



HAL
open science

UCG Grammars; the Control of their Descriptive Adequacy

Gabriel G. Bès, Pierre-François Jurie

► **To cite this version:**

Gabriel G. Bès, Pierre-François Jurie. UCG Grammars; the Control of their Descriptive Adequacy. Technical Documentation, ACORD ESPRIT Project 393, 1989. hal-01143084

HAL Id: hal-01143084

<https://hal.science/hal-01143084>

Submitted on 16 Apr 2015

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.

UCG Grammars; the Control of their Descriptive Adequacy

Gabriel G. Bès, Pierre-François Jurie
Formation doctorale Linguistique et informatique,
Université Blaise-Pascal, Clermont-Ferrand

Bès G. (dir.), *Technical Documentation*, ACORD ESPRIT Project 393,
Laboratoires de Marcoussis, December 8 1989.

Abstract

Several models of grammars were used in the ACORD project (LFG, UCG, GPSG, string grammars and CCG). Only the first two remained as the standard grammar models of the project. Furthermore, the UCG model can be understood as a general framework leaving open choices: these are represented in the English and French grammars in their several and evolving versions.

The justification of the choice of a particular model has been an important concern of the ACORD project, but the issue about the criteria of choice of alternative UCG grammars remains open. It is in this background that the work on the control of grammars grew, in the last part of the project.

Grammars must be checked and validated with observations and data which should be expressed in a both formal and theory neutral way. The control of grammars introduces a descriptive metalanguage, specifying a set of grammatical sentences, to which the set of sentences formalized by a grammar is to be compared.

See also

Pierre-François Jurie, Gabriel G. Bès, "The Control of UCG Grammars",
in Gabriel G. Bès, Thierry Guillotin (eds), *A Natural Language and Graphics Interface: Results and Perspectives from the ACORD Project*, Springer, p. 47-64, 1992.

Gabriel G. Bès. « La phrase verbale noyau en français. » *Recherches sur le français parlé*,
Publications de l'Université de Provence, 1999, n° 15, p. 237-358.

Gabriel G. Bès, Caroline Hagège, *Properties in 5P*, Research Report, GRIL,
Université Blaise-Pascal, 2001.

Chapter 12

UCG grammars ; the control of their descriptive adequacy

*Gabriel G. Bès,¹
Pierre-François Jurie
CLF*

12.1 Scientific Background and Development

Several models of grammars were used in the ACORD project (LFG, UCG, CCG, string grammars and CCG). Only the first two remained as the standard grammar models of the project. Furthermore, the UCG model can be understood as a general framework leaving open choices : these are represented in the English and French grammars in their several and evolving versions (see 3 and 4, but also in other UCG epigona (cf Bès et Gardent 89, Gardent et al. 89, Popowich 89, Sanfilippo 90, and Beaven 90).

The justification of the choice of a particular model has been an important concern of the ACORD project ; Deliverable T1.1 and Deliverable T1.2 discuss about grammar and semantic models respectively. The question of how to choose between particular UCG grammars was a major issue discussed in Clermont-Ferrand parsing meeting of September 87. In general, discussions about choices of formalisms consumed considerable amount of time and energy (see Jurie 88). Furthermore, reviewers of the project and participants to conferences where the ACORD project was presented, always formulated the same question: why two different models of grammars are in the same project ? Is it possible to justify them otherwise than by social pressure and/or academic tradition ?

¹Alphabetical order

No answer has ever been given to these questions. The issue about the criteria of choice of alternative UCG grammars, which manifest a foreseen and confirmed trend to proliferation, remains opened.

It is in this background that the work on the control of grammars grew. This work has BOTH theoretical (see the following) AND practical goals (see 12.4). It was developed in the last part of the project.

Taken for granted that (computational) linguistics is an empirical science, (computational) hypothesis, expressed in the form of grammars, must be checked and validated with observations and data. Although this point seems uncontroversial, little has been done to capture observations in a formal way theoretically neutral with respect more or less explicit choices introduced in particular models of grammars. The work on control of grammars attempts to overcome this situation: it pretends to formulate in a compact way natural language observations, and to check the adequacy of a particular grammar by a calculus on it, in order to evaluate its descriptive adequacy. Though descriptive adequacy is not the unique parameter with respect to which a particular grammar must be selected, it is worthwhile to recall that (computational) linguistics being empirical, no rational for choosing among competing formalisms exists without it.

12.2 Technical description

The control of grammars consists in three parts: (i) a descriptive metalanguage, specifying the set GS of grammatical sentences; (ii) a calculus on grammar, specifying the set FS of formalised sentences; (iii) a comparison of GS and FS .

