

# CARNAVAL : une base de données de régimes généralisés

Évelyne Arbour Barbaud

OLST – Université de Montréal  
evelyne.arbour.barbaud@umontreal.ca

## 1 Introduction

Lorsque les grammaires traitent des noms et de leurs divers dépendants, ceux-ci sont souvent classifiés selon leur catégorie grammaticale ou partie du discours (proposition, syntagme nominal, syntagme prépositionnel avec préposition sélectionnée – *de*, *envers*, *à cause de* ou autre –, etc.). Diverses théories linguistiques reconnaissent d'ailleurs l'utilité de cette technique dans la description du lexique (voir par exemple la c-sélection de la grammaire générative). On peut observer que cette méthode est parfois appliquée en lexicologie/lexicographie. Toutefois, en observant les grands dictionnaires, on se rend compte que lorsque cela est fait, on précise surtout certaines caractéristiques sémantiques des dépendants (qqn, qqch., etc.). Une autre pratique courante est de traiter du comportement syntaxique de la lexie vedette. Pour les verbes, on utilise la plupart du temps des étiquettes de type « transitif, intransitif, pronominal, etc. », ce qui permet de savoir par exemple qu'un verbe donné s'utilise avec ou sans objet direct. L'usage de telles étiquettes, bien que souvent utilisées avec les lexies verbales, n'est pas pratique courante dans la description du comportement syntaxique des unités lexicales des autres parties du discours. En ce qui concerne les noms, la précision quant à la nature de leurs compléments est souvent uniquement déductible des exemples donnés dans les articles.

Évidemment, toutes les techniques mentionnées ci-haut permettent de présenter de précieuses informations au sujet des dépendants syntaxiques des lexies décrites. Cependant, lorsqu'elles sont présentes dans le dictionnaire, de telles informations servent surtout à distinguer différents sens ou différents usages et ne sont pas fournies automatiquement. De plus, l'ensemble de ces pratiques ne semble pas permettre de rendre compte efficacement de la relation sémantique qui existe entre le nom gouverneur et ses dépendants. En lexicographie générale, on ne mentionne que de manière sporadique si un dépendant donné est un argument ou un modificateur du gouverneur, si un complément est facultatif ou obligatoire, etc. Il est rare que l'on précise systématiquement, surtout dans le cas des noms, quels sont les actants sémantiques de la lexie vedette et la façon dont chacun d'eux s'exprime dans une phrase vis-à-vis de celle-ci.

De plus, les descriptions en termes de simples étiquettes morphologiques (« *de N* » par exemple) ne permettent pas de rendre compte des différences entre des constructions superficiellement identiques qui possèdent pourtant des propriétés syntaxiques bien différentes. Pour illustrer ce propos, voici divers exemples où un dépendant en *de N* est gouverné par un nom (en gras) :

- |   |   |
|---|---|
| (1) [ <i>le</i> ] <b>chat</b> <i>de Marie</i>           | 'le chat qui appartient à Marie'                          |
| (2) [ <i>la</i> ] <b>dispute</b> <i>de Paul et Lili</i> | 'la dispute entre Paul et Lili'                           |
| (3) [ <i>une</i> ] <b>ceinture</b> <i>de pantalon</i>   | 'une ceinture qui se porte généralement avec un pantalon' |
| (4) [ <i>un</i> ] <b>échange</b> <i>de joueurs</i>      | 'un échange qui implique des joueurs'                     |
| (5) [ <i>cet</i> ] <b>imbécile</b> <i>de Gustave</i>    | 'Gustave, qui est un imbécile'                            |
| (6) [ <i>la</i> ] <b>ville</b> <i>de Paris</i>          | 'la ville dont le nom est Paris'                          |
| (7) [ <i>l'</i> ] <b>anniversaire</b> <i>de Gil</i>     | 'l'anniversaire de la date où Gil est né'                 |

Évidemment, les complémentations en *de* N ci-haut, bien que superficiellement identiques, ne sont pas de même nature. Elles impliquent des relations syntaxiques différentes puisque les constructions ont des propriétés « cachées » différentes. Par exemple, les *de* N des exemples (1), (2), (4) et (7) peuvent être remplacés par des déterminants possessifs (*son chat, leur dispute, [on a procédé à] leur échange, son anniversaire*), ce qui n'est pas le cas des exemples en (3), (5) et (6).

Conséquemment, nous désirons réfléchir sur un moyen rendant la désambiguïsation des constructions de complémentation possible. Dans une optique d'encodage de valence, un des moyens envisageables est d'inscrire la relation syntaxique précise qui permette l'expression des actants. Dans le cadre de notre projet, divers tests syntaxiques (comme la possibilité pour le syntagme du dépendant d'être remplacé par un déterminant possessif) serviront à l'identification des relations syntaxiques régies qui existent en français. De plus, leur caractérisation sera faite entre autres grâce aux sens qu'elles traduisent. Comme notre démarche est avant tout lexicographique, nous laisserons de côté dans cette présentation plusieurs travaux qui traitent entre autres des noms et des différents compléments du nom (Tesnière, 1959; Grevisse & Goose, 2007; Riegel et al., 2009; Knittel, 2009; etc.) ainsi que d'autres qui ont été menés afin de décrire certaines constructions particulières du français. Nous pouvons par exemple mentionner différents travaux ayant étudié les « noms d'insultes » comme dans *cet imbécile de Gustave* (Gaatone, 1988; Larrivée, 1994; Flaux & Van de Velde, 2000; Tellier & Valois, 2003). Cela n'implique aucunement que ces travaux (et d'autres) n'ont pas pu aider dans l'identification des propriétés pertinentes à l'identification de nos relations, ni à notre travail de caractérisation en général. Nous présenterons ici de façon sommaire les résultats auxquels nous sommes parvenus.

