



HAL
open science

Constructions pronominales dans Dicovalence et le lexique-grammaire – Intégration dans le Leff

Laurence Danlos, Benoît Sagot

► **To cite this version:**

Laurence Danlos, Benoît Sagot. Constructions pronominales dans Dicovalence et le lexique-grammaire – Intégration dans le Leff. *Linguisticae investigationes: International Journal of Linguistics and Language*, 2009, 32 (2), pp.293-304. 10.1075/li.32.2.11dan . inria-00515459

HAL Id: inria-00515459

<https://inria.hal.science/inria-00515459v1>

Submitted on 7 Sep 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Constructions pronominales dans Dicovalence et le lexique-grammaire — Intégration dans le *Lefff*

Laurence Danlos, Benoît Sagot
INRIA & Université Paris 7

Introduction

Dans le but de construire un lexique morpho-syntaxique du français à large couverture et utilisable dans les applications de Traitement Automatique des Langues (TAL), et dans le prolongement de travaux déjà effectués dans cette direction, en particulier sur les constructions impersonnelles (Danlos & Sagot, 2007; Sagot & Danlos, 2007), nous nous intéressons ici aux constructions pronominales. Notre objectif est triple :

1. réaliser une étude comparative du traitement des constructions pronominales dans le lexique-grammaire (Gross, 1975; Leclère, 2002) et dans DICOVALENCE (Van den Eynde & Mertens, 2006)¹,
2. proposer une formalisation des résultats de ces travaux qui soit adaptée au TAL, dans le cadre du formalisme utilisé par le *Lefff* 3, dernière version du *Lefff* décrite brièvement ci-dessous,
3. fusionner dans le *Lefff* les informations lexicales présentes du lexique-grammaire et de DICOVALENCE concernant les constructions pronominales.

Le *Lefff* (Lexique des Formes Fléchies du Français) est conçu pour pouvoir être directement utilisé dans plusieurs systèmes de TAL (Sagot *et al.*, 2006). Divers travaux, dont celui présenté ici, ont pour but d'améliorer sa qualité et sa couverture. La dernière version du *Lefff*, la version 3, fait usage de la notion de redistribution : le lexique *intensionnel*, édité par les développeurs du lexique, associe à chaque entrée un cadre de sous-catégorisation canonique, et liste les redistributions possibles à partir de ce cadre. Le processus de *compilation* du *Lefff* intensionnel en *Lefff* extensionnel construit différentes entrées pour ces différentes redistributions.

Soit par exemple l'entrée intensionnelle simplifiée suivante :

```
clarifier1  Lemma;v;<arg0:Suj:c|n|scompl|sinf|sn,  
           arg1:Obj:(cla|scompl|sn)>;;%actif,%se_moyen_impersonnel,  
           %passif_impersonnel,%passif
```

1. DICOVALENCE a été développé dans le cadre de l'approche pronominale, qui a été initiée par Claire Blanche-Benvéniste et Karel van den Eynde (van den Eynde et Blanche-Benvéniste 1978).

Elle décrit une entrée du lemme verbal *clarifier*, qui est transitif direct (deux arguments réalisés canoniquement par les fonctions syntaxiques Suj et Obj décrites entre les chevrons), et qui admet les redistributions fonctionnelles (préfixées par le symbole %) *actif* (la distribution par défaut), *se moyen impersonnel* (*il s'est clarifié de nombreuses choses à cette réunion*), *passif impersonnel* (*il a été clarifié par Pierre que Luc était le coupable*), et *passif*. À titre d'exemple, l'entrée extensionnelle pour la forme fléchie *clarifiés* et la redistribution *passif* a la forme (simplifiée) suivante :

```
clarifiés v [pred='clarifier₁<arg₁:Suj:cln|scompl|sn,
             arg₀:Obl₂:(par-sn)>',@passive,@pers,@Kmp]; %passif
```

Cette entrée résulte des travaux sur les constructions impersonnelles décrits dans (Sagot & Danlos, 2007). Nous excluons donc de ce travail les constructions impersonnelles pronominales. Le lecteur attentif aura remarqué qu'il manque la construction *se-moyenne* (*Un problème comme celui-ci ne se clarifie pas aisément*). C'est tout l'objectif de ce travail que de combler ce types de lacunes dans le *Lefff*.

