe the semantic universe:

(The boy, My sister) is a person, etc.

(A pie, This cake) is food, etc.

Classificatory and basic sentences are combined by syntactic operations such as

- relativization:

The person who is the boy eats food which is this pie

- *WH-is* deletion:

The person the boy eats food this pie

- redundancy removal:

The boy eats this pie

In this way, the semantic variations are explicitly attributed to lexical variations, and not to intuitive abstract features, that is, arbitrary features, or semes or the like. The requirement of using WORDS in such descriptions is a crucial means for controlling the construction of an empirically adequate linguistic system. In this respect, one is led to categorizing words by evaluating actual classificatory sentences. Hence, all the knowledge linguistically expressible (i.e. in terms of words) is represented by both the basic and the classificatory sentences. A good deal of the inferences that one has to draw in order to understand sentences are contained in the derivations that lead to the seemingly simple sentences.

From a formal point of view, the entries of the lexicon-grammar become much more specific. We have eliminated class symbols altogether, replacing them by specific nouns <5>. Entries are then of the type

$(person)_0$ *eat* $(food)_1$

$(person)_0$ *give* $(object)_1$ *to* $(person)_2$

$(person)_0$ *kick the bucket*

An application of this representation of simple sentences is the treatment of certain metaphors. Consider the two sentences

(1) *Jo filled the turkey with truffles*

(2) *Jo filled his report with poor jokes*

(1) is a proper use of *to fill*, while (2) is a metaphoric or figurative meaning. The properties of these sentences vary according to the lexical choices in the complements (Boons 1971). For example, the *with*-complement that can be occupied by an internal noun in the proper meaning can be omitted:

Jo filled the turkey with a certain filling

= *Jo filled the turkey*

This is not the case in the figurative meaning:

**Jo filled his report*

How to represent (1) and (2) is a problem in terms of number of entries. On the one hand, the two constructions have common syntactic and semantic features, on the other, they are significantly different in form and content. Setting up two entries is a solution, but not a satisfactory one, since both entries are left unrelated. A possible solution in the framework of lexicon-grammars is to consider having just one entry:

N_0 *fill* N_1 *with* N_2

and to specify N_1 lexically by means of columns of the matrix. For example

N_1 =: *food*

N_1 =: *text*

Then, the content of N_2 is largely determined and has to be roughly of the type

N_2 =: *stuffing*

N_2 =: *subtext*

An inclusion relation <6> holds between the two complements. We can write for this relation

N_2 *is in* N_1

But now, in our parsing procedure, we have to compensate for the fact that in the lexicon-grammar, the nouns that are represented in the free positions are not the ones that in general occur in the input sentences. In consequence, occurrences of nouns will have to undergo a complex process of identification that will determine whether they have been introduced by syntactic operations (e.g. restructuration), or by chains of substitutions defined by classificatory sentences, or by both processes.

3. SUPPORT AND OPERATOR VERBS

We have alluded to the fact that only a certain class of sentences could be reduced to entries of the lexicon-grammar as presented in 1. and 2. We will now give examples of simple sentences that have structures different of the structures of free and frozen sentences. In sentences such as

(1) *Her remarks made no difference*

(2) *Her remarks have some (importance for, influence) on Jo*

(3) *Her remarks are in contradiction with your plan*

it is difficult to argue that the verbs *to make*, *to have* and *to be in* semantically select their subjects and complements. Rather, these verbs should be considered as auxiliaries. The predicative element is here the nominal form in complement position. This intuition can be given a formal basis. Let us look at nominalizations as being relations between two simple sentences (Z.S. Harris 1964), as in

5 It is doubtful that actual nouns such as *food* will be available in the language for each distribution of each entry, but then, expressions such as *smoking stuff* can be used (in the object of *to smoke*), again avoiding the use of abstract features.

6 This relation is an extension of the *Vsup* relations of 3. *To fill* could be considered as a (causative) *Vop*.

Max walked

= *Max took a walk*

Her remarks are important for Jo

= *Her remarks are of a certain importance for Jo*

= *Her remarks have a certain importance for Jo*

Jo resembles Max

= *Jo has a certain resemblance with Max*

= *Jo (bears, carries) a certain resemblance with Max*

= *There is a certain resemblance between Jo and Max*

It is then clear that the roots *walk*, *important* and *resemble* select the other noun phrases. We call support verbs (*Vsup*) the verbs in such sentences that have no selectional function. Some support verbs are semantically neutral, others introduce modal or aspectual meanings, as for example in

Bob loves Jo

= *Bob is in love with Jo*

= *Bob fell in love with Jo*

= *Bob has a deep love for Jo*

to fall, as other motion verbs do, introduces an inchoative meaning. In this example, the main semantic relation holds between Bob and *love*, and the support verbs simply add their meaning to the relation.