## 2 Cadre et objectifs de la recherche

Bien sûr, différentes théories linguistiques ont fait des propositions intéressantes en ce qui concerne les descriptions lexicographiques. On pourra mentionner, par exemples, la HPSG (Sag & Pollard, 1989) et la *conceptual semantics* (Jackendoff, 1990). Toutefois, bien que de telles théories aient réfléchi sur la place qui doit être faite aux informations syntaxiques dans les descriptions lexicales, on voit rarement un réel travail lexicographique découler de ces approches. Nous tenons à souligner qu'il existe quelques projets de recherche en lexicographie qui ont cherché à explorer et améliorer les pratiques existantes et qui en sont ainsi arrivés à des résultats pertinents en ce qui concerne la modélisation des régimes lexicaux. On pourra entre autres penser à : FrameNet (Ruppenhofer et al., 2006) qui se base sur la *frame semantics*, le DICOVALENCE (Van den Eynde & Mertens, 2003) qui se base sur l'approche pronominale.

La recherche que nous présentons ici s'inscrit dans le cadre du projet de base lexicale du français DiCo (acronyme pour *Dictionnaire de combinatoire*) (Polguère, 2000) à l'Observatoire de Linguistique Sens-Texte de l'Université de Montréal (OLST). Dans le DiCo, chaque fiche lexicographique comporte les dérivations sémantiques qui lient la lexie décrite à d'autres lexies de la langue et les collocations qu'elle contrôle. Cette description s'accompagne d'une modélisation des structures syntaxiques régies par la lexie et d'une modélisation de son sens, sous forme d'étiquetage sémantique. Notre recherche s'intéresse plus précisément au développement de la section syntaxique des fiches lexicographiques ainsi qu'à la généralisation des informations qu'elle contient. Ce travail s'inscrit dans le cadre théorique de la lexicologie explicative et combinatoire (LEC) (Mel'čuk et al., 1995), la branche lexicologique de la théorie Sens-Texte (TST), ainsi que dans l'approche de syntaxe de dépendance postulée par cette théorie.

Présentement, les fiches du DiCo contiennent une modélisation des structures syntaxiques régies par la lexie, mais celle-ci demeure assez sommaire. Pour l'instant, on n'y indique pas explicitement les relations syntaxiques liant les actants à la lexie vedette. On utilise ce qu'on pourrait qualifier de notation abrégée. Le premier objectif de cette recherche est donc de vérifier les informations actuellement compilées et d'affiner les descriptions des régimes lexicaux des unités lexicales nominales (pour une présentation du schéma de régime en TST voir Miličević, 2009). Il s'agit d'une amorce d'un travail qui sera éventuellement complété pour les autres principales parties du discours (verbe, adjectif

et adverbe). Dans un deuxième temps, nous visons la création d'une base de données des différents régimes lexicaux du français. Comme nous venons de le mentionner, le corpus de départ pour l'élaboration de cette base sera les lexies nominales actuellement décrites dans le DiCo. La description rigoureuse et uniforme des régimes individuels de chacune de ces lexies permettra de factoriser les informations sur les régimes dans une base de régimes généralisés que l'on nommera CARNAVAL (VAL suggérant *valence*).

### 3 Dépendances Syntaxiques de Surface

Comme un des principaux objectifs de cette recherche est d'ajouter la spécification des relations syntaxiques précises à la description des régimes lexicaux du français, la première étape du travail est d'établir une liste de ces relations. Pour ce faire, nous avons utilisé la notion de relations syntaxiques de surface [RelSyntS] de la TST et nous avons débuté la construction de notre liste de RelSyntS régies par les noms du français selon une méthode développée dans Iordanskaja et Mel'čuk (2009). Nous présentons un peu plus loin les relations (qui, par volonté de précision, ont plutôt été identifiées comme étant des dépendances syntaxiques de surface [DépSyntS]) que nous avons décrites jusqu'à maintenant ainsi que leurs propriétés définitoires. Notons que deux dépendances qui sont typiquement non régies ont été incluses car, dans quelques cas particuliers, elles sont nécessaires à la description des régimes lexicaux.

