

Restrictions de sélection et réalisations syntagmatiques dans DICOVALENCE Conversion vers un format utilisable en TAL

Piet Mertens

Département de Linguistique, Université de Leuven, Belgique
Piet.Mertens@arts.kuleuven.be

Résumé Cet article décrit des modifications du dictionnaire de valence des verbes du français DICOVALENCE qui visent à le rendre neutre par rapport aux modèles syntaxiques, à expliciter certaines informations sur le cadre de sous-catégorisation et à le rendre ainsi directement utilisable en TAL. Les informations explicitées sont les suivantes : (a) les fonctions syntaxiques des arguments verbaux, (b) les restrictions de sélection portant sur ces arguments et (c) leurs réalisations syntagmatiques possibles. Les restrictions sont exprimées à l'aide de traits sémantiques. L'article décrit aussi le calcul de ces traits sémantiques à partir des paradigmes des pronoms (et d'éléments similaires) associés aux arguments. On obtient un format indépendant du modèle syntaxique, dont l'interprétation est transparente.

Abstract This paper describes modifications to the verbal valency dictionary for French, known as DICOVALENCE, in order to obtain a theory-independent syntactic lexicon, to make explicit certain information about the slots in the valency frame, to facilitate the use of the lexicon in natural language processing. The modifications make explicit the following aspects: (a) the syntactic function of the slots, (b) the selection restrictions on these verbal arguments, (c) their possible phrasal realizations. Selection restrictions are expressed using semantic features. The article describes how these features are obtained from the paradigms of pronouns (and similar elements) associated with the valency slots. This results in a format which is theory-independent, with a transparent interpretation.

Mots-clés : lexiques syntaxiques, restrictions de sélection, traits sémantiques

Keywords: lexical databases, selection restrictions, semantic features

1 Introduction

Depuis sa mise à disposition en 2006, le lexique syntaxique DICOVALENCE (Eynde & Mertens 2003) a suscité beaucoup d'intérêt dans le domaine du TAL. Certaines études ont comparé les données et les approches syntaxiques dans plusieurs lexiques syntaxiques tels que le Lexique-Grammaire (les tables du LADL, Leclère 1990), DICOVALENCE, Synlex (Gardent *et al.* 2006) et le *Lefff* (Sagot *et al.* 2006). DICOVALENCE a été utilisé comme référence pour valider l'acquisition automatique des cadres de sous-catégorisation (Falk *et al.* 2007, Gardent 2009). Il a été utilisé pour valider et compléter les informations dans d'autres lexiques, tel le *Lefff* (Danlos & Sagot 2007a, 2007b; Sagot & Danlos 2007). Il a été intégré dans des analyseurs syntaxiques pour le français, mettant en œuvre des approches syntaxiques variées (Mertens 2008, Nasr & Béchet 2009).

Dans les travaux mentionnant DICOVALENCE, certaines critiques reviennent à plusieurs reprises (Falk *et al.* 2007, Kupść & Abeillé 2008, Messiant *et al.* 2008, Nasr & Béchet 2009). D'abord, DICOVALENCE ne renseignerait pas sur les réalisations syntagmatiques possibles des arguments verbaux. Ensuite, l'information dans DICOVALENCE ne serait pas directement utilisable en TAL. Enfin DICOVALENCE serait dépendant du modèle syntaxique de l'approche pronominale, ce qui limiterait son utilisation. La plupart de ces critiques sont évidemment liées. Comme le montrera cette étude, elles reposent le plus souvent sur une connaissance limitée de l'approche en question. (Falk *et al.* 2007) soulèvent les questions de la couverture du dictionnaire, de son rappel et sa précision inconnues. Rappelons que les méthodes d'acquisition automatique actuelles amalgament les différents emplois d'un même verbe, ce qui pose évidemment des problèmes de rappel et de précision nettement plus graves.

La présente étude vise plusieurs objectifs : (1) de préciser sous quelle forme les informations (fonctions syntaxiques, réalisations syntagmatiques...) sont répertoriées dans DICOVALENCE et comment on peut les expliciter, (2) d'exploiter les paradigmes de pronoms pour calculer certains traits sémantiques fonctionnant comme des restrictions sélectionnelles, (3) de fournir un nouveau format de DICOVALENCE qui rend explicites tous ces aspects, et qui soit directement utilisable en TAL. La mise en œuvre des lexiques syntaxiques les plus riches suppose en effet leur conversion préalable (Constant & Tolone 2008).

