A descriptive and formal perspective for grammar development

Marie-Laure Guénot, Philippe Blache

▶ **To cite this version:**

Marie-Laure Guénot, Philippe Blache. A descriptive and formal perspective for grammar development. Foundations of Natural-Language Grammar, 2005, Edinburgh, United Kingdom. hal-00134236

HAL Id: hal-00134236
https://hal.science/hal-00134236

Submitted on 1 Mar 2007

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.

A descriptive and formal perspective for grammar development

Introduction. — For many linguistic theories, a grammar is a mechanism making it possible to define a natural language. The idea behind generative theories is more precisely to consider grammars as an enumerative process deriving a language. This idea is still very present, even in non-generative approaches, and a grammar is considered as a device used to check whether an input belongs or not to the language. This conception is very restrictive for many reasons. First, it has a lot of consequences on the way of representing linguistic information which is expressed in order to rule out ungrammatical utterances. This is particularly clear in the Optimality Theory (Prince and Smolensky [1993]) in which constraints (considered as universals) are stipulated precisely in this perspective. Second, considering grammar as a way of defining a language relies on a clear distinction between grammatical and ungrammatical productions. However, we know that such a distinction doesn't fit with the reality of language for many reasons. Corpus linguistics has shown for many years how often inputs can be ill-formed, and this characteristics concerns written as well as spoken productions. A linguist as well as an engineer has to deal with such material either to explain or to treat it. This means that we need an approach taking into account such aspects. Moreover the notion of grammaticality requires a total covering of the language : uncertainty, incompleteness, heterogeneity have no place in such approaches. However, in some cases, there is (almost) no information available at the syntactic level. This is the case for example between core and peripheral elements or in some configurations such as example (1). In this case, no relation (and no grammatical functions) can be given taking only into account the succession of words.

(1) Monday, washing, Tuesday, ironing, Wednesday, rest.

Our conception of grammar. — For these reasons, we think that grammar doesn't have to be restricted as a defining mechanism anymore. We need a more general conception in which grammar is an actual descriptive tool containing information from which the description of an input can be built, whatever its form. This claim is even strongest when taking into account the different domains of linguistic information. The classical way of defining interaction between these domains relies on structure mapping. In other words, the relations between different domains (for example between prosody and syntax) are expressed in terms of relations between parts of the respective structures (for example a prosodic and a syntactic tree). This incremental conception of building meaning actually doesn't work. Things have to be described in a more interactional way making it possible to take into account the fact that meaning is spread over the different linguistic domains. More precisely, meaning is the result of the interaction of elements of information spread over the domains.

Considering this, we think that a grammar is a set of information (that can even be non connected) capable of describing an input, whatever its form. The main role of a grammar is not to build a structure, but to specify the different characteristics of the linguistic objects. In this perspective, each piece of linguistic information has to be taken *per se*, which corresponds to a non holistic view of grammar (see Huddleston and Pullum [2002]). Such an approach is intrinsically monostratal in the sense that all information, including the relations between the different domains, is expressed at a unique level. It is also non-derivational especially because it doesn't necessarily propose a full coverage and can build non-connected structures. Last, it ignores deliberately the question of universality.

Constraints and Constructions. — In our conception, a grammar is then a set of information clusters indicating interaction between different objects. The idea is to bring together different sources of information as soon as they have an identified consequence on meaning. This conception is that of Construction Grammars (noted CxG, see for example Fillmore [1998] or Kay [2002]). We propose in our approach to integrate the CxG framework within the constraint-based perspective of Property Grammars (hereafter PG, see for instance Blache [2005]). From this CxG/PG point of view, a construction is a set of features plus a set of relations represented by means of constraints, as described in figure 1.

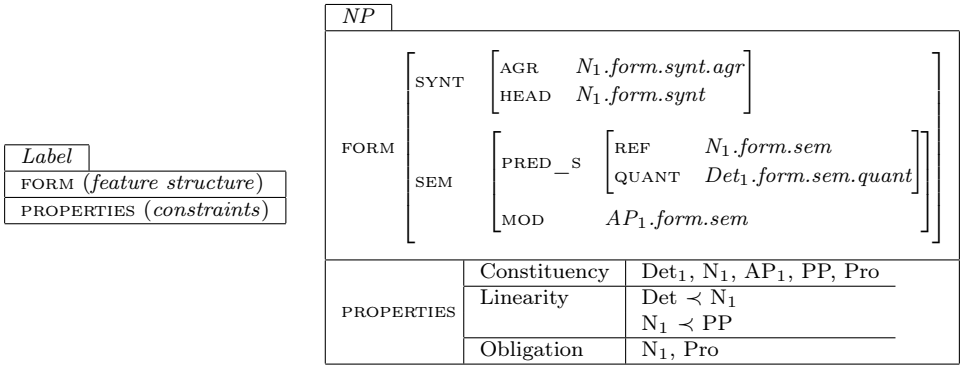

FIG. 1 – (Rough) Description of the NP construction in CxG/GP.

