# French Order Without Order

Gabriel G. Bès, Claire Gardent

# FRENCH ORDER WITHOUT ORDER*

Gabriel G. Bès
Université Blaise Pascal – Clermont II, Formation Doctorale Linguistique et Informatique
34 Ave. Carnot, 63037 Clermont-Ferrand Cedex, FRANCE

Claire Gardent
Université Blaise Pascal – Clermont II and University of Edinburgh, Centre for Cognitive Science,
2 Buccleuch Place, Edinburgh EH89LW, SCOTLAND, UK

## ABSTRACT

To account for the semi-free word order of French, Unification Categorial Grammar is extended in two ways. First, verbal valencies are contained in a set rather than in a list. Second, type-raised NP's are described as two-sided functors. The new framework does not overgenerate i.e., it accepts all and *only* the sentences which are grammatical. This follows partly from the elimination of false lexical ambiguities – i.e., ambiguities introduced in order to account for all the possible positions a word can be in within a sentence – and partly from a system of features constraining the possible combinations.

## INTRODUCTION

In the version of categorial grammar (henceforth, CG) developed by Bar-Hillel (Bar-Hillel 1953), categories encode both constituency and linear precedence. Linear precedence is encoded by (a) ordering valencies in a list and (b) using directional slashes indicating whether the argument is to be found to the left or to the right of the functor.

A similar approach is adopted in Unification Categorial Grammar (UCG) (Zeevat, Klein and Calder 1987) as regards word order whereby the directional slash is replaced by a binary *Order* feature with value *pre* or *post*. Thus, S/NP\NP in normal CG translates as S/NP:pre/NP:post in UCG, where *pre* indicates that the functor must precede the argument and *post* that it should follow it.

Our work on French syntax supports the claim that the complicated pattern of French linearity phenomena can be treated in a framework closely related to UCG but which departs from it in two ways. First, there is no rigid assignment of an order value (*pre* or *post*) to verb valencies. Second, following (Gunji 1986) verbal valencies are viewed as forming a set rather than a list. As a result, the syntactic behaviour of constituents is dissociated from surface ordering. Constraints on word order are described by a system of features as advocated in (Uszkoreit 1987) and (Karttunen 1986).

* The word reported here was carried out as part of ESPRIT Project 393 ACORD, "The Construction and Interrogation of Knowledge Bases using Natural Language Text and Graphics".

## 1. UCG

In UCG, the phonological, categorial, semantic and order information associated with a word is contained in a single grammar structure called a *sign*. This can be represented as follows.

(1) UCG sign
Phonology:Category:Semantics:Order

*or equivalently*

Phonology
:Category
:Semantics
:Order

where colons separate the different fields of the sign.

We need not concern ourselves here with the *Semantics* and the *Phonology* fields of the sign. More interesting for our purpose are the *Category* and the *Order* attributes. Following the categorial tradition, categories can be basic or complex. Basic categories are of the form *Head^Features* where *Head* is one of the atomic symbols *n*(oun), *np* or *s*(entence) and *Features* is a list of feature values. Complex categories are of the form *C/Sign*, where *C* is either atomic or complex and *Sign* is a sign, so that departing from traditional CG's, a functor places constraints on the whole sign of the argument rather than on its syntactic category only. The part of a complex category which omits the *Head^Features* information constitutes the *active part* of the sign. The first *accessible sign* in the active part is called the *active sign*, the idea being that e.g. verb valencies are ordered in a list so that each time a valency is consumed, the next sign in the active part becomes the new active sign. The *Order* attribute places constraints on the combination rule that may apply to a functor: *pre* on an argument sign Y indicates that the functor X/Y must precede the argument, while *post* indicates that the functor must follow the argument.

Using terms and term unification, the forward version[1] of functional application can then be stated as follows.