2 Rappel de quelques notions syntaxiques utilisées dans DICOVALENCE

2.1 Le cadre valenciel

L'ensemble des éléments régis par un verbe constitue sa *rection*. Celle-ci regroupe les *éléments valenciels* (le sujet et les compléments *essentiels*), les *circonstants* (de manière, de temps, de lieu, de raison, etc.), des éléments de négation (« ne, pas, point... ») et des éléments sans fonction sémantique, comme « en » dans « elle lui en veut énormément ». Les constituants syntaxiques correspondent le plus souvent à des arguments sémantiques, mais pas toujours : dans « il pleut » ou « il s'agit de décider », le sujet syntaxique « il » (dit *impersonnel*) n'a pas de fonction sémantique, ni de rôle thématique, il ne fonctionne pas comme agent ou référent.

Le *cadre valenciel* ou *cadre de sous-catégorisation* désigne l'ensemble des dépendants valenciels d'un verbe dans un emploi donné. Un même lemme peut présenter plusieurs emplois, chacun avec son propre cadre valenciel, comme dans les exemples suivants.

*Il ne manifeste jamais ses vrais sentiments.
Cette expression manifeste un dédain réel.*

RESTRICTIONS DE SELECTION

Aucun candidat ne s'est manifesté pour ce poste.

Les étudiants manifestent pour la réduction des gaz à effet de serre.

Le cadre valencielle comporte des places valencielles, appelées parfois *positions*, ou encore *paradigmes* dans l'approche pronominale et dans le DICOVALENCE. Dans cette dernière approche les cadres valenciels sont définis sur base de critères distributionnels, plutôt que sur une base sémantique. Dans DICOVALENCE les éventuels dépendants sans statut sémantique sont indiqués en dehors du cadre valencielle, par leur nom.

2.2 Les fonctions syntaxiques

Dans les lexiques syntaxiques, les places du cadre valencielle sont désignées de façons variées : le Lexique-Grammaire les désigne par leur nature syntagmatique et leur rang (N0, N1, Prép N2...), le *Lefff* et *SynLex* utilisent des fonctions syntaxiques. Cette dernière approche suppose un inventaire des fonctions retenues. Les lexiques syntaxiques reprennent généralement les fonctions syntaxiques classiques (sujet, objet direct, objet indirect, complément prépositionnel, attribut du sujet, attribut de l'objet...), avec quelques ajouts. Ainsi *Lefff* et *SynLex* distinguent deux objets indirects, selon la préposition utilisée dans le syntagme prépositionnel : *à-objet* et *de-objet*.

En revanche, l'approche pronominale (Blanche-Benveniste *et al.* 1984) définit les places valencielles de façon distributionnelle, à partir du paradigme de pronoms attesté dans la place en question. Le statut de place valencielle suppose en effet un *paradigme* comportant au moins deux pronoms, dont un interrogatif. Dans la plupart des cas, la correspondance entre paradigme et fonction syntaxique est transparente. Pour les exceptions, voir (Eynde & Mertens 2003) et la documentation du DICOVALENCE.

DV	Grammaire traditionnelle	Synlex	Lefff	Remarques
P0	sujet	SUJ	SUJ	sauf <i>il</i> impersonnel et discontinu
P1	objet direct	OBJ	OBJ	
P2	objet indirect	AOBJ	OBJA	aussi clitique sans prép.
P3	objet indirect	DEOBJ	OBJDE	aussi clitique sans prép.
PP	complément prépositionnel	POBJ	OBL	préposition fixe, pas de forme clitique
PQ				quantitatif
PMi	attribut du sujet	ATS	ATT	
PMi	attribut de l'objet	ATO	ATT	
PM	(circonstant de manière)	POBJ	OBL	pour le type valencielle uniquement
PL	(circonstant locatif)	POBJ	LOC	id.
PDL	(circonstant délocatif)	POBJ	DLOC	id.
PT	(circonstant de temps)	POBJ		id.
AE	-	-	-	adjectif évaluatif

Tableau 1 : Correspondance entre les termes utilisés dans plusieurs lexiques pour désigner les fonctions syntaxiques.