Pour débiter le travail de description, il est avant tout important d'observer différentes lexies nominales du français et leur comportement syntaxique respectif quant à leur valence active. Cette étape permet de dresser une première liste sommaire des DépSyntS régies par un gouverneur nominal. Un premier travail de repérage effectué par Lidija Iordanskaja (Iordanskaja, inédit) sur les données du DiCo a permis d'effectuer cette étape préliminaire. Cette ébauche a servi à établir une première liste des propriétés définitoires de ces DépSyntS nominales. Ce travail est analogue au travail de Marengo (2009) sur les dépendants syntaxiques des adjectifs. Dans son travail, Marengo part d'un groupe de lexies adjectivales témoins et regroupe les différentes relations faisant partie de leur régime sur la base de leurs propriétés particulières. Il semble cependant que ce travail n'ait pas un degré d'abstraction suffisant pour rendre compte de certains faits. Par exemple, une même relation peut servir à spécifier des constructions syntaxiques qui peuvent être ou non régies lexicalement (les dépendants peuvent exprimer des actants sémantiques de leur gouverneur ou être des modificateurs). Selon notre approche, on aura la même relation de surface dans les deux cas suivants :

*aide* —*r*→ *américaine*

*bœuf* —*r*→ *américain*

Dans le premier exemple, le dépendant correspond à un actant sémantique de son gouverneur, plus précisément le premier (grosso modo : *aide DE LA **personne** X À LA personne Y EN CE QUI CONCERNE LE problème Z AVEC LE moyen W*), ce qui n'est pas le cas dans le deuxième exemple. Bien que ces deux constructions puissent être associées à la même relation (soit la DépSyntS modificative), certaines de leurs propriétés spécifiques peuvent être légèrement différentes. Il en est de même lorsque la partie du discours du gouverneur diffère d'une construction à l'autre. Dans les deux cas suivants, la relation syntaxique *r* est la relation *agentive*.

[*L'homme est souvent*] *exploité* —*r*→ *par l'homme*

*exploitation* —[*de l'homme*]—*r*→ *par l'homme*

Ainsi, le premier travail effectué par Iordanskaja a certes servi de base, mais nous avons pris soin d'élargir notre perspective. Par la suite, le travail consistant à encoder les différents régimes du DiCo se situera dans la perspective plus pointue de l'analyse de données, ce qui nous permettra de vérifier la validité des critères d'identification qui auront été établis.

Le patron de description des DépSyntS que nous proposons se détaille donc en 5 étapes (qui diffèrent légèrement de celles explicitées dans Iordanskaja et Mel'čuk (2009, 169).

1. Identifier les Gouverneurs[G] possibles et les Dépendants[D] prototypiques de la relation
2. Identifier les propriétés définitoires de la relation
3. Identifier les propriétés descriptives de la relation
4. Justification (si nécessaire)

Dans les cas les moins évidents, il peut être utile de démontrer que la relation *r* ne pourrait pas être considérée comme étant une instance d'une relation différente (*r'*).

5. Commentaires (si nécessaire)

La première étape de notre travail de description est donc de cerner quels sont les G possibles et les D prototypiques des DépSyntS nominales, ainsi que leurs autres propriétés définitoires, ce qui les distingue fondamentalement les unes des autres. Nous ne nous sommes pas attardés dans un premier temps aux propriétés descriptives des relations. Pour qu'une propriété soit considérée comme définitoire, elle devait être assez générale pour être éventuellement applicable dans d'autres langues que le français. Certaines autres propriétés typiques du français ont été retenues. Par exemple, lorsqu'une relation a comme dépendant une lexie particulière du français, nous l'avons indiqué. Grosso modo, les propriétés définitoires sont celles qui sont nécessaires à la différenciation des relations et qu'une relation particulière possèdera toujours dans une construction syntaxique donnée. Voici donc le patron servant à l'identification des DépSyntS tel qu'il est pour l'instant. G signifie 'Gouverneur', D signifie 'Dépendant', PDD signifie 'partie du discours', RelSyntP signifie 'relation syntaxique profonde', RelSém signifie 'relation sémantique', ASém signifie 'Actant Sémantique' :

#### I. Taxinomie des G et D

- a. G = (<PDD>) [nature sémantique]
- b. D = <Lexème> (<PDD>)

#### II. Source de la dépendance

- a. <RelSyntP> <RelSém> [<Rôle sémantique de ASém>]
- b. Sémantème <'s'>
- c. Translation → Adj

#### III. Manifestation de la dépendance

- a. <Ordre linéaire>
- b. <Prosodie>
- c. <Dépendance morphologique>

Bien entendu, cette liste de propriétés définitoires devra être élargie au fur et à mesure que de nouvelles DépSyntS seront répertoriées. Celles qui y figurent sont celles qui ont été utiles jusqu'à présent. La « taxinomie des G et D » fait référence à la classification des gouverneurs et des dépendants selon différents critères. Pour l'instant, il semble pertinent d'identifier les gouverneurs et les dépendants d'une relation soit par leur partie du discours accompagnée d'autres précisions (ex. V transitif, N commun, etc.), leur nature sémantique (ex. prédicat, quasi-prédicat) ou par un lexème spécifique (ex. DE#1, QUE#1, etc.). La « source de la dépendance » est ce qui donne naissance à cette dépendance aux niveaux plus profonds (les niveaux SyntP ou Sém de la TST). Nous avons conclu que les RelSyntP, les RelSém et éventuellement certains rôles sémantiques, sémantèmes, et autres réalités syntaxiques telles que la translation (notion empruntée à Tesnière, 1959) peuvent être la source d'une DépSyntS particulière. Finalement, dans quelques cas, nous avons jugé que la façon générale dont la dépendance se manifeste dans la langue est une propriété définitoire de celle-ci. Par exemple, dans tous les cas où