A descriptive metalanguage (for short metalanguage) is a formalised knowledge about the object language but it differs from the particular grammar models associated with the same object language.

Besides formal properties, we suggest that a grammar, when referring to such objects as LFG UCG GPSG, FUG, HPSG particular grammars are intended to satisfy at least two basic requirements

- association to parsing and generation
- characterisation of NL

By the first of these requirements, we understand that the grammars must be integrated in parsing and generation mechanisms in efficient ways. For fulfilling this requirement,

the models of grammars use “constructive” rules, that is, rules which are not only inferencing rules, but which, in a hidden way, condition the algorithmic processes of parsing and generation.

None of the above is a basic requirement for a metalanguage. A metalanguage is not intended to say something clever about NL. The notions of “insight” and “efficiency”, the latter one related to parsing and generation, are definitely alien to it. It specifies grammatical sentences, but it leaves opened what the significant generalization are and it introduces no choice on algorithms which can be designed to analyse or generate effectively the specified sentences.

The metalanguage consists in two parts

i A formalisation of the lexicon

- (a) A finite set VAL
- (b) A relational structure in a first order language

$LEX = \langle E, C_0...C_n, P_0...P_n, R_0...R_n, \dots \rangle$

which is a model of

a set of axioms T1

[E denotes a set : the set of formalised lexical signs, C_0, C_1 , denote individual constants, $P_0...P_n$ denote unary predicates, $B_0...B_n$ binary predicates ...]

ii A formalisation of the sentence

A formalised sentence is a pair
 $\langle X, Cons \rangle$

where

- (a) X is a string
 $X = \langle X_0, \dots, X_n \rangle$
 - either an element of E
 - or a formalised sentence
- (b) Cons is a binary predicate such that $Cons(i, t) \Rightarrow n \geq i$, and $t \in VAL$
[$Cons(i, t)$ means that the sign which is in the place number i, is interpreted as being in the relation named by t]

Example :

VAL = {nom, obj, ...}

Individual constants : *ne, pas, que, ...*

Unary predicates ; *verb, np, cln, wh, lex, ...*

T1

- The predicates *verb, cln, wh, lex* are disjoint
- $np = Wh \cup lex$
- ...

T2

1. $\exists i \text{ verb}(xi)$
2. $\text{verb}(i) \ \& \ \text{verb}(j) \Rightarrow i = j$
3. $\text{cons}(i, \text{nom}) \ \& \ \text{cons}(j, \text{nom}) \ \& \ \text{np}(i) \ \& \ \text{np}(j) \Rightarrow i = j$
4. $\text{cons}(i, \text{obj}) \ \& \ \text{cons}(j, \text{obj}) \Rightarrow i = j$
5. $\text{cons}(i, \text{nom}) \Rightarrow \text{cln}(xi) \ \text{or} \ \text{np}(xi)$
6. $\text{cons}(i, \text{obj}) \Rightarrow \text{np}(xi)$
7. $\text{cln}(xi) \ \& \ \text{cln}(xj) \Rightarrow i = j$
8. $\exists j \text{ cons}(j, \text{nom})$
9. $\text{verb}(xi) \ \& \ \text{clnom}(xj) \Rightarrow j = i + 1$
10. $\text{verb}(xi) \ \& \ \text{cons}(k, \text{nom}) \ \& \ k < i \Rightarrow i = k + 1$
11. $\text{verb}(xi) \ \& \ \text{cons}(k, \text{nom}) \ \& \ i < k \Rightarrow k = i + 1$
12. $\text{verb}(xi) \ \& \ \text{cons}(k, \text{nom}) \ \& \ \text{np}(xk) \ \& \ i < k \Rightarrow \sim \exists j \text{ clnom}(xj) \ \& \ \text{lex}(xk)$
13. $\text{verb}(xi) \ \& \ \text{clnom}(xj) \ \& \ i < k \Rightarrow \sim \text{wh}(xk)$
14. $\text{verb}(xi) \ \& \ \text{clnom}(xj) \ \& \ k < i \ \& \ \text{cons}(k, \text{nom}) \Rightarrow \sim \text{wh}(xk)$
15. $\text{verb}(xi) \ \& \ \text{cons}(k, \text{nom}) \ \& \ \text{np}(xk) \ \& \ i < k \Rightarrow \exists j[\text{wh}(xj) \ \& \ j < k]$

Axioms	Paraphrasis
(1) + (2)	There exists exactly one place in the sentence which is occupied by a verb
(3)	There exists at most one nominative place occupied by a <i>np</i>
(5)	Only nominative clitics (=cln) or <i>np</i> can occupy a nominative place
(4) + (6)	There exists at most one obj-place and this place (if it exists) must be occupied by a <i>np</i>
(7)	There is at most one place which can be occupied by a cln
(8)	There is at least one nominative place
(9)	The place of cln (cf. it exists) is just after the place of the verb