2.3. Eléments sans statut sémantique

Les exemples suivants comportent un pronom sans statut sémantique. Celui-ci ne peut pas être remplacé par un autre pronom ou par un groupe nominal ; il n'admet pas les transformations applicables aux éléments valenciels (clivage, dislocation). Par conséquent, ce pronom ne sera pas considéré comme la réalisation d'un paradigme, mais plutôt comme un comme un complément apparent (ou *pseudo-paradigme*).

dehors il pleuvait

(pseudo-il, verbes météorologiques)

PIET MERTENS

<i>ça caille ici</i>	(<i>pseudo-ça</i>)
<i>pour s'élancer, il faut plus de place</i>	(<i>pseudo-se, verbes pronominaux</i>)
<i>je le prends sur moi ; j'emporte avec moi ce souvenir</i>	(<i>pseudo-soi</i>)
<i>il en arrive à déménager</i>	(<i>pseudo-en</i>)
<i>il s'y connaît en mathématiques</i>	(<i>pseudo-y</i>)
<i>elle l'emporte sur son frère</i>	(<i>pseudo-le</i>)
<i>il vous la coupe</i>	(<i>pseudo-la</i>)
<i>il planta là son travail</i>	(<i>pseudo-là</i>)

Le traitement est analogue pour certaines notions de la grammaire traditionnelle. Pour les *verbes pronominaux*, par exemple, on observe bien plusieurs pronoms alternatifs (« je m'évanouis, tu t'évanouis... »). Cependant le pronom réfléchi est prévisible à partir du sujet (« *je t'évanouis », etc.) ou à partir de la forme du verbe (« élance-toi », « pour s'élancer »). Le DICOVALENCE le désigne comme le *pseudo_se*.

Plusieurs types de *constructions impersonnelles* rentrent dans la catégorie mentionnée. Le sujet des verbes météorologiques (« il pleut, il neige... ») est représenté dans DICOVALENCE comme un *pseudo_il*. Les constructions du type « il faut qu'elle parte », « il s'agit de travailler » sont représentées comme des verbes à *sujet discontinu*, où le sujet est constitué de deux parties, d'une part le pronom « il », d'autre part la complétive (« qu'elle parte ») ou l'infinitive (« de travailler »). Cette analyse repose sur les observations suivantes : dans ces cas, « il » ne peut pas être remplacé par un autre pronom (« *elle s'agit de travailler ») ou par un pronom interrogatif (« *qu'est-ce qu'il s'agit de travailler »).

3 Réalisations syntagmatiques des arguments valenciels

Pour noter les réalisations syntagmatiques possibles de chaque place valencielle les lexiques syntaxiques utilisés en TAL distinguent généralement les catégories syntaxiques majeures suivantes : pronoms clitiques, syntagme nominal, syntagme prépositionnel, syntagme adjectival, proposition infinitive (introduite ou non par une préposition), complétive (introduite par la conjonction « que », avec mention du mode utilisé, indicatif ou subjonctif).

Le nombre de catégories retenues varie d'un lexique à l'autre. Dans SynLex elles sont relativement peu nombreuses : sn (syntagme nominal), pinf (infinitive), pcompl (complétive), qcompl (question indirecte). SynLex assimile donc les syntagmes nominal et prépositionnel, et fait abstraction de la présence de la préposition, information prévisible à partir de la fonction syntaxique. Le *Lefff* propose un inventaire de catégories syntaxiques très riche, qui distingue plusieurs types de pronoms (clitiques nominatifs, accusatifs, datifs, « en », « y », se réfléchi, se réciproque...), plusieurs types de syntagmes nominaux (sn, de-sn, à-sn, ...) et adjectivaux (sa, pour-sa, ...), qui reprennent la préposition éventuelle liée à la fonction syntaxique. Il en est de même des infinitives (inf, de-inf, ...) et de complétives (scompl = que-phrase, à-scompl = à ce que P, de-scompl = de ce que P, qcompl = question indirecte,...). Vu le caractère prévisible de la préposition, cet inventaire détaillé semble en partie redondant.

Quelles sont les réalisations syntagmatiques pouvant être dérivées des informations dans DICOVALENCE ? Pour chaque place valencielle, DICOVALENCE fournit une liste d'éléments pronominaux, à la fois clitiques et non clitiques (« ça, celui-ci, celle-ci, à lui, de lui, à eux, à ça, de ça, qui, quoi... »). Notons que la classe est plus large que les pronoms de la grammaire traditionnelle et inclut également des pronoms *indéfinis* (« autant, combien... ») et des adverbes (« comment, ainsi, là, ici, ailleurs, à ce moment-là »).