La section suivante passe en revue les différents types de constructions pronominales personnelles (objectifs 1 et 2 décrits plus haut), la section 2 décrit l'opération de fusion des informations lexicales (objectif 3), et la section 3 en décrit les résultats.

1. Classification des constructions pronominales

En s'inspirant de (Boons *et al.*, 1976b), nous classons les constructions pronominales en quatre classes :

- les constructions intrinsèques et autonomes, où le pronom *se* fait partie de l'unité lexicale verbale, et ne peut être mis en relation ni avec un des actants du verbe, ni avec un actant d'un autre emploi du même verbe,
- les constructions réfléchies et réciproques, où le pronom *se* réalise un actant du verbe coréférent à l'actant réalisé comme sujet,
- les constructions neutres et moyennes, où le pronom *se*, quoique n'étant pas un actant du verbe, peut être mis en relation avec un actant d'un emploi différent du même verbe,
- certains constructions, non traitées ici, où le pronom *se* est un argument non pas du verbe mais d'un des actants verbaux (par exemple, les constructions en *se...N_{pc}* (Nom partie du corps) de type *Pierre se lave les mains* qui correspond à *Pierre lave les mains de Pierre*).

Étudions successivement les trois premières classes de constructions pronominales.

1.1. Constructions intrinsèques et autonomes

Contrairement aux autres classes de constructions pronominales, les constructions intrinsèques et autonomes ne peuvent pas être mises en correspondance avec une construction non pronominale du même verbe. Le *se* fait partie intégrante du lexème verbal².

On parle de verbe *intrinsèquement* pronominal lorsqu'il ne peut être employé dans aucune construction non pronominale (*s'exclamer*; *s'évanouir*). On parle d'emploi pronominal *autonome* lorsque le même verbe peut prendre part à des constructions non pronominales, mais qui ne sont reliés aux emplois pronominaux par aucune relation productive claire, voir *se barrer* (*Luc s'est barré*) vs. *barrer* (*La police a barré la route*).

Dans le lexique-grammaire, les verbes intrinsèquement pronominaux et les emplois pronominaux autonomes « non productifs » font l'objet d'entrées propres réparties dans les tables appropriées, entrées dont la vedette inclut le pronom *se* (ou sa forme élidée *s'*). Toutefois, certaines constructions pronominales autonomes « productives » sont codées dans des colonnes. Par exemple, dans la Table 4 (N0 V N1_{hum}, N0 = Qu P + N, *Ceci attriste Marie*), la colonne intitulée N1 se V de ce que P (*Marie s'attriste de ceci*) code une construction pronominale autonome qui n'est pas en relation « transformationnelle » avec la construction transitive (Gross, 1975, 103); de même, la colonne intitulée N1 se V auprès de N_{hum} de ce que P (*Marie s'attriste auprès de Jean de ce que Luc ne le regarde pas*) code une construction autonome, qui compte un actant de plus par rapport à la construction transitive et qui prend un sens de communication verbale. On trouve des colonnes similaires dans les tables 11, 13 et 19 de (Gross, 1975).

Dans DICOVALENCE, toutes les entrées pronominales intrinsèques et autonomes (productives ou non) font l'objet d'entrées propres, typées par l'indication *pseudo_se* qui rappelle le caractère non valencielle de *se*. Toutefois, toutes les entrées de DICOVALENCE comportant l'indication *pseudo_se* ne sont pas à classer dans les pronominales intrinsèques et autonomes. Comme nous le verrons à la section 1.3, les constructions pronominales dites *neutres* font l'objet d'entrées autonomes, qui sont repérables par un lien explicite avec l'entrée transitive correspondante.

Dans le *Lefff*, le codage retenu répercute les caractéristiques énumérées ci-dessus. Le *se* est intégré à l'entrée (mais doit être exclu de la flexion), et surtout au prédicat qu'il dénote (et qui devient donc *se* Lemma ou *s'* Lemma et non Lemma

2. Pour les formes pronominales, l'auxiliaire de conjugaison est le verbe *être*. Mais pour certains locuteurs/auteurs, le *se* est tellement intégré au verbe que l'auxiliaire de conjugaison est *avoir*. Des requêtes effectuées sur tout Internet au moyen du moteur de recherche de Google (le 2 juin 2009) donnent ainsi « environ 410 » occurrences de *il a s'agi de*, à comparer aux « environ 127 000 » occurrences de *il s'est agi de*.

comme dans le cas d'une entrée non pronominale). Ainsi, le lemme verbal *évanouir* correspond à deux entrées dans le *Lefff* 3 :

```
évanouir1  s'Lemma;v;<arg0:Suj:c|n|sn>;@pseudo_se;%actif
évanouir2  s'Lemma;v;<arg0:Suj:c|n|sn,arg1:Loc:(loc-sn|y)>;
             @pseudo_se;%actif
```

La première correspond à *Luc s'évanouit / Luc est évanoui*, la seconde à *La voiture s'est évanouie (dans la brume)*.

