

Property Grammars and Model-Theoretic Syntax: a non-generative view of syntax

Philippe Blache

*LPL-CNRS, Université de Provence
29 Avenue Robert Schuman, 13621 Aix-en-Provence, France*

Linguistic theories as well as NLP techniques are now faced with important questions, in particular concerning robustness and flexibility. In particular, it is necessary to take into consideration the fact that linguistic information is typically spread through different domains such as prosody, pragmatics, syntax, etc. We then need theoretical and practical solutions capable of dealing with this problem. Classical generative approaches cannot give satisfaction in this perspective for several reasons. One is that they rely on the description of language as a set. The question consists then in finding a procedure making it possible to enumerate this set. Derivation represents such a procedure: strings belonging to the language are those generated from the grammar. There are two possible cases: either there exists such a derivation (then the string belongs to the set and is said to be grammatical), either not. In the second case, the string is ungrammatical and nothing else can be said about it. Moreover, for generative theories, syntactic information is represented within a system in which each information has to be interpreted with respect to the entire system: a phrase structure rule is not evaluable in itself. This is one of their main drawbacks. For modern linguistic theories, for example Model-Theoretic Syntax (see [Pullum03]) or Construction Grammars (see [Fillmore98]), language is considered as a possible infinite set of objects. These approaches take into consideration the fact that linguistic information is spread, partial and eventually incomplete. It is then necessary to adopt an integrative view of linguistics in which information is represented in a decentralized way. In this case, each information is autonomous in the sense that it can be evaluated independently from the entire system. Moreover, all linguistic domains, such as prosody, syntax or pragmatics, have to be taken into consideration at the same time. Each domain can bear part of information: interpreting an utterance consists in connecting these different sources of information. This is what is behind the notion of construction which is characterized as a set of properties coming from different domains. We propose in this talk a constraint-based framework for the implementation of these ideas. Each property is here considered as

Email address: pb@lpl.univ-aix.fr (Philippe Blache).

a constraint, the basic parsing mechanism is then constraint satisfaction. In this approach, parsing consists then in finding the different models for a given input. This means that the classical view of parsing as deduction is replaced here by a parsing as satisfaction approach.

Each linguistic domain can be described with general kinds of properties. Concerning syntax, we can for example observe some regular characteristics such as: words follow a certain (partial) order; some words are mutually exclusive in a close context; some words systematically cooccur in a close context; some words cannot be repeated in a close context; the form of some words covaries in a close context; the realization of some words is more facultative than some others, etc. Such information can be expressed in terms of relations between categories or, in other words, by means of constraints, defined as follows: *Linearity* (\prec): linear precedence constraints (e.g. $Det \prec N$); *Dependency* (\rightsquigarrow): dependency relations between categories (e.g. $AP \rightsquigarrow N$); *Obligation* (\mapsto): set of compulsory and unique categories (e.g. $N \mapsto NP$); *Exclusion* ($\not\Leftarrow$): restriction of cooccurrence between sets of categories (e.g. $N[pro] \not\Leftarrow Det$); *Requirement* (\Rightarrow): cooccurrence between sets of categories (e.g. $N[com] \Rightarrow Det$); *Uniqueness* (*Uniq*): set of categories which cannot be repeated (e.g. $Uniq(NP) = \{Det, N, AP, PP, Pro\}$). Constraints are an element of answer for the questions addressed before. In particular, each constraint can be evaluated independently from the others. This means that any kind of information, whatever its domain, can be represented and used independently. In this perspective, a construction is then a constraint system, a grammar being a set of constructions. The parsing mechanism consists then in evaluating the different constraint systems. The state of the system after evaluation, formed with satisfied and violated constraints, is the description of the input. The constraint-based framework proposed here constitutes then a framework for non-generative theoretical models such as Model-Theoretic Syntax or Construction Grammars.

References

- [Blache01] Blache P. & J.-M. Balfourier (2001). "Property Grammars: a Flexible Constraint-Based Approach to Parsing", in proceedings of *IWPT-2001*.
- [Blache04] Blache P. (2004). "Constraints: an operational framework for Construction Grammars", *International Conference on Construction Grammars*.
- [Fillmore98] Fillmore C. (1998) "Inversion and Constructional Inheritance", in *Lexical and Constructional Aspects of Linguistic Explanation*, Stanford University.
- [Pullum03] Pullum G. & B. Scholz (2003), *Model-Theoretic Syntax Foundations: Linguistic Aspects*, ESSLLI lecture notes, Vienna University of Technology.