



Note de lecture sur Philippe Blache, Les Grammaires de propriétés

François Trouilleux

► To cite this version:

François Trouilleux. Note de lecture sur Philippe Blache, Les Grammaires de propriétés. Traitement Automatique des Langues, ATALA, 2003, 44 (2), pp.256-259. <hal-00373465>

HAL Id: hal-00373465

<https://hal.archives-ouvertes.fr/hal-00373465>

Submitted on 6 Apr 2009

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.

Philippe Blache, *Les Grammaires de propriétés*, Hermès Science Publications, 2001, 228 pages, ISBN : 2-7462-0236-0.

par François Trouilleux

trouilleux@yahoo.fr

Les objectifs annoncés des Grammaires de propriétés, sous-titré des contraintes pour le traitement automatique des langues, sont de « faire le point sur l'utilisation des contraintes en linguistique-informatique » et de « proposer un nouveau formalisme pour le traitement automatique des langues ».

L'ouvrage est organisé en trois chapitres principaux. Dans la première partie (*Les théories linguistiques face aux contraintes*), P. Blache examine trois théories linguistiques : GPSG, HPSG et la théorie de l'optimalité, avec un intérêt plus particulier pour l'usage qu'elles font des contraintes. On retiendra que celles-ci se manifestent sous des formes multiples : règles LP, restrictions de cooccurrence de traits, principes des traits de tête, etc., dans GPSG et/ou HPSG, filtrage des *candidats output* dans la théorie de l'optimalité. P. Blache propose également deux implantations de GPSG et HPSG dans le paradigme de la programmation par contraintes, occasion pour l'auteur d'insister sur la difficulté « de décrire et d'implanter l'analyse syntaxique dans ces théories comme un véritable processus de satisfaction de contraintes ».

Philippe Blache propose dans le chapitre central du livre (*Présentation des grammaires de propriétés*) un « nouveau formalisme pour le traitement automatique des langues », qu'il considère « exclusivement fondé sur la notion de contrainte ». Ce formalisme a été conçu à l'origine par Gabriel Bès, qui en a fait une description détaillée dans un long article paru en 1999². P. Blache en donne ici sa version, qu'il présente comme un « prolongement » de celle de Gabriel Bès, mais qui en fait ne s'en distingue pas sensiblement.

Le formalisme des propriétés consiste en un ensemble de formules (les propriétés) portant sur des catégories décrites en termes de structures de traits et qui, interprétées conjonctivement, caractérisent un langage. Sept types de propriétés sont définis : (i) *constituance* (*alphabet du modèle* chez G. Bès) : ensemble des catégories pouvant apparaître dans une suite du langage (désormais S_L); (ii) *unicité* : catégories ne pouvant apparaître qu'une fois dans S_L ; (iii) *noyaux* (ou *obligation*) : ensemble des catégories qui peuvent être noyau de S_L (par définition, toute suite a un et un seul noyau); (iv) *exigence* : si un ensemble de catégories apparaît dans S_L , alors un autre ensemble de catégories doit aussi y apparaître; (v) *exclusion* :

²Gabriel G. Bès. « La phrase verbale noyau en français. » In *Recherches sur le français parlé*, vol. 15, p. 273–358, Université de Provence, 1999. Par la suite (Bès, 1999).

impossibilité de cooccurrence de deux ensembles de catégories dans S_L ; (vi) *linéarité* : contraintes de précédence linéaire entre ensembles de catégories à l'intérieur de S_L ; (vii) *dépendance* (*fléchage* chez G. Bès) : liens entre les catégories apparaissant dans les suites de L .

On peut voir dans ce formalisme une systématisation de la décentralisation de l'information initiée par le formalisme ID/LP de GPSG : les informations exprimées par un ensemble de règles ID seront exprimées par un ensemble de propriétés des types (i) à (v) ci-dessus. Conséquence de cette décentralisation : appartient au langage spécifié les suites qui satisfont *toutes* les propriétés, là où des règles de type ID sont interprétées disjonctivement.

On reconnaîtra au formalisme des propriétés deux grandes qualités : sa parfaite déclarativité et la souplesse de description qu'il permet, les propriétés étant indépendantes les unes des autres et véhiculant chacune une information très spécifique³. À ces deux caractéristiques, mises en avant aussi bien par P. Blache que G. Bès, P. Blache ajoute une troisième, qui est que « les grammaires de propriétés visent la description d'un *input* plutôt que sa grammaticalité », « changement de point de vue par rapport aux approches classiques » sur lequel nous reviendrons plus loin.

La dernière partie du livre (*Implantation : l'analyse par contraintes*) est consacrée à « l'utilisation des grammaires de propriétés pour l'analyse syntaxique automatique ». Philippe Blache propose plusieurs approches, dont la principale vise une analyse approfondie par un processus de satisfaction de contraintes. Le processus proposé est incrémental et de type « generate and test » : on considère toutes les suites possibles et on applique les contraintes sur ces suites (une technique assez proche de celle mise en œuvre dans la théorie de l'optimalité). Pour qui connaît l'article de Gabriel Bès, ce chapitre constitue la partie la plus originale du livre. À quelques imprécisions près, les techniques utilisées sont clairement décrites. On regrettera cependant l'absence d'évaluation des méthodes proposées sur des données conséquentes (absence qui rend purement théoriques les références à leur « efficacité »), ainsi que l'absence totale de référence au système d'analyse implanté par Caroline Hagège⁴ à partir du même formalisme des propriétés.

Dans sa présentation du formalisme des propriétés, P. Blache adopte des points de vue qui méritent discussion, d'autant plus qu'ils s'écartent sensiblement de ceux de G. Bès et que P. Blache ne met pas en perspective ses idées avec celles de son prédécesseur.