If we use a dependency tree to schematize the relations in simple sentences, we can oppose ordinary verbs with one object and support verbs of superficially identical structures such as in figure 1:

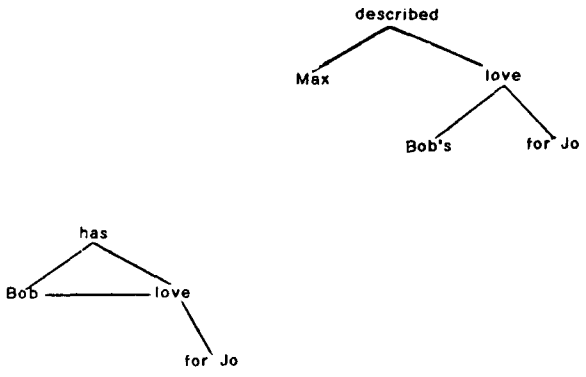

Figure 1

Two problems arise in connection with the distribution of support verbs:

- a noun or a nominalized verb accepts a certain set of support verbs and this set varies with each nominal;

- not every verb is a support verb; thus in the sentence

(4) *Max described Bob's love for Jo*

to describe is not a *Vsup*. The question is then to delimit the set of *Vsups*, if such a set can be isolated, or else to provide general conditions under which a verb acts as a *Vsup*.

One of the structural features that separates support verbs from other verbs is the possibility of clefting noun complements. For example, *for Jo* is a noun complement of the same type in both structures, but we observe

**It is for Jo that Max described Bob's love*

It is for Jo that Bob has a deep love

The main semantic difference between the two constructions lies in the cyclic structure of the graph. This cyclic structure is also found in more complex sentences such as

(5) *This note put her remarks in contradiction with your plan*

(6) *Bob gave a certain importance to her remarks*

Both verbs *to put* and *to give* have two complements, exactly as in sentences such as

(7) *Bob put* (*the book*)$_1$ (*in the drawer*)$_2$

(8) *Bob gave* (*a book*)$_1$ (*to Jo*)$_2$

While in (7) and (8), there is no evidence of any formal relation between both complements, in (5) and (6) we find dependencies already observed on support verbs (cf. figure 2).

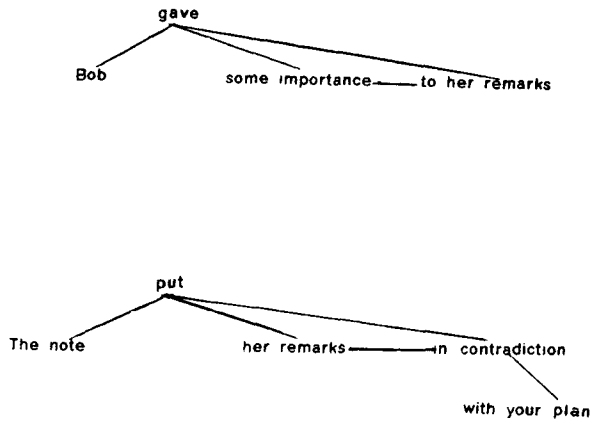

Figure 2

The verbs *to put* and *to give* are semantically minimal, for they only introduce a causative and/or an agentive argument with respect to the sentence with *Vsup*. We call such verbs operator verbs (*Vop*). There are other operator verbs that add various modalities to the minimal meanings, as in

The note introduced a contradiction between her remarks and your plan

Bob attributed a certain importance to her remarks

Other syntactic shapes are found:

Bob credited her remarks with a certain importance

Again, the set of nouns (supported by a *Vsup*) to which the *Vops* apply vary from verb to verb. As a consequence, we have to represent the distributions of *Vsups* and *Vops* with respect to nominals by means of a matrix such as the one in Table 4'.

In each row, we place a noun and each column contains a support verb or an operator verb. A preliminary classification of *Ns* (and *V-ns*) has been made in terms of a few elementary support verbs (e.g. *to have, to be Prep*).

In a sense, this representation is symmetrical with the representation of free sentences. With free sentences, the verb is taken as the central item of the sentence. Varying then the nouns allowed with the verb does not change fundamentally the meaning of the corresponding sentences. With support verbs, the central item is a noun. Varying then the support verbs only introduces a distributional-like change in meaning.

The recognition procedure has to be modified, in order to account for this component of the language:

- first, the look-up procedure must determine whether a verb is an ordinary verb (i.e. an entry found in a row of the lexicon-grammar) or a *Vsup* or a *Vop*, which are to be found in columns;

- simultaneously, nouns have to be looked up in order to check their combination with support verbs.