Selon l'idée centrale de l'approche pronominale, l'existence d'un paradigme de pronoms implique toujours la possibilité de réalisations comportant des mots lexicaux, sous la forme d'un groupe nominal, d'un groupe prépositionnel ou autre, selon la fonction syntaxique :

RESTRICTIONS DE SELECTION

{ *qui, je, tu, il, elle...* } → *cet homme, cet objet, la table, Pierre..*

{ *à qui, me, te, lui, leur, nous...* } → *à cet homme, à eux, à Jean..*

{ *où, là, ici...* } → *ailleurs, nullepart, dans cette salle, à Montréal, à TALN...*

Qui plus est, certains éléments de paradigme indiquent la *réalisation non pronominale* correspondant à un pronom. Il s'agit tantôt d'une proposition infinitive, éventuellement introduite par une préposition, tantôt d'une subordonnée introduite par « que » (*que-phrase*) ou par « si », ou d'une question indirecte. Dans le cas des subordonnées, le mode est également explicite. Pour des limitations de place, on se limitera ici à quelques exemples pour la fonction objet.

<i>PI : le(inf), ça(inf)</i>	<i>il admet l'avoir trompée</i>
<i>PI : le(de_inf)</i>	<i>je lui ai crié de revenir à temps</i>
<i>PI : le(à_inf)</i>	<i>il m'a demandé à voir ce que j'écrivais</i>
<i>PI : le(qpind), ça(qpind)</i>	<i>Luc a remarqué qu'il manquait plusieurs documents</i>
<i>PI : le(qpsubj), ça(qpsubj)</i>	<i>elle exige qu'on lui rende son argent</i>
<i>PI : le(spind), ça(spind)</i>	<i>on me demande si elle est rentrée</i>
<i>PI : le(indirq), ça(indirq)</i>	<i>elle demande qui était là, je sais comment faire, il sait de quoi parler</i>

Grâce à ces éléments, le passage des paradigmes de pronoms aux réalisations syntagmatiques possible s'avère trivial.

4 Restrictions sélectionnelles

4.1 Les traits sémantiques, leur définition et leur organisation

Le terme *référent* sera utilisé ici pour désigner ce à quoi renvoie un pronom ou un nom. Dans « je la connais », le pronom « la » peut renvoyer à une personne, à un animal, à une plante, à un objet, à un événement, à une notion abstraite (« je la connais, cette religion »), et ainsi de suite. Mais « je » ne désigne qu'un être humain. Pour distinguer les types de référents, on se sert souvent de *traits sémantiques*, tels que [humain], [animé], [concret] (Flaux & Van de Velde 2000). Ces mêmes traits sont couramment utilisés pour représenter des *restrictions de sélection* portant sur les arguments d'un prédicat verbal dans un emploi donné (Leclère 1990). Provisoirement on utilisera la notation suivante, qui sera commentée plus loin.

<i>homme (nom)</i>	<i>[+concret, +animé, +humain]</i>
<i>chat (nom)</i>	<i>[+concret, +animé, -humain]</i>
<i>religion (nom)</i>	<i>[-concret]</i>
<i>la (pronom)</i>	<i>[±concret, ±animé, ±humain]</i>
<i>je (pronom)</i>	<i>[+concret, +animé, +humain]</i>
<i>ceci (pronom)</i>	<i>[+concret, -animé, -humain]</i>
<i>celui-ci (pronom)</i>	<i>[±concret, ±animé, ±humain]</i>
<i>ça (pronom)</i>	<i>[±concret, ±animé, ±humain]</i>

Il existe plusieurs façons d'utiliser et de combiner les traits sémantiques. Certains lexiques syntaxiques, comme *VerbNet* pour l'anglais, adoptent une *organisation hiérarchique* des traits, comme dans une taxinomie, avec des classes emboîtées les unes dans les autres. Chaque classe se définit par un ou plusieurs traits. À chaque fois qu'on applique à une classe un trait supplémentaire, elle se subdivise en plusieurs sous-classes, pour lesquelles le nouveau trait prend des valeurs alternatives. Dans cette approche, une sous-classe hérite des traits de la super-classe, mais présente un (ou plusieurs) traits additionnels qui permettent

de la différencier des sous-classes complémentaires. Ainsi la classe des référents (ou des éléments lexicaux correspondants) dotés du trait [+concret] comporte plusieurs sous-classes, parmi lesquelles celle avec le trait [+animé], qui à son tour comporte la sous-classe [+humain], et ainsi de suite. Les traits seront utilisés pour contraindre les réalisations des places valencielles. Dans cette approche hiérarchique, lorsqu'un élément valenciel présente le trait [+humain], par définition il présente également les traits [+animé] et [+concret], bien que seulement une partie des référents concrets soient acceptés.

