

## Une grammaire du français pour une théorie descriptive et formelle de la langue

Marie-Laure GUÉNOT\*

Signes, Université Michel de Montaigne Bordeaux 3

marie-laure.guenot@u-bordeaux3.fr

**Résumé.** Dans cet article, nous présentons une grammaire du français qui fait l'objet d'un modèle basé sur des descriptions linguistiques de corpus (provenant notamment des travaux de l'*Approche Pronominale*) et représentée selon le formalisme des *Grammaires de Propriétés*. Elle constitue une proposition nouvelle parmi les grammaires formelles du français, participant à la mise en convergence de la variété des travaux de description linguistique, et de la diversité des possibilités de représentation formelle. Cette grammaire est mise à disposition publique sur le *Centre de Ressources pour la Description de l'Oral* en tant que ressource pour la représentation et l'analyse.

**Abstract.** In this paper I present a grammar for French, which is the implementation of a linguistic model based on corpus descriptions (notably coming from *Approche Pronominale*) and represented into the *Property Grammars* formalism. It accounts for a new proposition among formal grammars, taking part into the works that aim to promote convergence between the various researchs of descriptive linguistics and the diversity of formal representation possibilities. It is freely available on the *Spoken Data Resource Center* (CRDO), as a representation and analysis resource.

**Mots-clés :** développement de grammaire, ressource pour le TAL, grammaire du français, syntaxe, linguistique formelle, linguistique descriptive, grammaires de propriétés (GP).

**Keywords:** grammar development, resource for NLP, French grammar, syntax, formal linguistics, descriptive linguistics, property grammars (PG).

### 1 Introduction

Nous présentons ici une nouvelle ressource pour le TAL et pour la linguistique descriptive et formelle : une grammaire du français à large couverture s'intéressant à l'écrit et à l'oral, basée sur des positions théoriques originales et formalisée en *Grammaires de Propriétés* (Blache, 2005) (ci-après GP). Cette grammaire a été conçue dans le but de faire davantage converger la variété des descriptions linguistiques et la richesse des possibilités formelles offertes, afin de proposer à la fois d'élargir les points de vue théoriques des grammaires formelles, et de valider les descriptions linguistiques utilisées.

---

\*Je tiens à remercier très chaleureusement Philippe Blache, sans qui cette ressource n'aurait pu voir le jour.

**Une théorie descriptive et formelle.** Le développement de cette ressource est basé sur la mise au point d'une *théorie de la langue*, que l'on a orientée à la fois vers la linguistique descriptive et vers la linguistique formelle.

La linguistique descriptive construit des hypothèses explicatives à partir de l'observation des productions de langue ; celles-ci constituent la matière première de la grammaire. Il s'agit là, en premier lieu, d'opérer un choix parmi les théories, les méthodes, et les descriptions linguistiques que l'on souhaite retenir, puisque l'on sait que celles-ci présentent une variété considérable, aussi bien concernant les phénomènes auxquelles la discipline s'intéresse, la couverture des travaux, et la diversité des approches théoriques sous-jacentes. L'une des conséquences évidentes de cette variété est que toutes les descriptions ne sont pas compatibles entre elles ; en outre, puisque les descriptions ne sont pas systématiquement envisagées sous l'angle de leur cohésion générale, de leur interdépendance, il arrive que certaines d'entre elles répondant pourtant à une même théorie et une même méthodologie, engendrent quand on entreprend de les rassembler en une grammaire une certaine hétérogénéité, voire parfois des contradictions. Partant de cet état de fait, nous avons donc voulu dans notre grammaire *explicitier* les mécanismes mis en œuvre par les descriptions utilisées, et en proposer un ensemble *homogène* et *cohérent*.

La linguistique formelle, quant à celle, propose des modes d'expression logico-mathématique pour représenter les régularités linguistiques et leur fonctionnement, à partir desquels on peut élaborer des *modèles* de grammaire, *i.e.* des squelettes d'organisation des informations et de leur analyse. Or la discipline connaît elle aussi une importante variété des propositions, qui s'explique par le fait que les motivations sous-jacentes aux modèles peuvent être très différentes, par les domaines de recherche qui peuvent en être à l'origine, et par l'ensemble de postulats propre à chaque modèle et en faisant l'originalité. Tout ceci offre au linguiste une remarquable gamme de possibilités de représentation. Toutefois, toutes les pistes envisagées dans le cadre de la linguistique (strictement) descriptive ne sont pas, et dans certains cas même ne *peuvent* pas être représentées formellement, en dépit de la richesse des propositions. Nous avons donc souhaité également pour notre grammaire, proposer un modèle qui *prenne en compte* toute une partie de la linguistique descriptive qui n'a pas encore été formalisée, pour en retour *valider* la pertinence du formalisme linguistique utilisé en montrant qu'il permet de représenter de telles informations.