Exemples

<i>écrire</i> [G] <i>à</i> [D] <i>Jean</i>	<i>lettre</i> [G] <i>à</i> [D] <i>Jean</i>
<i>donner</i> [G] <i>un cadeau</i> <i>à</i> [D] <i>Marc</i>	<i>fidèle</i> [G] <i>à</i> [D] <i>sa blonde</i>
<i>dessiner</i> [G] <i>un mouton</i> <i>à</i> [D] <i>Lucien</i>	<i>Ça lui</i> [D] <i>est</i> [G] <i>égal</i> (Contrôlé par le régime de ÉGAL)
<i>abandon</i> [G] <i>de ses biens</i> <i>à</i> [D] <i>ses enfants</i>	
<i>don</i> [G] <i>aux</i> [D] <i>pauvres</i>	

**Objectale oblique**Propriétés définitoires

- I. a. G = (prédicat ou quasi-prédicat)  
 b. D = (Prép ≠ Prép<sub>datif</sub> ou DE#1) ou QUE#2  
 II. b. ASéma ≠ 1

Exemples

<i>se rapprocher</i> [G] <i>de</i> [D] <i>Jean</i>	<i>rapprochement</i> [G] <i>avec</i> [D] <i>Jean</i>
<i>parler</i> [G] <i>de</i> [D] <i>politique</i>	<i>différent</i> [G] <i>de</i> [D] <i>Marcel</i>
<i>rêve</i> [G] <i>de</i> [D] <i>partir</i>	<i>le même</i> [G] <i>que</i> [D] <i>le tien</i>

**Subjectale adnominale**Propriétés définitoires

- I. a. G = (Ncomm, prédicat)  
 b. D = DE#1 ou « fam » À ou ENTRE  
 II. b. ASém 1 si DE ou À; ASém 1+2 si ENTRE

Exemples

<i>désir</i> [G] <i>de</i> [D] <i>Paul</i>	<i>lutte</i> [G] <i>entre</i> [D] <i>Paul et Pierre</i>
<i>faute</i> [G] <i>de</i> [D] <i>Marc</i>	<i>lutte</i> [G] <i>de</i> [D] <i>Paul contre l'ignorance</i>
« fam » <i>faute</i> [G] <i>à</i> [D] <i>Marc</i>	

**3.2 DépSyntS non régies (ou Non-valence-controlled SSyntDep)****Appositive numérative**Propriétés définitoires

- I. a. G = (Ncomm)  
 b. D = (Num)  
 II. b. sémantème 'numéro'

Exemples

<i>place</i> [G] <i>de stationnement</i> <i>18</i> [D]	<i>paragraphe</i> [G] <i>vingt</i> [D]
<i>abattoir</i> [G] <i>cinq</i> [D]	<i>version</i> [G] <i>3.2</i> [D]

**Circonstancielle**Propriétés définitoires

- I. a. G = (V) ou (N) ou (Adj)  
 b. D = (Adv) ou (Prép) ou (N)  
 II. b. Sémantème 'lieu' ou 'moment' ou 'durée' ou 'manière' ou 'cause' ou 'but'

III. c. Pas de dépendance morphologique

Exemples

*partir*[G] *ce matin*[D]  
*se sentir*[G] *bien en*[D] *Italie*  
*marcher*[G] *vite*[D]  
*maison*[G] *sur*[D] *la falaise*

*séjour*[G] *chez*[D] *Jean*  
*exil*[G] *en*[D] *Angleterre*  
*né*[G] *en*[D] *juin*  
*beau*[G] *de*[D] *façon bizarre*

**Modificative**

Propriétés définitives

I. a. G = (N sauf Pron)

b. D = (Adj, non déterm.)

III. c. Dépendance morphologique G→D

Exemples

*excellent*[D] *comportement*[G]  
*troisième*[D] *porte*[G] *à gauche*

*aide*[G] *américaine*[D]

**3.3 DépSyntS Mixtes (ou *Mixed SSyntDep*) (sont typiquement régies ou non-régies)**

**Appositive compositive**

Propriétés définitives

I. a. G = (Ncomm)

b. D = (Ncomm) ou (Npr)

III. c. Pas de dépendance morphologique G→D

Exemples

*un ticket*[G] *restaurant*[D]  
*des tickets*[G] *restaurant*[D]  
*espace*[G] *fumeur*[D]

*espace*[G] *fumeurs*[D]  
*ministère*[G] *Donadieu*[D]  
*(un) vert*[G] *émeraude*[D]

**Attributive adnominale**

Propriétés définitives

I. a. G = (Ncomm)

b. D = (Prép)

II. c. Translation→Adj<sup>1</sup>

Exemples

*gâteau*[G] *au*[D] *chocolat*  
*bouton*[G] *de*[D] *nacre*  
*congé*[G] *de*[D] *maladie*  
*viande*[G] *de*[D] *cheval*  
*moulin*[G] *à*[D] *café*  
*prison*[G] *de/pour*[D] *femmes*  
*problème*[G] *à*[D] *résoudre*

*fleur*[G] *des*[D] *champs*  
*femme*[G] *d'*[D] *une grande beauté*  
*tir*[G] *à*[D] *l'arc*  
*morceau*[G] *de*[D] *sucre*  
*mangeur*[G] *de*[D] *pommes de terre*  
[='personne qui mange régulièrement des  
pommes de terre']