1.2. Constructions réfléchies et réciproques

Les constructions pronominales réfléchies et réciproques se distinguent des autres par le fait que le pronom *se* mérite pleinement son appellation de pronom : il est la réalisation de surface d'une fonction syntaxique qui fait partie de la valence du verbe. Sa particularité est qu'il coréfère à l'actant réalisé comme sujet. Dans le cas d'une construction réfléchie (*Marie se ment (à elle-même)*), il y a identité entre l'« agent » et le « patient » (ou le « destinataire ») du procès. Dans le cas d'une construction réciproque, (*Pierre et Marie s'engueulent (l'un l'autre)*), le sujet est obligatoirement syntaxiquement pluriel et chacun des actants sujets pris comme agent peut être le patient du procès vis-à-vis des autres (*Pierre engueule Marie et Marie engueule Pierre*). Pour un verbe permettant la construction réfléchie et la construction réciproque, les formes pronominales à sujet pluriel sont ambiguës en l'absence d'une locution pronominale de type (*Prép*) *PRO-même* ou *l'un (Prép) l'autre*. Par exemple, *Pierre et Marie se caressent* reçoit l'interprétation réfléchie (*Pierre se caresse et Marie se caresse*) et l'interprétation réciproque (*Pierre caresse Marie et Marie caresse Pierre*).

Le lexique-grammaire ne renseigne pas explicitement la possibilité de réalisations réfléchies ou réciproques. Par contre, celles-ci sont codées systématiquement dans *DICOVALENCE* (ce qui est cohérent avec l'approche pronominale sur laquelle repose ce lexique) : *se refl.* et *se réc.* sont deux éléments possibles (et distincts) des paradigmes P1 (qui correspond approximativement à la fonction syntaxique 0bj) du *Lefff* et P3 (qui correspond approximativement à 0bj à)³.

Dans le *Lefff*, nous avons adopté une position similaire à celle de *DICOVALENCE*⁴ : *se_{refl}* et *se_{rec}* sont des réalisations possibles des fonctions

3. Dans *DICOVALENCE*, la locution pronominale *l'un (Prép) l'autre* est codée comme une réalisation syntaxique possible d'un paradigme (\approx fonction syntaxique). Elle apparaît le plus souvent dans des entrées permettant la forme pronominale réciproque. Toutefois, elle apparaît aussi dans des entrées intrinsèquement pronominales (*Luc s'acharne sur Marie*), pour coder une interprétation réciproque, *Ils s'acharnent l'un sur l'autre*, qui ne peut être marquée par le pronom *se* (**Ils se s'acharnent l'un sur l'autre*).

4. Par contre, nous n'avons pas encore codé la possibilité de réaliser une fonction syntaxique par une locution pronominale du type *l'un (Prép) l'autre*, voir note précédente.

syntaxiques Obj et Objâ. Ainsi, l'entrée décrivant l'emploi le plus fréquent de *parler* est la suivante :

```
parler1 Lemma;v;<arg0:Suj:c|n|sn,
arg1:Objde:(de-scompl|de-sinf|de-sn|en),
arg2:Objâ:(cld|serefl|serec|â-sn)>;@CtrlSujObjde;%actif
```

1.3. Constructions neutres et moyennes

Les constructions neutres et moyennes sont des constructions intransitives où le pronom *se* n'est pas la réalisation d'un actant du verbe. Toutefois, ces constructions peuvent être mises en rapport avec un autre emploi transitif du même verbe, selon le mécanisme suivant : l'objet de la construction transitive devient sujet de la construction pronominale, alors que le sujet de la construction transitive n'est plus réalisé. Deux cas se présentent (Ruwet, 1972) :

- les constructions dites *se-moyennes* ou à *agent fantôme*. Le sujet (agent) de la construction transitive est omis en syntaxe⁵ mais sous-entendu en sémantique, comme en témoigne la possibilité d'ajouter un adverbial orienté-agent, *Les grenouilles se mangent (avec les doigts + avec enthousiasme + énergiquement)* (Boons *et al.*, 1976b, 131), ou une infinitive causale, *Ce genre de maison, ça se brûle pour toucher la prime d'assurance* (Abeillé, 2002, 193).
- les constructions dites *se-neutres*. Le sujet de la construction transitive ne laisse aucune trace dans le procès, qui acquiert une sémantique d'« activité indépendante », généralement situable dans le temps. Aucun adverbial orienté-agent ne peut être ajouté à une construction *se-neutre*, *Le pont-levis s'est abaissé (de lui même + tout seul + *énergiquement)*⁶.