La *ressource* que nous présentons ici est basée sur un *modèle*, lui-même élaboré à partir d'un ensemble de *positions théoriques* (§2.1) et du choix d'un *formalisme* (§2.2) ; après avoir introduit ceci, nous présenterons les points les plus originaux de la grammaire développée pour le français (§3). Nous aborderons (§4) sa mise à disposition publique et évoquerons d'autres ressources complémentaires à diffuser dans le même cadre. Enfin, nous concluons sur les futurs développements de la grammaire.

## 2 Le modèle

### 2.1 Aspects théoriques

**Une approche non générative.** Là où les approches génératives (*Generative-Enumerative Syntax*), les plus répandues actuellement, définissent une langue comme une *liste* récursivement énumérable d'éléments obtenus à partir d'un nombre fini de règles, et définissent une grammaire comme un *outil* permettant de générer toutes et uniquement les phrases d'une langue, les

théories basées sur des modèles (*Model-Theoretic Syntax*) considèrent la grammaire comme un ensemble de *contraintes* portant sur la structure des expressions de langue, et la langue comme un ensemble par définition infini, tenant ainsi compte de la variabilité des degrés de grammaticalité et d'interprétabilité, ainsi que de la créativité linguistique (Pullum & Scholz, 2001). Pour ces raisons, le modèle de que l'on propose fait partie de cette seconde classe.

**Une grammaire non présomptive.** Nous adoptons une position qui consiste à ne pas anticiper sur la *structure externe* d'une construction au sein de la description de sa *structure interne* (Deulofeu, 2006) : autrement dit, on ne présume pas des relations qu'une construction entretient avec les autres éléments de l'énoncé, uniquement sur la base des relations qu'entretiennent ses constituants. La conséquence de ceci qui tranche le plus avec les autres grammaires formelles actuelles, est que la nôtre n'est *pas lexicalisée* : en effet, s'il est indéniable que le lexique contraint effectivement en partie la structuration syntaxique, comme le font par exemple les restrictions de sélection lexicale ou la valence des objets, nous refusons pour autant de radicaliser la vision lexicaliste et de systématiser l'expression des informations syntaxiques dans le lexique.

**Un traitement non modulaire.** A l'observation des corpus, on constate que l'information y est éparse et inconstante (Blache, 2004) : elle est dispersée à travers les domaines (phonétique, morphologie, syntaxe, sémantique, *etc.*), et pour chaque domaine elle est présente en quantité (de même qu'en qualité) variable. Par exemple, on trouve des indications syntaxiques en plus grande quantité et de meilleure qualité dans (1a), que dans (1b, 1c) où il manque des informations, ou bien dans (1d, 1e) où figurent des écarts à la grammaire, qu'ils soient volontaires et acceptés (1d) ou non (1e).

- (1) a. Deux scientifiques travaillent actuellement sur un matériau qu'ils ont baptisé "claytronics", qui sera composé de nanomachines capables de s'organiser pour reproduire à distance en 3D, et via internet, n'importe quelle forme, y compris la vôtre. (*fr.sci.philo*, 21 septembre 2006)
- b. lundi lavage mardi repassage mercredi repos (Mertens, 1993)
- c. il y a mon frère sa moto le guidon eh ben complètement naze quoi (Cappeau & Deulofeu, 2001)
- d. T'occupe!
- e. c'est parcequ'on est 3 à bosser sur le site et que le troisième à casser son portable alors la com passe mal! (<http://www.onpeutlefaire.com/forum/index.php?showtopic=6164>, 26 juillet 2006)

Notre grammaire est donc conçue comme une ressource unique, rassemblant des constructions provenant de tous les domaines selon une représentation homogène.