(2) Forward Application

| *Functor* | , *Argument* |
|---|---|
| PhonologyF | |
| :CategoryF/PhonologyA | PhonologyA |
| : CategoryA | : CategoryA |
| : SemanticsA | : SemanticsA |
| : pre | : pre |
| :SemanticsF | |
| :OrderF | |

⇒ *Result*

PhonologyF PhonologyA
:CategoryF
:SemanticsF
:OrderF

were upper letters indicate Prolog variables. In effect, the rule requires that the active part of the functor sign term unifies with the argument sign. The *Result* is a sign identical to the functor sign, but where the complex category is stripped from its active part and where variables shared by the active part of the functor and the rest of the functor sign may have become ground as a result of the active part unifying with the argument. The resulting phonology consists of the phonology of the functor followed by the phonology of the argument. An illustrative combination is given in (3) below for the sentence *Jean marche*.

(3) Derivation of *Jean marche*

| jean | | |
|---|---|---|
| : C/_ | | marche |
| :(C/(_:np:jean':O) | | :s^[fin]/(_:np:X:pre) |
| : S | | : marche'(X) |
| : O | | :_ |
| : S | | |
| :_ | | |

⇒ jean marche
:s^[fin]
:marche'(jean')
:_

where lines represent the information flow determining ordering : shared variables ensure that *pre* in a verb valency constrains the functor NP that consumes this valency to precede the verb carrying this valency.

## 2. LINGUISTIC OBSERVATIONS

Word order in French is characterised by three main facts. First, the positioning - left or right - of a particular argument with respect to the verb is relatively free. As illustrated in (4), the subject can appear to the left (4a) or to the right (4b,c) of the verb, or between the auxiliary and the verb (4d), depending on the morphological class of the NP and on the type of the sentence (declarative, subject-auxiliary inversion, wh-question, etc).

(4) (a) *Jacques* aime Marie.
(b) Aime-t-*il* Marie ?
(c) Quel livre aime *Jacques* ?
(d) A-t-*il* aimé Marie ?

All other arguments can also appear to the left or to the right of the verb under similar conditions. For example, a lexical non-nominative NP can never be to the left of the verb, but clitics and wh-constituents can.

(5) (a) **Marie* a regardée Jacques ?
(with *Marie* = Obj)
(b) *Quelle revue* a regardée Jacques ?
(with *Quelle revue* = Obj)
(c) Jacques *l'*a regardée

Second, there seems to be no clear regularities governing the relative ordering of a sequence of arguments. That is, assuming that only adjacent constituents may combine and taking the combinations left-to-right, the combination pattern varies as indicated below of each example in (6). Here again, the permissible distributions are influenced by factors such as the morphological class of the constituents and the verb mood.

(6) (a) Pierre donne à Marie un livre.
[Subj,IObj,Obj]
(b) Pierre donne un livre à Marie.
[Subj,Obj,IObj]
(c) Le lui donne-t-il ?
[Obj,IObj,Subj]
(d) Se le donne-t-il ?
[IObj,Obj,Subj]

Third, coocurrence restrictions hold between constituents. For example, clitics constrain the positioning and the class of other arguments as illustrated in (7)[2]

(7) (a) Pierre le *lui* donne.
(b) Pierre *lui* en donne.
(c) Pierre *lui* donne un livre.
(d) *Pierre *lui* le donne.
(e) *Pierre *lui* y donne.

Since the ordering and the positioning of verb arguments in French are very flexible, the rigid orde-

[1] Backward application is just the symmetric of (2) where the argument precedes the functor and *pre* becomes *post*.

[2] In italics : the word whose coocurrence restriction is violated (starred sentences) or obeyed (non-starred sentences). For instance, (7d) is starred because *lui* may not be followed by *le*.

ring forced by the UCG active list and the fixed positioning resulting from the *Order* attribute are rather inadequate. On the other hand, word order in French is not free either. Rather it seems to be governed by conditional ordering statements such as:

(8) IF (a) the verb has an object valency, and
(b) the object NP is a wh-constituent, and
(c) the verbal constituent is the simple inflected verb, and
(d) the clitic *t-il/elle* has not been incorporated
THEN the object can be placed to the left or to the right of the verb.

If say, (8d) is not fulfilled, the wh-NP can be placed only to the left, witness: **Jacques a-t-il regardé quelle revue ?*, and *mutatis mutandis* for the other conditions. More generally, five elements can be isolated whose interaction determine whether or not a given argument can occupy a given position in the sentence.