Axioms	Paraphrasis
(10) + (11)	Nominative places are just next to the verb
(13) + (14)	If there exists a place for <i>cln</i> (inversion 1) then no place after the verb and no nominative place before the verb can be occupied by a <i>wh</i>
(12) + (15)	If a nominative place is after the verb and is occupied by a <i>np</i> (inversion 2) then there is no place for a nominative clitic, the <i>np</i> in the nominative place is <i>lex</i> and there exists a place before the verb which is occupied by a <i>wh</i>

The calculus on grammar is founded on the notion of closure table, defined in the following.

Definition. A closure table is of the form:

$$R_1 A_1 B_1 C_1$$

$$R_2 A_2 B_2 C_2$$

...

$$R_n A_n B_n C_n$$

where the following conditions are satisfied

(a) A_i, B_i, C_i are templates (Categories + features)

(b) R_i is a binary rule

(c) if A_i is a template of a sign X , and

B_i is a template of a sign Y

then there exists a sign Z such that:

$$X + Y \xrightarrow{R_i} Z \text{ and } C_i \text{ is a template of } Z$$

(d) if

$$X + Y \xrightarrow{R} Z$$

then there exists i such that

$$R_i = R$$

A_i is a template of X

B_i is a template of Y and

C_i is a template of Z

The closure table can be used immediately to determine the set of all the strings $X_1 \dots X_n$ of lexical signs such that

$$X_1 + \dots + X_n \stackrel{GR}{\Rightarrow} sent$$

More precisely

definition A closure tree is a tree in which

- (a) the root is the template *sent*
- (b) every vertex *C* has two sons *A* and *B* and there exists a binary rule *R* such that

$$RABC$$

is a line of the closure table (a straightforward extension is needed for unary rules)

- (c) every leaf is lexical

Result

- (a) the set of all the closure trees can be automatically founded from the closure table, and:

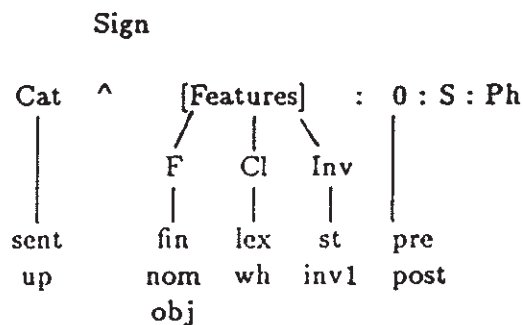
(b) $X_1 + \dots + X_n \stackrel{GR}{\rightarrow} sent$

iff

there exists a closure tree which string of leaves is A_1, \dots, A_n and for every i A_i is a template of X_i

For all presently examined UCG grammars it is in principle possible to construct a closure table. Furthermore, we conjecture that, given the formal properties of UCG the construction of such a table is always possible. The conjecture is based in particular on the fact that UCG grammars (like pure categorical grammars) generate only a finite number of syntactically different signs from a finite number of entries. For grammars which use more intensively type-raising and computational rules, for instance, the situation may be completely different.

Example. The Grammar *G* :



12.3 Relations to other ACORD components

The methodology developed in relation to the grammar can be applied in principle to other ACORD components and particularly to the control of the Resolver and the KB conceptual knowledge.

12.4 Potential of development

Today nobody knows explicitly what the coverage of a parser or a generator is. Demos and current specifications are, at the best, hints indicating roughly what the system can or must perform. When a lot of money is spent in the implementation of grammars and parsers, it is a big disappointment to find in a haphazard way evident but completely unforeseen counter-examples to some system. The point is that when some unexpected counter-examples are established, the reliability of the whole system decreases dramatically. And reliability is *sine qua non* requirement of any technological product: we are not aware of any long-lasting technological market founded on the marketing of an unreliable product. Control of grammars can add reliability to natural language systems.

With an explicit metalanguage it will be possible to specify accurately corpus of linguistic observations and, by the way, to put in perspective achievements of particular models of grammars, evaluating thus what they can effectively do. Control of grammars is thus an important step toward the elaboration of NL standards and grammatical models evaluation and comparison.

Currently particular grammars overgenerate. On the other hand, any non-exclusively natural language system must be prepared to deal with a corrupt input, i.e. conditions on grammaticality must be relaxed in specified situation. It is difficult to foresee how this can be done if the effective descriptive adequacy of particular grammar remains obscure. Control of grammars looks like an unescapable tool for robust natural language understanding.