**Une analyse multi-dimensionnelle.** Le dernier point théorique concerne les axes de structuration de l'analyse linguistique. Nous pensons, pour ce qui nous concerne, que la structure arborescente telle que définie par (Chomsky, 1957) (et qui équivaut à la plupart des structures de représentation des grammaires formelles, y compris celles de dépendance (Kahane, 2006)) ne suffit pas. En effet, l'une des premières choses que l'on apprend en linguistique est que l'information y est organisée selon deux axes, syntagmatique et paradigmatic (De Saussure, 1916). Cette idée fondamentale a été suivie dans certaines descriptions (Meillet, 1924; Bally, 1965; Perrot, 1994), mais elle est jusqu'à présent demeurée ignorée dans les grammaires formelles : la dimension paradigmatic, considérée *a priori* comme restant « virtuelle » (Dubois *et al.*, 1994), n'est pas prise en compte et ne peut y être représentée. Pourtant il a été montré (Blanche-Benveniste *et al.*, 1990) que des objets produits dans un énoncé peuvent entretenir des

relations s’inscrivant dans cet axe, comme les disfluences et les coordinations. En conséquence, nous avons tenu dans notre modèle à bien considérer l’analyse syntaxique comme s’articulant simultanément selon les deux dimensions syntagmatique et paradigmatique et à permettre l’introduction de constructions des deux sortes.

## 2.2 Aspects formels

Une grammaire<sup>1</sup> en GP se présente sous la forme d’un couple constitué d’un *ensemble de descriptions linguistiques* et d’une *spécification des types de propriétés*.

**Ensemble de descriptions linguistiques.** La première partie se présente sous la forme d’un réseau complexe de descriptions de *constructions* linguistiques (Fillmore, 1985), héritant les unes des autres. Chaque objet *y* est représenté sous la forme d’un double ensemble, de caractéristiques intrinsèques d’une part (*i.e.*, les qualités inhérentes à la construction, p.ex. le fait de porter un genre pour une construction nominale), et extrinsèques d’autre part (*i.e.*, les qualités qui lui viennent de ses constituants, p.ex. l’accord, entre un déterminant et un nom et/ou entre un nom et un adjectif, *etc.*) (Guénot, 2005a). Les informations intrinsèques prennent la forme d’une matrice de traits, et les extrinsèques d’une liste de contraintes : les *propriétés*.

Adj															
HERIT X															
INTR.	<table border="1"> <tr> <td>ID   NATURE</td> <td> <table border="1"> <tr> <td>TYPE</td> <td>Adj</td> </tr> <tr> <td>CATÉGORIE</td> <td>Adj</td> </tr> </table> </td> </tr> <tr> <td>SYN</td> <td> <table border="1"> <tr> <td>CPLT</td> <td> <table border="1"> <tr> <td>RECTEUR</td> <td>①</td> </tr> <tr> <td>DÉPENDANT</td> <td>②</td> </tr> </table> </td> </tr> </table> </td> </tr> </table>	ID   NATURE	<table border="1"> <tr> <td>TYPE</td> <td>Adj</td> </tr> <tr> <td>CATÉGORIE</td> <td>Adj</td> </tr> </table>	TYPE	Adj	CATÉGORIE	Adj	SYN	<table border="1"> <tr> <td>CPLT</td> <td> <table border="1"> <tr> <td>RECTEUR</td> <td>①</td> </tr> <tr> <td>DÉPENDANT</td> <td>②</td> </tr> </table> </td> </tr> </table>	CPLT	<table border="1"> <tr> <td>RECTEUR</td> <td>①</td> </tr> <tr> <td>DÉPENDANT</td> <td>②</td> </tr> </table>	RECTEUR	①	DÉPENDANT	②
ID   NATURE	<table border="1"> <tr> <td>TYPE</td> <td>Adj</td> </tr> <tr> <td>CATÉGORIE</td> <td>Adj</td> </tr> </table>	TYPE	Adj	CATÉGORIE	Adj										
TYPE	Adj														
CATÉGORIE	Adj														
SYN	<table border="1"> <tr> <td>CPLT</td> <td> <table border="1"> <tr> <td>RECTEUR</td> <td>①</td> </tr> <tr> <td>DÉPENDANT</td> <td>②</td> </tr> </table> </td> </tr> </table>	CPLT	<table border="1"> <tr> <td>RECTEUR</td> <td>①</td> </tr> <tr> <td>DÉPENDANT</td> <td>②</td> </tr> </table>	RECTEUR	①	DÉPENDANT	②								
CPLT	<table border="1"> <tr> <td>RECTEUR</td> <td>①</td> </tr> <tr> <td>DÉPENDANT</td> <td>②</td> </tr> </table>	RECTEUR	①	DÉPENDANT	②										
RECTEUR	①														
DÉPENDANT	②														
Majeur	①(Adj-q ∨ Adj-n ∨ V-p ∨ N ∨ SPrep ∨ Adv-d)														
Mineur	②SPrep														
Unicité	②														
Exigence	①.Adj ⇒ ②														
Adjacence	① ↔ ②														
Accord	①.GENRE ↔ ②.GENRE ①.NOMBRE ↔ ②.NOMBRE														