**Complétive adnominale**Propriétés définitives

- I. a. G = (N, quasi-prédictat)  
b. D = DE#1 ou « fam » À

Exemples*jouet*[G] *de*[D] *Marc*« fam » *jouet*[G] *à*[D] *Marc**visage*[G] *de*[D] *Marc**morceau*[G] *de*[D] *ce sucre**voisine*[G] *de*[D] *Gustave**(Il est la) fierté*[G] *de*[D] *la famille*« fam » *bâtiment*[G] *à*[D] *Pierre***Déterminative**Propriétés définitives

- I. a. G = (N sauf Pron)  
b. D = (Dét)  
III. a. Position extrême dans le syntagme de G  
c. Dépendance morphologique G→D

Exemples*le/un/ce*[D] *dégoût*

]²

*son*[D] *dégoût*[G]

Tel que le mentionne notre patron de description des DépSyntS, après avoir cerné les propriétés définitives des relations, la deuxième étape pour une caractérisation complète des relations est d'en cerner les propriétés descriptives. À l'heure actuelle, les DépSyntS ne sont pas toutes associées de façon satisfaisante à l'ensemble de leurs propriétés descriptives respectives. Ce travail de caractérisation est toujours en cours. À titre indicatif, la description détaillée de la relation *attributive adnominale* est donnée dans Arbour Barbaud (en préparation). Toutefois, ces informations ne feront pas partie intégrante de l'ontologie GTN. De plus, pour la tâche qui consiste à ajouter le nom des DépSyntS dans les régimes lexicaux du DiCo, les propriétés définitives se sont avérées suffisantes. Les caractéristiques particulières des DépSyntS, seront utiles dans une grammaire du français ou dans un protocole plus vaste d'annotation de corpus français. Nous donnons donc en quelque sorte cette étape à titre indicatif seulement. Pour quelqu'un dont le but serait de véritablement caractériser les DépSyntS et d'en donner la liste exhaustive de propriétés, cette étape serait cruciale. En effet, il existe de nombreuses propriétés spécifiques aux dépendants de gouverneurs nominaux. On pourra se référer au travail effectué au LADL pour se rendre compte que de cerner la quasi-totalité de ces propriétés et de les associer aux dépendants correspondants dans une démarche lexicographique est une tâche ardue et complexe qui, bien qu'intéressante, est en dehors de la portée de la présentation actuelle.

**4 Présentation de la base de données CARNAVAL**

L'objectif général derrière notre travail est d'en arriver éventuellement à la création d'une base de données qui contienne l'ensemble des régimes de la langue française. Bien évidemment, il est plus réaliste à court terme de débiter la création de cette base par un nombre relativement restreint de régimes. Pour notre part, nous nous sommes fixés comme objectif de départ les régimes des lexies nominales du DiCo. Nous avons choisi cet échantillon, d'abord parce que le DiCo a surtout été développé avec ce type de lexies. Ceci est dû au fait que le DiCo est avant tout un dictionnaire de



combinatoire. Comme nous l'avons mentionné précédemment, chacune des lexies décrites dans le DiCo est accompagnée, entre autres, des collocations qu'elle contrôle : verbes supports, modificateurs, etc. Bien que les lexies de toutes les parties du discours puissent en principe contrôler ce type de relations syntagmatiques, c'est dans la description des noms prédicatifs et quasi-prédicatifs (grosso modo qui ont des actants) que cela s'avère le plus pertinent. D'autre part, comme les constructions syntaxiques gouvernées par des lexies nominales n'ont été que peu étudiées dans une perspective lexicographique, elles semblent donc offrir de nombreuses possibilités de réflexions intéressantes.

La base CARNAVAL est développée sous forme de fichier .xml grâce au logiciel oXygen. Avant de débiter les modélisations, nous avons d'abord analysé les régimes afin de bien identifier leurs différentes composantes. Nous avons ensuite hiérarchisé ces composantes en leur attribuant chacune un nom qui deviendrait éventuellement le nom d'une balise dans notre fichier .xml. Avant d'entamer la conception de nos fiches, un schéma .rnc (RELAX NG (*REgular LAnguage for XML, New Generation*), *Compact Syntax*) a été créé. Ce schéma permet la vérification automatique de la structure des fiches de la base de données. Le nom de chacune des balises se retrouve en majuscules dans nos descriptions, ce qui permet de les distinguer facilement du texte de nos fichiers. De plus, leur nom est en anglais, ce qui permet de différencier le métalangage de la langue sur laquelle porte les descriptions. Cela procure également l'avantage de permettre à un plus grand nombre de personnes de pouvoir comprendre la structure de nos fichiers, même si la langue décrite ne leur est pas accessible. Ce schéma détaille donc la modélisation des régimes lexicaux que nous avons conçue.