Example. — Let's now consider the example of an analysis that intrinsically has to make use of several levels of information : the treatment of disfluencies. This phenomenon has two main characteristics that allow to consider it as specific. First, disfluencies bear neither syntactic nor semantic role : they do have a linguistic form, but this set of words (or beginning of words) isn't a (syntactic) phrase-structure (most of time it's the beginning of a phrase) ; moreover they aren't a way to express a semantic variation in comparison with a production without disfluency (e.g., *il il* in ex. (2) has exactly the same meaning as one only utterance of *il* would have had).

(2) **il il** a quand-même **un une** fibre pédagogique **assez assez** euh enfin réelle quoi [1]

The second characteristic is that the disfluency phenomenon must be considered through a multi-level feature set : to be complete, it must be described as a set of (morpho-)syntactic (POS, sub-categorization), semantic (sense, reference) and prosodic (intonation, pauses) features.

The disfluency construction in our CxG/GP approach will have the description of figure 2. This construction, which takes place in our mono-stratal resource, shows that a disfluency is a set of objets, consisting in one object x (of any category) and one of several objet(s) x' (that can be different). The "agreement" property tells that the x' has to have a feature set very close to x's one : every value[2] of x' must agree with the corresponding value of x, which means that their entire FORM AVMs must be almost the same.

Disfluency		
FORM [...]		
PROPERTIES	Obligation	x
	Constituency	x', *"euh"*
	Requirement	$x \Rightarrow x'$
	Exclusion	$x \neq$ Coord
	Linearity	$\{ x'$, *"euh"* $\} \prec x$
	Agreement	x'.*feature* $\approx x$.*feature, (feature* $\neq$ *index)*

FIG. 2 – CxG/GP representation of the *disfluency* construction.

The consequences of this description in a analysis process are the following : if all the evaluated properties are satisfied, then a "completed"[3] disfluency is characterized, like *il il* or *assez assez* in example (2). If some of the evaluated properties are violated, then a "modified" disfluency is characterized : *un une* in (2) is an example of a syntactically modified disfluency, where the two determiners only differ in the value of their gender (*un* is masculine and *une* is feminine), whereas in example (3) we can see a semantically modified disfluency, where *un peu* and *pas mal* have exactly the same feature set, except their intensity value.

(3) ils sont pas à l'abri de ça quoi mais c'est **un peu pas mal** d'hypocrisie quand-même à ce niveau-là [4]

Perpectives. — A formal grammar for French is currently being developed in this CxG/PG framework. The aim of this grammar is to represent a set of fine-grained descriptions coming from corpus linguistics. Some of the represented phenomenon are characteristic of spoken productions. This grammar is the result of the theoretical reflections we have presented here, about the nature and the aims of formal representation of linguistic information, and about the way that very heterogeneous problems can be considered through a single homogeneous system.

Références

Philippe Blache. Property grammars : A fully constraint-based theory. In H Christiansen, P Skadhauge, and J Villadsen, editors, *Constraint Satisfaction and Language Processing*. Springer-Verlag, 2005.

Claire Blanche-Benveniste. Syntaxe, choix du lexique et lieux de bafouillage. *DRLAV*, 36-37 :123–157, 1987.

Charles Fillmore. Inversion and constructional inheritance. In *Lexical and Constructional Aspects of Linguistic Explanation*, 1998.

Sandrine Henry and Berthille Pallaud. Word fragments and repeats in spontaneous spoken french. In R. Eklund, editor, *Proceedings of DiSS'03*, pages 77–80, Göteborg University, 2003.

R. Huddleston and Geoffrey Pullum. *The Cambridge Grammar of English Language*. Cambridge University Press, 2002.

Paul Kay. An informal sketch of a formal architecture for construction grammar. *Grammars*, 1(5), 2002.

A. Prince and P. Smolensky. Optimality theory : Constraint interaction in generative grammars. Technical Report RUCCS TR-2, Rutgers Center for Cognitive Science, 1993.

[1] ***he he*** *has anyway* ***a a*** *teaching fibre* ***rather rather*** *hum well actual say.* (approximative word-to-word translation)

[2] Except the one corresponding to the INDEX attribute - which would have meant that x' was x.

[3] For a description of "completed" vs. "modified" disfluencies, cf. for example Henry and Pallaud [2003].

[4] *they are not safe from that but it's* ***a little quite*** *hypocrite anyway concerning this.* (approximative word-to-word translation)