FIG. 1 – Représentation d’une construction en GP : l’Adjectif construit (Adj).

La Figure 1 montre la forme que revêt une construction en GP : il s’agit de l’exemple de la construction adjectivale dans la grammaire qui nous intéresse. Elle est constituée d’un cartouche permettant de l’identifier dans la grammaire, qui indique l’étiquette de la construction (Adj) ainsi que son héritage (ici, Adj hérite de la construction étiquetée X). On y voit ensuite les deux blocs de caractéristiques : en haut la structure de traits rassemblant les informations intrinsèques, indiquant p.ex. le fait qu’il s’agit d’une construction de CATÉGORIE adjectivale ; en bas la liste des propriétés représentant les informations extrinsèques, comme p.ex. l’indication selon laquelle les deux constituants possibles de la construction, s’ils sont présents, doivent être

<sup>1</sup>Ou un module dans le cadre d’un modèle modulaire. P.ex., notre modèle n’a qu’une grammaire, mais une GUST (Kahane, 2002) a autant de grammaires que de modules.

contigus (quel que soit leur ordre relatif), ou le fait que l'objet Mineur doit s'accorder en genre et en nombre avec le Majeur, si toutefois il possède les traits correspondants.

**Spécification des types de propriétés.** La seconde partie de la grammaire rassemble la définition des contraintes pouvant s'établir entre les objets de la grammaire, dans un énoncé. A chaque *ensemble de descriptions* correspond une *spécification* donnée, indiquant les relations que la grammaire observe, les objets qu'elles mettent en relation (leur nombre, leur forme, etc.), et la façon dont elles sont évaluées (leurs conditions d'évaluation et de satisfaction).

TYPE	DÉFINITION INFORMELLE	EXEMPLE
Majeur	Influe sur la nature	$M(Det)$ dans SN
Mineur	N'influe pas sur la nature	$m(SPrep)$ dans SN
Unicité	Objets uniques	Objet direct dans SV
Exigence	Obligation de cooccurrence	$SPrep \leftrightarrow N$ dans SN
Exclusion	Restrictions de cooccurrence	$Clitique\ accusatif \not\leftrightarrow SN\ objet$ dans SV
Précédence	Ordre relatif	$Det \prec Nom$ dans SN
Adjacence	Contiguïté	$Adj \leftrightarrow Nom$ dans SN
Accord	Correspondance de traits	$Nom.genre \rightsquigarrow Adj.genre$ dans SN

FIG. 2 – Spécification des types de propriétés pour le modèle de grammaire présenté.

Par exemple, le modèle que nous avons défini contient dans sa spécification les types de propriétés de la Figure 2 : les deux premières propriétés (Majeur et Mineur) expriment les relations de constituance ; les trois suivantes (Unicité, Exigence, Exclusion) les relations de cooccurrence ; les deux suivantes (Précédence, Adjacence) les relations de position ; enfin, la dernière (Accord) exprime une relation de dépendance.

Il est possible de faire varier cet ensemble : dans d'autres grammaires<sup>2</sup> les types de propriétés ne sont pas les mêmes, et même les propriétés à la dénomination identique n'ont pas nécessairement le même fonctionnement (p.ex. la Précédence peut être définie comme immédiate, ou non).