En somme, la base est une suite de GP et de MULTIPLE\_GP (GP est utilisé pour *government pattern*). Un gouvernement multiple est utilisé pour la description des lexies qui permettent les modifications de régimes (pour une explication de cette notion voir entre autres Mel'čuk *et al.*, 1984, 22). Chaque régime (GP ou MULTIPLE\_GP) a un numéro d'identification unique. Un régime simple GP est constitué d'un élément VALENCY et d'un élément DIATHESIS. L'élément VALENCY permet d'associer les variables sémantiques (SEM\_VAR) de la forme propositionnelle de la lexie (X, Y, Z, etc.) à la relation sémantique (SEM\_REL) impliquée (1, 2, 3, etc.). À l'intérieur de l'élément DIATHESIS l'expression de chacun d'eux est décrite dans une balise ACTANT\_EXPRESSION. L'élément DIATHESIS est donc ce que l'on pourrait considérer comme le régime en tant que tel. On y retrouve toutes les correspondances entre les différentes relations du régime, soit la SEM\_REL, DSYNT\_REL (relation syntaxique profonde) et la/les SSYNT\_DEP (dépendance syntaxique de surface). On y retrouve également l'ensemble des syntagmes qui servent à exprimer chacun des actants. Dans les cas où il est impossible d'exprimer directement un actant vis-à-vis de la lexie décrite, on retrouve la balise NO\_REALIZATION. Il est également possible d'ajouter des conditions à divers endroits du régime, ce qui permet notamment d'indiquer la présence obligatoire d'un élément.

La plupart des éléments qui composent le texte de nos fichiers sont également contraints par ce schéma. Les valeurs possibles ont été listées de façon à limiter l'introduction d'erreurs dans la base de données. En cours d'annotation, lorsque nécessaire, il est aisé d'introduire de nouvelles possibilités à ce schéma. Par exemple, cet extrait du schéma indique le texte qui peut être inséré à l'intérieur de la balise SSYNT\_DEP :

```
SSYNT_DEP-ELEM = element SSYNT_DEP {"agentive" | "appositive  
compositive" | "appositive numérative" | "appositive énonciative" | "attributive  
adnominale" | "circonstancielle" | "complétive adnominale" | "déterminative" |  
"modificative" | "objectale adnominale" | "objectale indirecte" | "objectale oblique"  
| "subjectale adnominale"}
```

Pour terminer la présentation, voici donc deux régimes de la base CARNAVAL. Le r.N.2.002 (pour : régime Nominal à 2 actants numéro 002) et le r.N.2.020 (pour : régime Nominal à 2 actants numéro 020) correspondent respectivement à plusieurs lexies dont les suivantes : CRIME#I.b, FAUTE#I.1a d'une part et ANIMOSITÉ, AVERSION d'autre part. Nous souhaitons préciser que cette première mise en page, qui est quelque peu simpliste, sera améliorée prochainement. Nous avons ajouté en fin d'article le texte des fichiers .xml du r.N.2.002 et r.N.2.020 incluant l'ensemble de leurs balises

respectives. Ceci permettra au lecteur (initié à ce langage) de voir les différentes composantes de ces régimes.

<b>r.N.2.002</b> 1 → X 2 → Y
1 ⇔ I
subjectale adnominale <input type="text" value="de N"/>
déterminative <input type="text" value="A-poss"/>
2 ⇔ II
—

3

<b>r.N.2.020</b> 1 → X 2 → Y
<b>Rég.1</b>
1+2 ⇔ I
subjectale adnominale <input type="text" value="entre N-coord"/>
déterminative <input type="text" value="A-poss"/>
<b>Rég.2</b>
1 ⇔ I
subjectale adnominale <input type="text" value="de N"/>
déterminative <input type="text" value="A-poss"/>
2 ⇔ II
objectale oblique <input type="text" value="Prép-envers N"/>

4

## 5 Conclusion

Grâce au travail de description que nous avons effectué, les noms des dépendances syntaxiques de surface ont pu être ajoutés aux tableaux de régimes des lexies nominales du DiCo. Après vérification des informations compilées, les régimes servent actuellement à la conception de la base de données CARNIVAL, que nous avons brièvement présentée précédemment. Lorsque la construction de la base sera complétée, un travail futur pourra établir un hyperlien dans chacune des fiches lexicographiques du DiCo et du LAF (Polguère & Mel'čuk, 2007) vers son régime particulier. Ceci permettra également d'avoir accès à l'information sur le régime via l'interface DiCoPop (version en ligne du LAF, non disponible pour le public actuellement), ce qui n'est pas le cas présentement. Éventuellement, ces informations serviront entre autres à comparer les unités lexicales qui partagent le même régime. Ceci permettrait de voir jusqu'à quel point il existe des corrélations entre les régimes lexicaux et les classes sémantiques (présentement établies grâce aux étiquettes sémantiques) auxquelles appartiennent les lexies. Ainsi, il serait possible de faire des généralisations sur les liens entre la sémantique lexicale et la

syntaxe régie qui se basent sur un large éventail de données. Ce type de généralisations, également connu sous le terme d'*argument linking*, a déjà soulevé de nombreuses réflexions (Gisborne, 2001 ; Jackendoff, 1990 et autres références dans Milićević, 2009). Nous croyons toutefois que les études sur le sujet se basent en général sur des intuitions ou sur un nombre restreint de données. Nous travaillons dans l'optique de fournir un vaste bassin de données qui pourra servir de départ pour de telles études.

Nous envisageons également que notre base de données pourra être utile pour la création d'articles de dictionnaire généralisés, où certains aspects (dont le régime) de lexies sémantiquement apparentées pourront être décrits simultanément. Nous croyons que cette perspective de généralisation est particulièrement intéressante au point de vue de la didactique de la langue (soit en tant que langue maternelle, seconde ou étrangère). Finalement, il est évident qu'une base de données où les régimes sont décrits de façon aussi méticuleuse pourrait être utile en TAL, que ce soit pour la génération automatique de textes ou de résumés de textes, pour la correction automatique, etc. Lorsque complet, l'inventaire décrivant l'ensemble des DépSyntS du français (qu'elles soient régies ou non) pourra également servir à de telles applications.