Première idée : les propriétés constituent pour P. Blache un « formalisme fondé sur les contraintes » et l'implantation qu'il en propose « repose exclusivement sur un

³À cet égard, (Bès, 1999) montre comment il est aisé d'intégrer différents niveaux de granularité dans la description de la phrase verbale noyau, par ajout ou suppression de propriétés.

⁴cf. Caroline Hagège. *Analyse syntaxique automatique du portugais*. Thèse de doctorat. GRIL, Université Blaise-Pascal, Clermont-Fd, 2000. Par la suite (Hagège, 2000).

processus de satisfaction de contraintes ». On reprochera à l'auteur de ne pas donner à son lecteur les éléments qui lui permettraient d'évaluer cette affirmation, la notion de contrainte restant difficile à cerner tout au long de l'ouvrage. Pour l'usage du terme *contrainte* en linguistique, P. Blache part d'une définition très large ⁵. C'est seulement au fur et à mesure qu'il avance dans le livre que la notion se dessine pour le lecteur, à travers la qualification de tel ou tel aspect d'une théorie comme « contrainte ». Or, nous l'avons évoqué plus haut, les contraintes, dans les théories examinées, sont multiformes. Pour ce qui est du terme *contrainte* en informatique, où l'on peut penser que la notion est mieux circonscrite, P. Blache aide peu le lecteur non initié à bien la cerner. Ce dernier aura le sentiment qu'à peu près tout peut être vu comme une contrainte et il n'en sera que plus perplexe sur l'utilité de cette notion, lorsqu'il constatera que G. Bès ne fait jamais appel à elle dans sa présentation du même formalisme.

Autre idée importante pour P. Blache : l'implantation qu'il propose constitue une « interprétation directe » ⁶ du formalisme des propriétés, ce qui « valide le modèle théorique proposé ». Cette idée contraste avec les objectifs de G. Bès : ce dernier inscrit le formalisme des propriétés dans un paradigme plus général, appelé 5P, et dans lequel il sépare nettement les propriétés (formalisme utilisé pour décrire une langue) des processus de traitement automatique, les premières n'étant pas la source déclarative de ces derniers (cf. Bès, 1999, p. 348) ⁷. L'enjeu est aussi de disposer d'un formalisme qui permette la formulation de descriptions utilisables par différents systèmes ⁸.

Enfin, P. Blache place « au centre de sa conception de l'analyse linguistique » l'idée qu'il faut « abandonner la notion de *grammaticalité* pour celle de *caractérisation* » et signale que « cette modification de point de vue remet en cause une caractéristique fondamentale de la plupart des théories linguistiques : celui de la générativité. » La modularité d'une description faite en terme de propriétés est telle qu'on pourra, confronté à un énoncé agrammatical, assouplir la description par suppression d'une ou plusieurs propriétés de sorte qu'une analyse soit produite pour

⁵« La linguistique fait un usage systématique du terme de *contrainte* pour exprimer les limites au-delà desquelles un objet linguistique sort du domaine étudié (la langue). [...] Dans tous les cas l'idée reste la même : il s'agit de fournir une propriété permettant d'exclure des structures correspondant à des productions n'appartenant pas à la langue. »

⁶Par opposition à une interprétation indirecte, qui consisterait, par exemple, à traduire un ensemble de propriétés en une grammaire hors-contexte, puis à utiliser les implantations développées pour ces dernières.

⁷L'analyseur implanté par Caroline Hagège, déjà évoqué plus haut, met en pratique cette idée ; cf. (Hagège 2000), en particulier le chapitre 6, où est définie une fonction qui permet de passer des propriétés aux « feuilles », structures de données utilisées par l'algorithme d'analyse.

⁸Dans cette perspective, G. Bès et C. Hagège décrivent comment une grammaire HPSG peut être déduite d'un ensemble de propriétés. Cf. Caroline Hagège et Gabriel G. Bès. Encoding and Reusing Linguistic Information Expressed by Linguistic Properties. In *Workshop on Grammar Engineering and Evaluation, COLING 2002*, Taiwan, 2002.

cet énoncé. Que cette modularisation possible des propriétés conduise à abandonner la notion de grammaticalité est cependant sujet à caution. En effet, cette notion est celle par quoi on vérifiera l'adéquation d'une description de la langue avec les observations. Philippe Blache l'abandonne, mais ne dit pas comment les caractérisations produites par une description seront alors évaluées. En ce qui concerne la générativité, on trouvera dans (Bès, 1999) la spécification d'un « générateur de modèles », « outil informatique qui doit permettre de calculer, étant donné un lexique de catégories [...] et un ensemble de propriétés, tous et seulement les modèles qui satisfont ces propriétés ». L'abandon de la générativité par P. Blache n'est donc pas une nécessité et on regrettera que l'auteur n'ait pas poussé plus loin son argumentation sur ce sujet qu'il qualifie lui-même de « fondamental ».

Au vu des quelques éléments de discussion que nous avons abordés ici, il apparaît que nous disposons aujourd'hui de deux présentations très contrastées d'un même formalisme, qui s'avère très intéressant pour sa capacité à exprimer des descriptions modulaires et déclaratives des langues naturelles. En schématisant, nous dirions que le point de vue de P. Blache s'inscrit plutôt dans une problématique du développement informatique, tandis que celui de G. Bès s'inscrit plutôt dans une démarche méthodologique qui vise à inscrire la linguistique dans le paradigme des sciences du réel. Selon l'inclination de son esprit, le lecteur pourra préférer l'une ou l'autre des deux approches. Notre conseil serait qu'il n'en néglige aucune.