Karel van den Eynde (communication personnelle) propose également un classement hiérarchique, mais organisé autrement. Au premier niveau, le trait [humain] oppose les référents humains aux non-humains, et à l'intérieur de la deuxième classe le trait [concret] oppose les référents abstraits aux référents concrets non humains.

G. Gross (1994) montre les difficultés liées à l'utilisation des traits sémantiques en général, et implicitement à leur organisation hiérarchique. Mentionnons quelques objections majeures. D'abord les traits sémantiques ne sont pas assez précis pour représenter les propriétés pertinentes sur le plan syntaxique. Ensuite, ils sont basés sur des distinctions ontologiques, utilisées de façon intuitive et définies a priori. Certains éléments lexicaux sont à cheval sur plusieurs critères de classification ; ainsi la « cote de maille » est à la fois un *objet métallique* et un *vêtement*. Les définitions des traits seraient insatisfaisantes à la fois pour la description linguistique, et pour leur application en TAL. Selon l'auteur les classes devraient être établies « de façon interne, à l'aide des moyens que la langue fournit elle-même, c'est-à-dire des outils syntaxiques ». Des critères distributionnels permettent en effet de définir des *classes d'objets*, telles que « moyen de transport, profession, fonction, collectif, maladie, locatif, lieu de travail, qualité » (Gross 1994:17). Cependant, à un niveau préalable aux classes d'objets, il maintient 8 classes majeures, délimitées par les traits suivants : *humain, animal, végétal, inanimé concret, inanimé abstrait, locatif, temps, événement*. Pour la classe majeure des humains il y aurait 54 classes d'objets, par exemple. Dans l'approche de Gross ces 8 traits ne présentent pas d'organisation hiérarchique, mais constituent un classement à un niveau, comportant 8 classes indépendantes, définies à partir de critères syntaxiques.

L'organisation des traits a un impact sur leur représentation et sur leur utilisation. En effet, dans un système hiérarchique où $[+humain] \subset [+animé] \subset [+concret] \subset [\pmconcret]$, les caractérisations formulées plus haut doivent être lues comme dans la colonne A du tableau ci-dessous. La notation sous B, plus concise et plus lisible, évite la redondance.

	A	B
homme (nom)	$[+concret] \cap [+animé] \cap [+humain]$	[+humain]
la (pronom)	$[\pmconcret] \cup [\pmanimé] \cup [\pmhumain]$	[]
je (pronom)	$[+concret] \cap [+animé] \cap [+humain]$	[+humain]
ceci (pronom)	$[+concret] \cap [-animé] \cap [-humain]$	[-humain]

Les difficultés de notation se manifestent par exemple quand un terme de valence accepte à la fois les référents [+humain] et les référents [-concret], comme pour l'objet indirect de « s'accoutumer, se consacrer, céder » ou l'objet direct de « comprendre, dénoncer ». La notation « $[\pmconcret, +humain]$ », assez imprécise, devrait être interprétée comme « $([+concret] \cap [+humain]) \cup [-concret]$ ». Cette difficulté sera évitée avec le classement non hiérarchique, comme pour les 8 classes majeures de G. Gross.

4.2 Le calcul de traits dans DICOVALENCE

Notre objectif est d'obtenir automatiquement certains traits sémantiques à partir d'informations déjà présentes dans DICOVALENCE, plus particulièrement à partir des paradigmes de pronoms associés aux places de valence (Eynde *et al.* 1988). En effet, la présence de certains éléments dans le paradigme

RESTRICTIONS DE SELECTION

permettra d'associer à celui-ci un ou plusieurs traits. Lorsqu'un pronom peut dénoter des référents présentant certaines propriétés, il est compatible avec ces propriétés et avec les traits sémantiques qui les explicitent. Ainsi les pronoms « je, nous, qui » désignent des êtres humains, mais pas des objets. Ils portent donc le trait [+humain]. Le pronom « ceci », en revanche, ne désigne que des référents concrets non humains. Ces informations seront mises à profit dans le calcul des traits associés aux places valencielles.

Dans cette approche, seuls peuvent être calculés les traits pour lesquels il existe des pronoms (ou des éléments, cf. infra) qui permettent de les identifier. Il n'existe pas de pronom réservé aux animaux ou aux végétaux ; aussi ces classes seront-elles regroupées avec les concrets non humains. On obtient ainsi 3 classes majeures, sans hiérarchisation : [humain], [concret non humain], [abstrait]. On obtient également le trait indiquant la propriété [complexe], qui rejoint la notion de *collectif* (cf. infra), et qui se combine avec les 3 classes majeures.