### Références bibliographiques

- Arbour Barbaud, É. (en préparation). *Identification et modélisation des dépendances syntaxiques régies du français*. Mémoire de maîtrise. Université de Montréal.
- Flaux, N. & Van de Velde, D. (2000). *Les noms en français: esquisse de classement*. Paris : Ophrys.
- Gaatone, D. (1988). Cette coquille de construction : remarques sur les trois structures affectives du français. *Travaux de linguistique*, 17, 159-176.
- Gisborne, N. (ed). (2001). Perspectives on argument structure and argument linking. *Special issue of Language Sciences*, 23, 4-5.
- Grevisse, M. & Goose, A. (2007). *Le Bon usage*. 14e édition. Bruxelles : De Boeck Université.
- Iordanskaja, L. & Mel'čuk, I. (2009). Establishing an Inventory of Surface-Syntactic Relations: Valence-controlled Surface-Syntactic Dependents of the Verb in French. Dans Mel'čuk, I. & Polguère, A. (eds.) *Dependency in Linguistic Description*. Amsterdam/Philadelphie : John Benjamins, 151-234.
- Jackendoff, R. (1990). *Semantic Structures*. Cambridge: MIT Press.
- Knittel, M. L. (2009) Le statut des compléments du nom en [de NP]. *Revue canadienne de linguistique*, 54 (2), 255-290.
- Larrivée, P. (1994). Quelques hypothèses sur les structures syntaxique et sémantique de *Ce fripon de valet*. *Revue québécoise de linguistique*, 23 (2), 101-113.
- Marengo, S. (2009). Les dépendants syntaxiques de l'adjectif en français: vers un inventaire des relations syntaxiques de surface. *Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'09)*, 263-273.
- Mel'čuk, I., Arbatchewsky-Jumarie, N., Elnitsky, L., Iordanskaja, L. & Lessard, A. (1984). *Dictionnaire explicatif et combinatoire du français contemporain : recherches lexico-sémantiques I*. Montréal : Les Presses de l'Université de Montréal.
- Mel'čuk, I., Clas A. & Polguère A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Louvain-la-Neuve : Duculot.
- Milićević, J. (2009). Schéma de régime --- le pont entre le lexique et la grammaire. *Langages*, 176, 94-116.
- Polguère, A. (2000). Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. *Proceedings of EURALEX'2000*, 517-527.
- Polguère, A. & Mel'čuk, I. (2007). *Lexique actif du français. L'apprentissage du vocabulaire fondé sur 20 000 dérivations sémantiques et collocations du français*. Bruxelles : De Boeck & Larcier.

- Riegel, M., Pellat, J.-C. & Rioul, R. (2009). *Grammaire méthodique du français*. Paris : Presses Universitaires de France.
- Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R. & Scheffczyk, J. (2006). *Framenet II: Extended Theory and Practice*. *ICSI*. [http://framenet.icsi.berkeley.edu/book/book.html]
- Sag, I. & Pollard, C. (1989). Subcategorization and Head-Driven Phrase Structure. Dans Baltin, M. & Kroch, A. (eds.) *Alternative Conceptions of Phrase Structure*. Chicago : University of Chicago Press, 139-181.
- Tellier, C. & Vallois, D. (2003). *Constructions méconnues du français*. Montréal : Les Presses de l'Université de Montréal.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. Paris : Klincksieck.
- Tremblay, O. (2010). *Une ontologie des savoirs lexicologiques pour l'élaboration d'un module de cours en didactique du lexique*. Thèse de doctorat. Université de Montréal.
- Van den Eynde, K. & Mertens, P. (2003). La valence: l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, 13, 63-104.

<sup>1</sup> Ici, « Translation→Adj » ne fait référence qu'à un type particulier de translation adjectivale dont Tesnière a traité. Il s'agit des cas où le D correspond soit à un « adjectif qualificatif » ou à un « adjectif de matière » selon les appellations de Tesnière.

<sup>2</sup> Nous ne prétendons pas à la complétude de cette liste de dépendances. Nous la donnons ici dans son état actuel. Nous n'écartons pas la possibilité d'ajouter des dépendances ou d'en modifier certains critères si notre travail d'annotation actuel ou si de futurs travaux sur les dépendances le démontrent nécessaire.

<sup>3</sup><GP id="r.N.2.002">

```

<VALENCY>
  <SEM_ACTANT_NUMBER>
    <SEM_REL>1</SEM_REL>
    <SEM_VAR>X</SEM_VAR>
  </SEM_ACTANT_NUMBER>
  <SEM_ACTANT_NUMBER>
    <SEM_REL>2</SEM_REL>
    <SEM_VAR>Y</SEM_VAR>
  </SEM_ACTANT_NUMBER>
</VALENCY>
<DIATHESIS>
  <ACTANT_EXPRESSION>
    <SEM_CORRESPONDANCE>
      <SEM_REL>1</SEM_REL>
      <DSYNT_REL>I</DSYNT_REL>
    </SEM_CORRESPONDANCE>
    <SSYNT_REALIZATION>
      <SSYNT_CORRESPONDANCE>
        <SSYNT_DEP>subjectale adnominale</SSYNT_DEP>
      <PHRASES>
        <PHRASE>
          <PHRASE_GOV_LEX>de</PHRASE_GOV_LEX>
          <ACTANT_SSYNT_EXPR>N</ACTANT_SSYNT_EXPR>
        </PHRASE>
      </PHRASES>
    </SSYNT_CORRESPONDANCE>
  </SSYNT_CORRESPONDANCE>
  <SSYNT_DEP>déterminative</SSYNT_DEP>
  <PHRASES>
    <PHRASE>
      <ACTANT_SSYNT_EXPR>A-poss</ACTANT_SSYNT_EXPR>
    </PHRASE>
  </PHRASES>