(9) (a) Position - left or right - with regard to the verb,
(b) Verbal form and sentence type,
(c) Morphological class (lexical, wh-constituent or clitic) of the previous constituent having concatenated to the left or to the right of the verb,
(d) Morphological class of the current constituent (lexical, wh-constituent or clitic),
(e) Case.

We claim that it is possible to extend UCG in order to express the above conditioning variables. The resulting grammar can account for the preceding linguistic facts without resorting either to lexical ambiguity or to jump rules[3].

## 3. EXTENSIONS TO UCG

To account for the facts presented in section 2, UCG has been modified in two ways. Firstly, the active part of a verb category is represented as a set rather than a list. Secondly, a feature system is introduced which embodies the interactions of the different elements conditioning word order as described in (9) above.

### 3.1 SIGN STRUCTURE AND COMBINATION RULE : FROM AN ACTIVE LIST TO AN ACTIVE SET.

To accomodate our analysis, the sign structure and the combination rule had to be slightly modified. In the French Grammar (FG), a sign is as follows.

(10) French Grammar Sign
Phonology
: Category
: Features
: Semantics
: Optionality
: Order

*Semantics* and *Phonology* are as in UCG. *Optionality* indicates whether the argument is optional or obligatory[4]. The *Category* attribute differs from UCG in that (i) there are no *Features* associated with the *Head* and (ii) the active part of a verb is viewed as a set rather than as a list.

The *Features* attribute is a list of features. In this paper, only those relevant to order constraints are mentioned. They are: case, verb mood, morphological class of NP's (i.e., lexical, clitic or wh-constituent) and last concatenation to the left *(Lastleft)* or to the right *(Lastright)*. The latter features indicate the morphological status of the last concatened functor and are updated by the combination rule (cf. (13)). For instance, the sign associated with *Jean lui a donné un livre* will have *lex* as values for *Lastleft* and *Lastright* whereas *lui a donné un livre* has *lui* and *lex* respectively. The *Features* attribute can be represented as in (11) below, where the same feature may occupy a different position in the feature list of different linguistic units, e.g., feature list of verb valencies and feature list of NP signs.

(11) The *Features* attribute
For valencies (active sign of NP's and verbs) :
[Case, Lastleft, Lastright]
For verb signs :
[Mclass, Lastleft, Lastright, Vmood]

As illustrated in (12), the *Order* attribute has two parts, one for when the functor combines forward, the other for when it combines backward.

(12) The *Order* attribute
Cdts ⇒ pre ⇒ Resfeat,
Cdts ⇒ post ⇒ Resfeat,

where *Cdts* and *Resfeat* are lists of feature values whose order and content are independent from those of the *Features* attribute. The intuition behind this is that functors (i.e., type-raised NP's) are two-sided i.e., they can combine to the left and to the right but under different conditions and with different results. The features in *Cdts* place constraints on the features of the argument while the features in *Resfeat* are inherited by the resulting sign. These effects are obtained by unifi-

[3] A jump rule as used in (Baschung et al. 1986), is of the form X/Y, Y/Z => X/Z.

[4] In the rest of this paper, the Semantics and the Optionality attributes will be omitted since they have no role to play in our treatment of word order while Phonology will only be represented when relevant.

cation of shared variables in the rules of combination. Omitting Semantics and Optionality attributes, the forward combination rule is as follows.

(13) Forward Combination[5] (FC)

*Functor* ,
PhonologyF
:[1] CategoryF / PhonologyA
:CategoryA
:[MClassA, ...]
:([Lastleft,Vmood]⇒pre⇒[Vmood2],
_)
:[MClassF ...]
:_
*Argument* ,
PhonologyA
:[2] CategoryA'
:[MClassA,Lastleft,Lastright,Vmood]
:_
{combine ([1], [2], Category')}

⇒ *Result*
PhonologyF PhonologyA
: Category'
:[MClassA, MClassF, Lastright, Vmood2]