Il est à noter que GP ne fait pas usage d'informations grammaticales au-delà de la grammaire (*i.e.*, elle n'exprime pas de « principes » comme en HPSG ou en TAG) : avec le jeu des héritages et l'indépendance d'expression des types de propriétés par rapport aux constructions, toutes les informations sont exprimées au sein même de la grammaire et sont traitées selon une même procédure.

### 3 La grammaire

Le *modèle* introduit dans la section précédente a servi de structure au développement de la *grammaire* que nous présentons dans cette section. Dans son état actuel, celle-ci est constituée d'un ensemble de descriptions de constructions syntaxiques du français. Il ne s'agit pas là de son contenu définitif puisque celle-ci a vocation à être complétée, aussi bien en ce qui concerne la finesse des descriptions syntaxiques que l'ajout d'informations provenant d'autres domaines.

<sup>2</sup>Cf. par exemple la grammaire développée pour les besoins de la campagne EASY (Balfourier *et al.*, 2005), ou les extraits de grammaires basées sur d'autres modèles dans (Guénot, 2006a).

Elle constitue toutefois d'ores et déjà une ressource à large couverture pour le français, intégrant notamment des descriptions de phénomènes tels que les SN sans tête nominale (2a), un traitement original des clitiqes, les coordinations elliptiques ou de catégories différentes (2b), ou même les disfluences (2c).

- (2) a. le vrai, le moi, le boire, les avants, le dessus, des mais, un sans faute, un je ne sais quoi, ...  
 b. Un pas de plus et tu es mort  
 c. un jeune homme me l'avait **dans lancé lancé dans** la figure

Plutôt que d'en présenter un extrait nécessairement très limité compte tenu de la place, nous préférons proposer ici une description des grandes originalités de notre ressource. L'ensemble de la grammaire prend la forme d'une hiérarchie de constructions : elle comporte une construction *racine*, dont héritent tous les autres objets. Les constructions qui en héritent peuvent être de deux sortes : *syntagmatique* (§3.1) ou *paradigmatique* (§3.2). L'une des conséquences de cette différenciation est que l'on a pu ainsi proposer une description paradigmatique des disfluences et des coordinations (§3.3), ce qui constitue l'une des plus grandes originalités de notre grammaire formelle.

### 3.1 Redéfinition des constructions syntagmatiques

Les constructions de la grammaire sont basées sur une redéfinition du *syntagme*, elle-même provenant d'une réflexion critique sur la notion de *tête* : là où la lecture classiquement adoptée de Bloomfield (1961) est que les syntagmes peuvent ou non porter une tête selon qu'il sont endocentriques ou exocentriques (Pollard & Sag, 1994), nous posons pour notre part que tous les syntagmes ont une tête, et que c'est la *portée* de l'influence de cette tête qui est variable en fonction de l'endocentricité ou exocentricité de la construction. En effet le texte de Bloomfield peut être compris de manières assez différentes (Zwicky, 1985) et l'on peut également en conclure que tout syntagme, qu'il soit endocentrique ou exocentrique, porte une tête c'est-à-dire un constituant dont la présence influe directement sur la nature de la construction. Toutefois, cette influence a une portée plus ou moins large suivant les autres constituants de la construction (notamment, suivant leur fonction au sein du syntagme).

**Syntagmes endocentriques.** Les constructions endocentriques occupent, « *globalement* » comme le disait Bloomfield, le même paradigme que leur tête<sup>3</sup>. Elles portent donc la CATÉGORIE ainsi que le TYPE de leur TÊTE : ainsi, des constructions syntaxiques telles que (3) seront de la même CATÉGORIE (*Adj*) et du même TYPE (*Adj*) que leur TÊTE (resp. *provocateur*, *rouge* et *vert*).

- (3) a. ...Thierry était **légèrement provocateur**  
 b. Maturité du potiron « **Rouge vif** d'Etampes »  
 c. d'un beau **vert émeraude**

De même, les constructions nominales de (4) et les verbales de (5) sont endocentriques et portent la CATÉGORIE et le TYPE de leur TÊTE (resp. *N* et *V*).

<sup>3</sup>On dit « globalement » parce que parce que pour l'instant on s'accorde à employer la définition de la notion de paradigme de Saussure (1916), tout en concédant qu'à l'observation des corpus on constate qu'elle mérite une redéfinition qui soit plus claire et surtout plus rigoureuse.