La comparaison des classes proposées par G. Gross avec celles calculables à partir des pronoms permet plusieurs observations : dans DICOVALENCE, (1) les classes *locatif* et *temps* de Gross sont réservées aux fonctions syntaxiques correspondantes (à savoir PL et PT) ; (2) la classe *événement* est regroupée avec les référents abstraits ; (3) enfin, les classes *végétal* et *animal* sont regroupées avec les concrets non humains.

La qualité de l'attribution des traits dépend du trait envisagé. Le trait [humain] peut être attribué pour toutes les places valencielles de la quasi-totalité des entrées du lexique verbal. Pour [abstrait], l'attribution n'est possible que dans une partie des cas.

Les pronoms les plus importants pour l'attribution des traits sont énumérés ci-dessous, avec leurs interprétations possibles. (« qui » désigne le pronom interrogatif.) L'élément *EVENT* est une macro qui accepte les différentes réalisations sous forme de subordonnée mentionnées plus haut : « ça(inf), ça(de_inf), ça(à_inf), ça(qpind), ça(qpsubj)... »

<i>qui, je, nous, vous, te...</i>	→ [humain]
<i>ceci</i>	→ [concret non humain]
<i>ça</i>	→ [abstrait] ∨ [concret non humain] ∨ [humain] (<i>désindividualisant</i>)
<i>ça(inf), EVENT</i>	→ [abstrait]

4.2.1 Le trait [humain]

Si le paradigme sujet d'une entrée lexicale comporte le pronom « je », « nous » ou « qui », alors cette place admet des référents humains et elle reçoit le trait [humain]. On le notera comme suit :

sujet : (*je* ∨ *nous* ∨ *qui*) → [humain]

La disjonction (« je » ∨ « nous ») permet de détecter les sujets qui ne s'emploient qu'au pluriel (« s'affronter, s'entraider, s'allier contre... » ; cf., pour l'objet, « accumuler, énumérer, confondre... »).

Pour les sujets discontinus, on utilise une règle similaire portant sur l'élément discontinu « il_de_qui ».

sujet : (*il_de_qui*) → [humain]

Pour les autres fonctions syntaxiques, les règles sont identiques, sauf que les pronoms utilisés seront évidemment ceux de la fonction syntaxique en question :

objet : (<i>te</i> ∨ <i>nous</i> ∨ <i>qui</i>)	→ [humain]
objet_à : (<i>qui</i> ∨ <i>me</i>)	→ [humain]

objet_de : (*qui*) → [humain]
 objet_prep : (*qui*) → [humain]

4.2.2 *Le trait [concret non humain]*

La présence de « ceci » en position sujet déclenche l'attribution du trait [concret non humain]. Une deuxième règle (sans « ceci ») permet dans certains cas d'attribuer ce trait aux référents sujets concrets incompatibles avec « ceci », comme dans « le bateau accoste et rembarque des passagers », « l'éléphant barrit ». Bien que le pronom « il » soit ambigu (il permet en effet les interprétations [humain], [non humain] ou [abstrait]), l'absence de « je, nous, qui » dans les cas visés exclut l'interprétation de « il » comme [humain], et l'absence de « ça, EVENT » celle de « il » comme [abstrait]. Des règles analogues sont données pour l'objet direct (où « que » désigne aussi « qu'est-ce que ») et pour l'objet indirect.

sujet : (*ceci* ∨ *il_de_ceci*) → [concret non humain]
 sujet : (*il* ∨ *ils*) ∧ ¬ (*je* ∨ *nous* ∨ *qui* ∨ *ça* ∨ *EVENT*) → [concret non humain]
 objet, objet indirect : (*ceci*) → [concret non humain]
 objet : (*que*) ∧ ¬ (*te* ∨ *vous* ∨ *qui* ∨ *ça* ∨ *EVENT*) → [concret non humain]
 objet indirect : (*ça* ∨ *y*) ∧ ¬ (*qui* ∨ *EVENT*) → [concret non humain]

La règle sans « ceci » ne sera appliquée que si la règle avec « ceci » ne s'applique pas au paradigme analysé. Cette situation demande l'application ordonnée des règles, qui sera interrompue après la première règle remplissant les conditions.