The rule requires (i) that the functor category [1] *combines* with the argument category [2] to yield the result category *Category'*. The notion of combination relies on the idea that the active part of a verb is a set rather than a list. More precisely, given a type-raised NP *NP1* with category C/(C/NP$_i$) where NP$_i$ is a valency sign, and a verb *V1* with category *s/ActSet* where *ActSet* is a set of valency signs, *NP1* combines with *V1* to yield *V2* iff NP$_i$ unifies with some NP-valency sign in the active set *ActSet* of the verb. *V2* is identical to *V1* except that the unifying NP$_i$ valency sign has been removed from the active set and that some features in *V1* will have been instantiated by the rule. Forward combination further requires (ii) that the two features in the condition list to *pre* unify with the *Lastleft* and *Vmood* features of the argument (the features conditioning *post* are ignored since they are relevant only when the functor combines backwards), and (iii) that the features of the resulting sign be as specified. Note in particular that the *MClass* of the resulting sign is the *MClass* of the argument, that *Lastright* which indicates the morphological class of the last sign to have combined with the verb from the right, is transmitted from the argument, and that *Lastleft* is assigned as value the *MClass* of the functor. Features of the resulting sign which are conditional on the combination order are inherited from the *Resfeat* field. This percolation of features by rule is crucial to our treatment of word order. It is illustrated by (14) below where the sign S1 associated with the clitic *le* combines with the sign for S2 *regarde* to yield a new sign S3 *le regarde*.

(14) Derivation of *le regarde*

S1 le
:C/(C/(np:[obj ...]:_)
:[obj ...]
:_)
:[vb, ...]
:([lui or i, ind] => pre => [ind],
[i, imp] =>post => [imp])
:[le,...]
:_

S2 regarde
:s/{ np :[obj ...] :_ , np :[subj ...] }
:[vb, i, i, ind]
:_

S3 le regarde
:s/{ np :[subj ...] :_ }
:[vb, le, i, ind]
:_

When *le* is used as a forward functor, the conditions on *pre* require that the argument i.e., the verb bears for the feature *Lastleft* the value *lui or i* where *i* stands for *initial* state thus requiring that the verb has not combined with anything on the left. When it combines by BC, the conditions on *post* ensure that the argument has not combined with anything on its right and that it has mood *imperative*. In this way, all sentences in (15) are parsed appropriately.

(15) (a) Il le lui donne.
(b) *Il lui le donne.
(c) Donne le lui.
(d) *Donne lui le.

The backward combination rule (BC) functions like FC except for two things. First, the argument must be to the left of the functor and second, the condition field considered is that of *post* rather than of *pre*. There is also a deletion rule to eliminate optional valencies. No additional rule is needed.

## 3.2 EXPRESSING THE VARIABLES UNDERLYING WORD ORDER CONSTRAINTS

In our grammar, there are no *post* and *pre* primitive values associated with specific verb valencies. Instead, features interact with combination rules to enforce the

[5] In this figure, numbers inside square denote the following attribute. For instance, [2] denotes CategoryA'.

constraints on word order described in (9). (9a) is captured in the two-sided order field. (9b - verb mood) and (9c - morphological class of preceding concatenating functor) are accounted for in that in a functor, the features conditioning order include the verb mood and the last concatenation attribute.

(9d) is accounted for in that *conditions which are invariant for a particular class of constituent* (clitic, wh-constituent, lexical NP) are expressed in the Order field of these constituents. For example, wh-constituents reject through their conditions to *pre* a wh-value for the *Lastleft* feature of the argument and an *invl*[6] value in their condition to *post*. As a result, the following sentences are parsed appropriately.

(16) (a) *A qui qui a téléphoné ?
(b) *A-t-il téléphoné a qui ?
(c) A qui a-t-il téléphoné ?
(d) Il a téléphone a qui ?

*Conditions which vary depending on the class of the concatenating constituent* are expressed in the *Features* attribute of the verb valencies. This allows us to express constraints on the position of a given type of NP (lex, wh or clitic) relative to the valency it consumes. For instance, a lexical NP can be subject or object. If it is subject and it is to the left of the verb, it cannot be immediately followed by a wh-constituent. If it is subject and it is placed to the right of the verb, it must be immediately adjacent to it. These constraints can be stated using unification along to the following lines.

A verb valency is of the form

(17) (np:[... X,Y ...]:Ord)

where X and Y are either the anonymous variable or a constant. They state constraints, among others, on possible values of *Lastleft* and *Lastright* features of the verb. Recall that a valency is a sign which is a member of a set in the *Category* attribute of a verbal sign.