4. CONCLUSION

We have shown that simple sentence structures were of varied types. At the same time, we have seen that their representation in terms of the entries of traditional "linear" dictionaries, that is, in terms of words alphabetically or otherwise ordered, is inadequate. An improvement appears to involve the look-up of two-dimensional patterns, for example the matrices we proposed for frozen sentences and their generalization to support verbs and operator verbs. More generally, syntactic structures are determined by combinations of a verb morpheme with one or more noun morpheme(s). Hence, the general way to access the lexicon will have to be through the selectional matrix of Tables 3 and 4.

In practice, syntactic computations are context-free computations in natural language processing. Context-free algorithms have been studied in many respects by computer scientists, theoreticians and specialists of programming languages. The principles of these algorithms are clearly understood and currently in use, even for natural languages where new problems arise because of the numerous ambiguities and the various terminologies attached to each theoretical viewpoint.

The fact that context-free recognition is a mastered technique has certainly contributed to the shaping of the grammars used in automatic parsing. The numerous sample grammars presented so far are practically all context-free. There is also a deep linguistic reason for building context-free grammars: natural languages use embedding processes and tend to avoid discontinuous structures.

Much less attention has been paid to the complex syntactic phenomena occurring in simple sentences and to the organization of the lexicon. The fact that we could not separate the syntactic properties of verbs from their lexical features has led us to construct a representation for linguistic phenomena which is more specific than the current context-free models. A context-free component will still be useful in the parsing process, but it will be relevant only to embedded structures found in complex sentences, with not much incidence on meaning.

To summarize, the syntactic patterns are determined by pairs (verb, noun):

- the frozen sentence N_0 *kick the bucket* is thus entirely specified, while the pair (*take, bull*) needs to be disambiguated by the second complement *by the horns*, requiring thus a more complex device to be identified;

- (*take, walk*) and (*take, food*) are support sentences, so are (*have, faith*) and (*have, food*);

- the verbs *have, kick* and *take* together with *concrete object* select ordinary sentence forms.

But the selectional process for structures may not be direct. The words in the previously discussed pairs may not appear in the input text. Words appearing in the input are then related to the words in the selectional matrix by:

- classificational relations:

  food classifies *cake, soup*, etc.

  concrete object classifies *ball, chair*, etc.

- relations between support sentences, such as

  Jo (had, took,threw out) some food

  Jo (took, was out for, went out for) a walk

  Jo (has, keeps, looses) faith in Bob

- relations between support and operator sentences:

  This gave to Jo faith in Bob

All these relations in fact add a third dimension to the selectional matrix.

The complete selectional device is now a complex network of relations that cross-relates the entries. It will have to be organized in order to optimize the speed of parsing algorithms.

REFERENCES

Boons, J.-P. 1971. Metaphore et baisse de la redondance, *Langue française* 11, Paris: Larousse, pp. 15-16.

Boons, J., Guillet, A. and Leclère, Ch. 1976a. *La structure des phrases simples en français. Constructions intransitives*, Droz, Geneva, 377 p.

Boons, J., Guillet, A. and Leclère, Ch. 1976b. *La structure des phrases simples en français. Classes de constructions transitives*, Rapport de recherches No 6, Paris: University Paris 7, L.A.D.L., 143 p.

Freckleton, P. 1984. *A Systematic Classification of Frozen Expressions in English*, Doctoral Thesis, University of Paris 7, L.A.D.L.

Giry-Schneider, J. 1978. *Les nominalisations en français. L'opérateur FAIRE*, Geneva: Droz, 414 p.

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

Gross, Maurice 1982. Une classification des phrases figées du français, *Revue québécoise de linguistique*, Vol. 11, No 2, Montréal : Presses de l'Université du Québec à Montréal, pp. 151-185.

Guillet, A. and Leclère, Ch. 1981. Restructuration du groupe nominal, *Langages*, Paris : Larousse, pp. 99-125.

Harris, Z.S. 1964. The elementary Tranformations, Transformations and Discourse Analysis Papers 54, in Harris, Zellig S. 1970, *Papers in Structural and Transformational Linguistics*, Reidel, Dordrecht, pp. 482-532.

Harris, Zellig 1983. *A Grammar of English on Mathematical Principles*, New York : Wiley Interscience,429 p.

Meunier, A. 1977. Sur les bases syntaxiques de la morphologie dérivationnelle, *Lingvisticae Investigationes* I:2, John Benjamins B.V., Amsterdam, pp. 287-331.

Négroni-Peyre, D. 1978. Nominalisations par ETRE EN et réflexivation, *Lingvisticae Investigationes* II:1, John Benjamins B.V., Amsterdam, pp. 127-163.