```

---

```
</PHRASES>
</SSYNT_CORRESPONDANCE>
</SSYNT_REALIZATION>
</ACTANT_EXPRESSION>
<ACTANT_EXPRESSION>
  <SEM_CORRESPONDANCE>
    <SEM_REL>2</SEM_REL>
    <DSYNT_REL>II</DSYNT_REL>
  </SEM_CORRESPONDANCE>
  <SSYNT_REALIZATION>
    <NO_REALIZATION>---</NO_REALIZATION>
  </SSYNT_REALIZATION>
</ACTANT_EXPRESSION>
</DIATHESIS>
</GP>
4<MULTIPLE_GP id="r.N.2.020">
  <VALENCY>
    <SEM_ACTANT_NUMBER>
      <SEM_REL>1</SEM_REL>
      <SEM_VAR>X</SEM_VAR>
    </SEM_ACTANT_NUMBER>
    <SEM_ACTANT_NUMBER>
      <SEM_REL>2</SEM_REL>
      <SEM_VAR>Y</SEM_VAR>
    </SEM_ACTANT_NUMBER>
  </VALENCY>
  <GP_OPTION name="Rég.1">
    <DIATHESIS>
      <ACTANT_EXPRESSION>
        <SEM_CORRESPONDANCE>
          <SEM_REL>1+2</SEM_REL>
          <DSYNT_REL>I</DSYNT_REL>
        </SEM_CORRESPONDANCE>
        <SSYNT_REALIZATION>
          <SSYNT_CORRESPONDANCE>
            <SSYNT_DEP>subjectale adnominale</SSYNT_DEP>
          <PHRASES>
            <PHRASE>
              <PHRASE_GOV_LEX>entre</PHRASE_GOV_LEX>
              <ACTANT_SSYNT_EXPR>N-coord</ACTANT_SSYNT_EXPR>
            </PHRASE>
          </PHRASES>
        </SSYNT_CORRESPONDANCE>
        <SSYNT_CORRESPONDANCE>
          <SSYNT_REL>déterminative</SSYNT_REL>
        <PHRASES>
          <PHRASE>
            <ACTANT_SSYNT_EXPR>A-poss</ACTANT_SSYNT_EXPR>
          </PHRASE>
        </PHRASES>
      </SSYNT_CORRESPONDANCE>
    </SSYNT_REALIZATION>
  </ACTANT_EXPRESSION>
</DIATHESIS>
</GP_OPTION>
<GP_OPTION name="Rég.2">
  <DIATHESIS>
```

---

```

<ACTANT_EXPRESSION>
<SEM_CORRESPONDANCE>
  <SEM_REL>1</SEM_REL>
  <DSYNT_REL>I</DSYNT_REL>
</SEM_CORRESPONDANCE>
<SSYNT_REALIZATION>
  <SSYNT_CORRESPONDANCE>
    <SSYNT_DEP>subjectale adnominale</SSYNT_DEP>
    <PHRASES>
      <PHRASE>
        <PHRASE_GOV_LEX>de</PHRASE_GOV_LEX>
        <ACTANT_SSYNT_EXPR>N</ACTANT_SSYNT_EXPR>
      </PHRASE>
    </PHRASES>
  </SSYNT_CORRESPONDANCE>
  <SSYNT_CORRESPONDANCE>
    <SSYNT_DEP>déterminative</SSYNT_DEP>
    <PHRASES>
      <PHRASE>
        <ACTANT_SSYNT_EXPR>A-poss</ACTANT_SSYNT_EXPR>
      </PHRASE>
    </PHRASES>
  </SSYNT_CORRESPONDANCE>
</SSYNT_REALIZATION>
</ACTANT_EXPRESSION>
<ACTANT_EXPRESSION>
<SEM_CORRESPONDANCE>
  <SEM_REL>2</SEM_REL>
  <DSYNT_REL>II</DSYNT_REL>
</SEM_CORRESPONDANCE>
<SSYNT_REALIZATION>
  <SSYNT_CORRESPONDANCE>
    <SSYNT_DEP>objectale oblique</SSYNT_DEP>
    <PHRASES>
      <PHRASE>
        <PHRASE_GOV_POS>Prép-envers</PHRASE_GOV_POS>
        <ACTANT_SSYNT_EXPR>N</ACTANT_SSYNT_EXPR>
      </PHRASE>
    </PHRASES>
  </SSYNT_CORRESPONDANCE>
</SSYNT_REALIZATION>
</ACTANT_EXPRESSION>
</DIATHESIS>
</GP_OPTION>
</MULTIPLE_GP>

```