4.2.3 *Le trait [abstrait]*

Pour le calcul du trait [abstrait] on exploite les éléments *EVENT*, réservé aux abstraits, et « ça ». Le pronom « ça » étant ambigu, on vérifie l'absence d'éléments qui montrent l'interprétation comme [humain] (à savoir « je, nous, qui ») et comme [concret non humain] (à savoir « ceci »). Pour les autres fonctions syntaxiques, les règles sont analogues. On ajoute une troisième règle qui indique que l'interprétation [abstrait] est probable étant donné la présence de « ça », sans être certaine, vu la présence de « ceci ».

sujet, objet : (*EVENT*) → [abstrait]
 sujet : (*ça*) ∧ ¬ (*je* ∨ *nous* ∨ *qui* ∨ *ceci*) → [abstrait]
 objet : (*ça*) ∧ ¬ (*te* ∨ *vous* ∨ *qui* ∨ *ceci*) → [abstrait]
 objet : (*ça*) ∧ (*ceci*) → [?abstrait]

4.2.4 *Le trait [complexe]*

Ce trait sera attribué aux termes dont la réalisation se limite à un référent au pluriel, à la coordination de plusieurs référents ou à un *collectif* : « ces droites convergent », « *cette droite converge » (mais : « elle converge avec celle-là » et « elle converge vers l'équation d'une droite »), « les preuves contre lui abondent », « *la preuve contre lui abonde », « on va accoler les mots », « les cheveux collent ensemble ».

sujet : (*ils* ∨ *elles*) ∧ ¬ (*il* ∨ *elle*) → [complexe]
 objet : (*nous* ∨ *les*) ∧ ¬ (*le* ∨ *la*) → [complexe]
 objet-à : (*leur*) ∧ ¬ (*lui* ∨ *lui_ton* ∨ *ceci*) → [complexe]

RESTRICTIONS DE SELECTION

objet-de : $(eux) \wedge \neg (lui \vee lui_ton \vee ceci) \rightarrow$ [complexe]

objet-prep : $(eux) \wedge \neg (lui \vee lui_ton \vee ceci) \rightarrow$ [complexe]

5 Représentation des informations dans le lexique syntaxique

L'extrait suivant illustre le résultat obtenu. La cadre valencielle apparaît entre chevrons. Pour chaque place de valence sont mentionnés la fonction syntaxique, la liste des réalisations syntagmatiques possibles (séparées par '|'), les traits sémantiques (hum, nhum (= concret non humain), abs (= abstrait), complexe), y compris le mode de la subordonnée (indicatif ou subjonctif), si présente. Le point d'interrogation précédant la place indique son caractère facultatif. Celui qui précède un trait indique sa probabilité relative.

```
54980, manifester, pred, < subj:pron|n:[hum], obj:pron|n:[abs], ?objà:pron|n:[hum] >
  il ne manifeste jamais ses vrais sentiments
54990, manifester, pred, < subj:pron|n:[abs], obj:pron|n|compl|indq:[abs,mood:ind] >
  cette expression manifeste un dédain réel
55000, se manifester, pred, < pseudo_se, subj:pron|n:[hum] >
  aucun assistant ne s'est manifesté pour ce poste
55010, manifester, pred, < subj:pron|n:[hum], ?objp<contre,pour>:pron|n: [hum,?nhum,abs] >
  on manifeste pour une terre plus verte et plus saine
55020, manigancer, pred, < subj:pron|n:[hum], obj:pron|n:[abs] >
  il parvient toujours à manigancer quelque chose
55030, manipuler, pred, < subj:pron|n:[hum], obj:pron|n:[nhum,?abs] >
  il aime manipuler les gens
55040, manoeuvrer, pred, < subj:pron|n:[hum], ?obj:pron|n:[nhum,?abs] >
  les officiers manoeuvrent le bateau
55050, manquer, pred, < subj:pron|n|compl|de_inf:[hum,nhum,abs,mood:subj],
?objà:pron|n:[hum] >
  ma femme me manque beaucoup
55060, manquer, pred, < subj:pron|n:[hum], obj:pron|n:[hum], ?quant<de>:pron|n:[] >
  il a encore manqué le car de 10 heures
55070, manquer, pred, < subj:pron|n:[hum,nhum,?abs], ?loc<>:pron|n:[] >
  ces produits manquent dans ce magasin
55080, manquer, pred, < subj:pron|n:[abs] >
  cette épreuve a manqué totalement
55100, manquer, pred, < subj:pron|n:[hum], ?objà:pron|n:[hum,abs] >
  un honnête homme ne peut manquer à ses obligations
55117, manquer de, adjunct:5:plain, < >
  ça ne manque jamais de me tomber dessus
```

6 Conclusion

Dans cet article nous avons décrit un ensemble de modifications du DICOVALENCE qui faciliteront son utilisation en TAL, mais aussi, d'une façon plus générale, dans la recherche en syntaxe. Ces modifications concernent trois propriétés des places valencielles : leur fonction syntaxique, leurs restrictions sélectionnelles et leurs réalisations syntagmatiques possibles. Des limitations de place empêchent qu'on détaille ici les autres modifications envisagées, pour la plupart déjà réalisées.