- (4) a. un **livre ennuyeux**  
b. une **tarte maison**  
c. la **personne venue me remplacer**
- (5) a. l'homme **a pris** une cigarette  
b. **il ne mange pas** mais grossit tout de même  
c. Pierre **lui avait fait donner** un cadeau

**Syntagmes exocentriques.** Les constructions exocentriques portent bien le TYPE de leur TÊTE, mais pas sa CATÉGORIE. C'est le cas par exemple des Syntagmes Nominaux (2a, 6) et des Syntagmes Verbaux (7) qui sont de TYPE *N* ou *V* mais de CATÉGORIE *SN* ou *SV*.

- (6) a. Le petit chien de ma grand-mère  
b. on peut être en désaccord avec **les ceusses qui gèrent l'Usenet français** sans user d'un tel vocabulaire qui n'apporte rigoureusement rien au débat (Fr.usenet.abus.d, 28 septembre 2000)
- (7) a. l'homme **a pris une cigarette**  
b. je sais où je dois aller  
c. Pierre **lui avait fait donner un cadeau**

Notons que d'après ceci, par exemple, il n'y a pas de Syntagmes Adjectivaux dans la grammaire, puisqu'il n'existe pas en français de syntagmes qui soient de TYPE adjectival mais de CATÉGORIE différente (en d'autres termes, qui soient de TYPE adjectival mais qui occupent un paradigme différent de l'adjectif lexical).

## 3.2 Introduction des constructions paradigmatiques

La plupart des grammaires admettent (et nous en faisons partie) que certaines constructions n'ont pas de tête ; mais dans le cas habituel, la différence entre constructions à tête et construction sans tête<sup>4</sup> les mène à opposer syntagmes endocentriques (avec tête) et exocentriques (sans tête ou à tête particulière). Or on vient de voir que nous adoptons une position différente puisque nous posons que tous les syntagmes ont bel et bien une (et une seule) tête (régulière). Néanmoins, nous admettons également que toutes les constructions n'ont pas de tête, ce qui signifie que dans la grammaire, toutes les constructions syntaxiques ne sont pas des syntagmes.

En effet, en plus des syntagmes on définit un autre type de construction syntaxique : les *constructions paradigmatiques*. Celles-ci sont caractérisées par le fait que leurs constituants n'entretiennent pas des relations hypotactiques, à la différence des syntagmes (où les relations, quelles qu'elles soient, se font hiérarchiquement entre un objet *recteur* et un *dépendant*), et par conséquent ne portent pas de tête.

## 3.3 Formalisation des disfluences et traitement original des coordinations

Deux grands types de constructions font partie des paradigmes : les *coordinations* et les *disfluences*. Cette position descriptive est basée sur les travaux de l'Approche Pronominale

<sup>4</sup>Ou alors, suivant les variantes possibles de la définition de la notion de tête, faisant l'objet de classes spéciales telles que les « *têtes faibles* » d'Abeillé (2003) par exemple.

(Blanche-Benveniste, 1987), dont nous avons proposé la description, la représentation formelle et l'intégration à notre grammaire.

Il s'agit ici aussi d'une proposition nouvelle et originale, sous deux aspects. D'une part, les disfluences ne sont quasiment jamais formalisées dans les grammaires actuelles ; et quand elles le sont, alors il n'est pas tenu compte de leurs caractéristiques et de ce qu'il a pu être montré de leur fonctionnement et de leurs apports à l'interprétation, ce qui implique des pertes d'information et un manque de rigueur, ainsi qu'une absence de consensus dans leur traitement (Guénot, 2005b). D'autre part, la fait de ne pas considérer les coordinations comme des syntagmes est un point de vue tout à nouveau en linguistique formelle. Nous avons montré (Guénot, 2006a) l'intérêt de cette approche en en illustrant le fonctionnement simple et efficace sur des exemples variés tels que les coordinations mettant en jeu des coordonnés de catégories différentes (2b) ou des coordonnants non standard (8a), ou encore les formes elliptiques (8b).