The active sign of a type raised NP is of the form:

(18) C/(np:[... V1, V2...]:_)
:[vb|_]
:([V1 ...] => pre => Z, [V2 ...] => post => W)

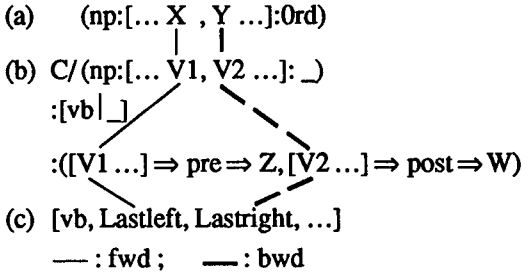
By rule, V1 and V2 in the *Category* attribute of (18) must unify with X and Y, respectively, in the verb valency (17). Being shared variables, they transmit the information to the *Conditions* on concatenation by FC (pre) and BC (post), respectively.

Furthermore, V1 and V2 in the *Ord* attribute of the functor must unify, by rule, with some specified features in the verb *Features* attribute represented in (19).

(19) [vb, Lasleft, Lastright, ...]

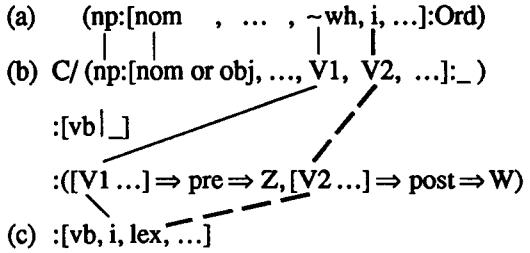
The flow of information between (17), (18) and (19) is represented graphically in (20), where (20a), (20b) and (20c) correspond to (17), (18) and (19) respectively. (20a) and (20c), which express the *Category* and the *Features* attibutes of the same verbal sign, have been dissociated for the sake of clarity.

(20) Flow of information between functor and argument




For example, suppose the nominative valency (21a), in the verbal sign *téléphone à la fille*, whose *Features* attribute is as in (21c), and the lexical sign *Jean* (21b).

(21) Flow of information between *Jean* and *téléphone à la fille*




The concatenation by FC is allowed (*~wh* is compatible with *i*), the requirement extracted from the verbal valency being that the Lastleft concatenated contituent with the verb is not a wh-constituent. But a concatenation by BC will fail (*i* does not unify with *lex*). Thus examples in (22) are correctly recognised (see Appendix).

(22) (a) Jean téléphone à la fille.
(b) *Téléphone à la fille Jean.
(c) *Jean à quelle fille téléphone ?
(d) A quelle fille téléphone Jean ?

## 4. IMPLEMENTATION

The UCG formalism and the corresponding computational environment were developped at the Centre for Cognitive Science, University of Edinburgh by (Calder et al. 1986). They include facilities for defining templates and path-equations as in PATR-2 and a shift-reduce parser. The extensions to the original framework have been implemented at the Université Blaise

[6] The value *invl* for the *Lastleft* feature of a verb results from a backward combination of the nominative clitic *-t-il* with this verb.

Pascal, Formation Doctorale Linguistique et Informatique, Clermont-Ferrand (France). The system runs on a Sun and has been extensively tested.

## 5. COVERAGE AND DISCUSSION

The current grammar accounts for the core local linearity phenomena of French i.e., auxiliary and clitic order, clitic placement in simple and in complex verb phrases, clitic doubling and interrogative inversions. Unbounded dependencies are catered for without resorting either to threading (UCG), functional uncertainty (Karttunen) or functional composition (Combinatory Categorial Grammar, Steedman 1986). Instead, the issue is dealt with at the lexical level by introducing an embedding valency on matrix verbs. Finally, non local order constraints such as constraints on the distribution of negative particles and the requirement for a wh-constituent to be placed to the left of the verb when a lexical subject is placed to the right (see example (22d)) can also be handled.