Une partie de l'article décrit comment les traits sémantiques, qui expriment les restrictions de sélection portant sur les arguments valenciels, sont calculés à partir des paradigmes des pronoms et d'autres éléments similaires, associés à ces arguments. Pour la plupart des utilisateurs actuels et futurs, ces ajouts apparaîtront comme une véritable extension du dictionnaire.

Ces modifications ont résulté dans une nouvelle version du lexique, quasi-indépendante du modèle syntaxique initial, où les informations apparaissent dans une forme transparente, et pour cette raison directement utilisable en TAL. Cependant, ce nouveau format résulte d'une conversion automatique de la version originale du DICOVALENCE, qui reste donc la source unique et suffisante.

Références

- BLANCHE-BENVENISTE C., DELOFEU J., STEFANINI, J., EYNDE, K. VAN DEN (1984). *Pronom et syntaxe. L'approche pronominale et son application au français*. Paris: SELAF, pp. 245.
- CONSTANT M., TOLONE E. (2010). A generic tool to generate a lexicon for NLP from Lexicon-Grammar tables. *Actes 27e Colloque international sur le lexique et la grammaire* (L'Aquila, 10-13 septembre 2008).
- DANLOS L., SAGOT B. (2007). Comparaison du Lexique-Grammaire des verbes pleins et de Dicovalence : vers une intégration dans le Leff. *Actes TALN 2007*, 229-238 (Toulouse, 5-8 juin 2007).
- DANLOS L., SAGOT B. (2007). Constructions pronominales dans Dicovalence et le Lexique-Grammaire : intégration dans le Leff. *Actes Colloque Lexique et Grammaire 2008*, L'Aquila.
- EYNDE K. VAN DEN, BROEDERS E., EGGERMONT C., VANGILBERGEN, L. (1988). The pronominal approach in NLP: a pronominal feature analysis of coordination in French. *Computers and Translation* 3, 177-213.
- EYNDE K. VAN DEN, MERTENS P. (2003). La valence : l'approche pronominale et son application au lexique verbal. *French Language Studies* 13, 63-104.
- FALK I., FRANCOPOULO G., GARDENT CL. (2007). Evaluer Synlex. *Actes TALN 2007*, 335-344.
- FLAUX N. & VAN DE VELDE D. (2000). *Les noms en français : esquisse de classement*. L'essentiel français. Gap : Ophrys.
- GARDENT C. (2009). Technique d'acquisition de connaissances lexicales à partir de corpus analysés en syntaxe. Manuscrit : Délivrable du projet Passage Passage/2009/D10/v1.0
- GARDENT C., GUILLAUME B., PERRIER G., FALK I. (2006). Extraction d'information de sous-catégorisation à partir des tables du LADL. *Actes TALN 2006*.
- GROSS G. (1994). Classes d'objets et description des verbes. *Langages* 115, 15-30.
- KUPSC A., ABELLE A. (2008). Growing TreeLex. in: GELBUKH A. (ed.), 9th Int. Conf., CICLing 2008, (Haifa, Israel, February 2008). *Lecture Notes in Computational Linguistics no 4919*, pp. 28-39.
- LECLERE C. (1990). Organisation du Lexique-Grammaire des verbes français. *Langue française* 87, 112-122.
- MESSIANT C., KORHONEN A., POIBEAU, TH. (2008) LexSchem: A Large Subcategorization Lexicon for French Verbs. *Actes LREC* (Marrakech).
- MERTENS P. (2008). Factorisation des contraintes syntaxiques dans un analyseur de dépendance. *Actes TALN 2008* (Avignon, 9-13 juin 2008).
- NASR A., BECHET, F. (2009). Analyse syntaxique en dépendances de l'oral spontané. *Actes TALN 2009*.
- SAGOT B., CLÉMENT L., DE LA CLERGERIE É., BOULLIER P. (2006). The Leff 2 syntactic lexicon for French: architecture, acquisition, use. *Actes de LREC 06* (Gênes, Italie).
- SAGOT B., DANLOS L. (2007). Améliorer un lexique syntaxique à l'aide des tables du lexique-grammaire: Constructions impersonnelles. *Cahiers du Cental*.