- (8) a. Il a été habiter à côté de chez Rosalie **que** Rosalie elle savait pas (Deulofeu, 1999)  
 b. Pierre aime le cinéma et Marie le théâtre

## 4 Mise à disposition

La grammaire que nous venons de présenter est mise à la disposition publique sur le *Centre de Ressources pour la Description de l'Oral*<sup>5</sup>, en tant que ressource pour la description et le traitement des données. Elle fait l'objet d'une licence *Creative Commons* (Paternité - Partage des conditions initiales à l'identique<sup>6</sup>). Elle sera régulièrement mise à jour à chaque étape validée de son développement. Elle est pour l'instant présentée sous une forme textuelle, qui sera prochainement complétée par son équivalent électronique<sup>7</sup>.

**Autres ressources.** La diffusion de cette grammaire est une première étape de mise à disposition de ressources GP pour le TAL ; en effet nous ferons suivre à cette grammaire la diffusion du modèle lui-même (sous la forme d'un squelette de développement de grammaire) afin de permettre, le cas échéant, son utilisation pour le développement d'autres grammaires (p.ex., portant sur d'autres langues et/ou mettant en œuvre d'autres positions théoriques).

Nous ajouterons également à cela la grammaire que nous avons développée dans le cadre de la campagne EASY (Paroubek, 2005), représentée dans les mêmes formats. Celle-ci est également formalisée en GP mais est basée sur un modèle théoriquement éloigné de notre présente proposition (Balfourier *et al.*, 2005).

## 5 Conclusion

Nous avons présenté ici nous nouvelle ressource pour le TAL : une grammaire du français à large couverture qui s'intéresse à la fois aux phénomènes écrits et oraux. Elle est basée sur

<sup>5</sup><http://crdo.fr>. Le service étant en cours d'élaboration et la grammaire n'y étant à l'heure actuelle pas encore disponible, elle est pour l'instant téléchargeable à l'adresse <http://mlguenot.googlepages.com/Grammaire.pdf>.

<sup>6</sup><http://creativecommons.org/licenses/by-sa/2.0/fr/>.

<sup>7</sup>En l'espèce d'un document XML généré à partir du même source L<sup>A</sup>T<sub>E</sub>X que celui qui est à l'origine du PDF.

un ensemble de positionnements théoriques qui constitue une nouveauté pour la linguistique formelle (non-généralisme, non-lexicalisation, non-modularité et multi-dimensionnalité), et est représentée sous la forme de *Grammaire de Propriétés*.

Elle a vocation à participer à la mise en convergence de la variété considérable des descriptions linguistiques de corpus et des possibilités formelles, apportant aux premières une validation par leur explicitation et leur régularisation (Chomsky, 1957; Pollard & Sag, 1994), et aux secondes un élargissement des phénomènes pris en considération.

Dans son état actuel, elle est vouée à être affinée et complétée, aussi bien concernant la description des phénomènes syntaxiques (on y affinera progressivement le contenu en enrichissant la hiérarchie jusqu'à proposer des descriptions de figements), que l'ajout d'informations provenant d'autres domaines (sémantique, prosodie, discours, voire descriptions multimodales (Guénot & Bellengier, 2004)). Le modèle sur lequel elle se base a été conçu afin que de tels ajouts soient facilités par la forme même de la grammaire.

Elle est mise à la disposition du public afin d'en permettre la diffusion, l'utilisation, la comparaison, la complétion avec la participation de la communauté. Ceci constitue une première étape dans la mise à disposition d'un ensemble de ressources GP pour le TAL.