Thus, it appears that insights from phrase structure and categorial grammar can be fruitfully put together in a lexical framework. Following GPSG, our formalism does not associate verb valencies with any intrinsic order. An interesting difference however is that LP statements are not used either. This is important since in French, clitic ordering (Bès 1988) shows that order constraints may hold between items belonging to different local trees. Another difference with GPSG is that as in UCG, no explicit statement of feature instantiation principles is required: the feature flow of information is ensured by the concatenation rules. Last but not least, it is worth underlining that our approach (1) keeps the number of combination rules down to 2 (plus a unary deletion rule) and (2) eliminates unjustified lexical ambiguity i.e., ambiguity not related to categorial or semantic information on the other hand.

Though there are - or so we argue - good linguistic reasons for representing verb valencies as a set rather than as a list, it is only fair to stress that this rapidly leads to computational innefficiency while parsing. Typically, given 3 adjacent signs NP1 V NP2 there will be two ways of combining each NP with the verb and thus two parses. In a more complex sentence, so-called "spurious ambiguities" - i.e., analyses which yield exactly the same sign - multiply very quickly. We are currently working on the problem.

## REFERENCES


Bar-Hillel, Yehoshua (1953) A quasi-arithmetical notation for syntactic description. *Language*, 29, 47-58.

Baschung, Karine, Bès, Gabriel G., Corluy, Annick, and Guillotin, Thierry (1986) Auxiliaries and Clitics in French UCG Grammar. *Proceedings of the Third European Chapter of the Association for Computational Linguistics.* Copenhague, 1987, 173-178.

Bès, Gabriel G. (1988), Clitiques et constructions topicalisées dans une grammaire GPSG du français. *Lexique*, 6, 55-81.

Calder, Jonathan, Moens, Marc and Zeevat, Henk (1986) An UCG Interpreter. ESPRIT Project 393 ACORD. Deliverable T2.6, Centre for Cognitive Science, University of Edinburgh.

Gunji T. (1986) Subcategorisation and Word Order. In *International Symposium on Language and Artificial Intelligence,* Kyoto, 1986.

Karttunen, Lauri (1986) Radical Lexicalism. Report n°. CSLI-86-68, Center for the Study of Language and Information, December 1986. Paper presented at the Conference on Alternative Conceptions of Phrase Structure, July 1986, New York.

Steedman, Mark J. (1986) Incremental Interpretation in Dialogue. ESPRIT Project 393 ACORD Deliverable 2.4, Centre for Cognitive Science, University of Edinburgh.

Uszkoreit, Hans (1987) Word order and Constituent Structure in German. CSLI Lecture Notes.

Zeevat, Henk, Klein, Ewan and Calder, Jonathan (1987) An Introduction to Unification Categorial Grammar. In Haddock, Nicholas J., Klein, Ewan and Morrill, Glyn (eds.) *Edinburgh Working Papers in Cognitive Science,* V. 1: *Categorial Grammar, Unification Grammar, and Parsing.*


## APPENDIX. Order constraints

The following matrix represents features in nominative (a) and non-nominative (b) valencies in verbal signs (i.e., they correspond to (21a)), and features in the valency of NP's active signs, lexical NP(c) and wh-NP(d); see (21b). Columns stand for specified slots for both types valencies (see (11)).

| | Lleft | Lright | Lleft | Ligth |
|---|---|---|---|---|
| (a)nom valency | ~wh | i | ~wh | k |
| (b)~nom valency | k | _ | ~wh | _ |
| (c)Lexical NP | V1 | V2 | _ | _ |
| (d)Wh-NP | _ | _ | V1 | V2 |

The matrix express the following constraints (in italics the constituent inducing the constraints).

(a) A lexical subject NP to the left of the verb cannot be

immediately followed by a wh-constituent :
**Jean* quel homme regarde ? (*Jean* = subject)

(b) A lexical subject placed to the right of the verb must be immediately ajacent to it :
*Quel cadeau présente à Marie *Pierre* ?

(c) A wh-subject to the left of the verb forbids a wh-constituent to its immediate right :
**Qui* quel homme regarde ?

(d) There may be no wh-subject to the right of the verb :
*Jean regarde *qui* ?

(e) Lexical non-subject NP's cannot be placed to the left of the verb :
**Marie* Pierre regarde

(f) A wh-NP non-subject to the left of the verb cannot be immediately followed by a wh-constituent :
**Qui* qui regarde ?