## Références

- ABEILLÉ A. (2003). A lexicalist and construction-based approach to coordinations. In S. MUELLER, Ed., *Proceedings of the HPSG03 Conference*, p. 5–24, Michigan State University, East Lansing: CSLI Online Publications.
- BALFOURIER J.-M., BLACHE P., GUÉNOT M.-L. & VANRULLEN T. (2005). Comparaison de trois analyseurs symboliques dans une tâche d'annotation syntaxique. In *Actes de TALN 2005 - Workshop EASY*, Dourdan, France.
- BALLY C. (1965). *Linguistique générale et Linguistique française*. Paris: Leroux.
- BLACHE P. (2004). Constraints: an operational framework for construction grammars. In *Proceedings of the 3rd International Conference on Construction Grammars (ICCG3)*, Marseille, France.
- BLACHE P. (2005). Property grammars: A fully constraint-based theory. In H. CHRISTIANSEN, P. SKADHAUGE & J. VILLADSEN, Eds., *Constraint Satisfaction and Language Processing*. Springer.
- BLANCHE-BENVENISTE C. (1987). Syntaxe, choix du lexique et lieux de bafouillage. *DR-LAV*, **36-37**, 123–157.
- BLANCHE-BENVENISTE C., BILGER M., ROUGET C. & EYNDE K. V. D. (1990). *Le français parlé: Etudes grammaticales*. Sciences du langage. Paris: CNRS Editions.
- BLOOMFIELD L. (1961). *Language*. New York: Holt.
- CAPPEAU P. & DEULOFEU H.-J. (2001). Partition et topicalisation : *il y en a* "stabilisateur" de sujets et de topiques indéfinis. *Cahiers de praxématique*, **37**.
- CHOMSKY N. (1957). *Syntactic Structures*. la Hague: Mouton.
- DE SAUSSURE F. (1916). *Cours de linguistique générale*. Paris: Payot. réédition de 1990.
- DEULOFEU H.-J. (1999). Problèmes méthodologiques de l'analyse morphosyntaxique de *que* en français contemporain. *Recherches sur le Français Parlé*, **15**.

- DEULOFEU H.-J. (2006). Les consécutives construites avec *tellement* ont-elles une syntaxe scalaire ? In *La scalarité: autant de moyens d'expression, autant d'effets de sens*, Bruxelles, Belgique.
- DUBOIS J., GESPIN L., GIACOMO M., MARCELLESI C., MARCELLESI J.-B. & MÉVEL J.-P. (1994). *Dictionnaire de linguistique et des sciences du langage*. Paris: Larousse.
- FILLMORE C. (1985). Syntactic intrusions and the notion of grammatical construction. *BLS*, **11**, 73–86.
- GUÉNOT M.-L. (2005a). Des constructions à l'interface entre lexique et grammaire. In S. KAHANE, Ed., *Journée de l'Atala Interfaces Lexique-Grammaire et lexiques syntaxiques et sémantiques*, Paris.
- GUÉNOT M.-L. (2005b). Parsing de l'oral: traiter les disfluences. In *Actes de TALN 2005*, p. 323–332, Dourdan, France.
- GUÉNOT M.-L. (2006a). *Eléments de grammaire du français pour une théorie descriptive et formelle de la langue*. Thèse de doctorat, Université de Provence, Aix-Marseille I.
- GUÉNOT M.-L. (2006b). La coordination considérée comme un entassement paradigmatique: description, représentation et intégration. *Actes de TALN 2006 - Cahiers du Cental*, **2**(1), 178–187.
- GUÉNOT M.-L. & BELLENGIER E. (2004). Quelques principes pour une grammaire multimodale du français. In B. BEL & I. MARLIEN, Eds., *Actes de RECITAL 2004*, p. 51–60, Fès, Maroc.
- KAHANE S. (2002). *Grammaire d'unification sens-texte: Vers un modèle mathématique articulé de la langue naturelle*. Document de synthèse de l'habilitation à diriger les recherches, Université Denis Diderot, Paris VII.
- KAHANE S. (2006). On the status of phrases in head-driven phrase-structure grammar: Illustration by a totally lexical treatment of extraction. In A. Polguère (ed.). *Benjamins*. A paraître.
- MEILLET A. (1924). *Traité de grammaire comparée des langues indo-européennes*. Paris: Champion.
- MERTENS P. (1993). Accentuation, intonation et morphosyntaxe. *Travaux de Linguistique*, **26**.
- PAROUBEK P. (2005). EASY: Campagne d'évaluation des analyseurs syntaxiques. In *Actes de TALN 2005 - Workshop EASY*, Dourdan, France.
- PERROT J. (1994). Eléments pour une typologie des structures informatives. In *Mémoires de la Société de Linguistique de Paris, tome II: La phrase: Énonciation et information*, p. 13–26. Leuven: Peeters.
- POLLARD C. & SAG I. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.
- PULLUM G. & SCHOLZ B. (2001). On the distinction between model-theoretic and generative-enumerative syntactic frameworks. In P. DE GROOTE, G. MORRILL & C. RETORÉ, Eds., *Lecture Notes in Artificial Intelligence - Proceedings of the 4th LACL International Conference*, p. 17–43, Berlin: Springer.
- ZWICKY A. (1985). Heads. *Journal of Linguistics*, **21**, 1–29.