Dans le lexique-grammaire, seule la table 32C de (Boons *et al.*, 1976a) (N0 V N1, N1 concret, passif résultatif possible) comporte une colonne N1 *se* V décrivant la possibilité d'une construction *se-neutre* : *La misère avait émacié son visage / Son visage s'est émacié*. Il est dit dans (Gross, 1975, 102) que « (la transformation) *se-moyen* s'applique en général aux complétives directes des tables 6, 9, 10, 12 et 16 », voir *On dit partout qu'il a fait cela, Qu'il a fait cela se dit partout*. Cependant, cette propriété n'est pas codée dans les tables citées.

DICOVALENCE représente les constructions pronominales neutres et moyennes de façon différente. Une construction *se-moyenne* est considérée comme une reformulation de la construction transitive (cette reformulation est appelée *se*

5. Contrairement au passif, un complément d'agent en *par* est difficilement acceptable avec une construction *se-moyenne* : ? *Ce théorème se comprend par quiconque veut s'en donner la peine* (Boons *et al.*, 1976b, 131).

6. Il existe également des constructions neutres non pronominales. Dans certains cas, les deux constructions neutres sont possibles, et elles sont alors quasiment interchangeables, voir *Le plomb (E + se) liquéfie à 327,4 degrés Celsius* (Boons *et al.*, 1976b, 130).

passif). En revanche, une construction *se*-neutre est répertoriée comme une entrée distincte, qui partage les caractéristiques des entrées pronominales autonomes (indication *pseudo_se*), tout en s'en distinguant par un lien vers l'entrée représentant l'emploi transitif correspondant. Cette différence de traitement est à rapprocher de celle avancée par (Ruwet, 1972) — les constructions *se*-moyennes sont analysées par transformation, les *se*-neutres sont « engendrées dans la base » — ou de celle avancée par (Abeillé, 2002) — un noeud sujet N_0 vide pour les *se*-moyennes, pas de noeud N_0 pour les *se*-neutres. Chez (Dubois & Dubois-Charlier, 1997), les *se*-moyennes et les *se*-neutres sont codées comme des reformulations des formes transitives, avec la différence suivante : les *se*-moyennes (appelées *pronominales passives*) sont dérivées de la forme transitive, tandis que les *se*-neutres sont des formes premières dont sont dérivées les formes transitives (à sens factitif).

Dans le *Lefff*, nous avons appliqué des principes similaires à ceux de DICOVALENCE. Ainsi, une construction *se*-moyenne y est représentée comme une redistribution de l'entrée transitive, la redistribution *%se_moyen*. Cette redistribution a trois effets :

- l'arg₀ (réalisé comme sujet dans la construction transitive) est conservé, mais il n'est plus associé à la moindre fonction syntaxique (le cadre est donc de la forme $\langle \text{arg}_0 : : , . . . \rangle$) : il est présent en sémantique mais absent en syntaxe,
- l'objet de la construction transitive (arg₁) devient le sujet de la construction moyenne, moyennant les transformations nécessaires sur les différentes réalisations possibles,
- la macro *@se_moyen* est rajoutée, qui implique l'emploi de la particule pré-verbale *se*.

En revanche, les constructions *se*-neutres font l'objet d'entrées séparées, qui se présentent donc comme des entrées pronominales autonomes. Lorsque le *Lefff* comportera des identifiants sémantiques, il est prévu qu'une entrée *se*-neutre et l'entrée transitive associée partagent le même identifiant sémantique, à l'image des liens présents dans DICOVALENCE.

2. Fusion des ressources lexicales sur les constructions pronominales

Pour prendre en compte toutes ces données linguistiques dans le *Lefff*, nous avons exploité les informations présentes dans DICOVALENCE et le lexique-grammaire. Il s'agit d'un processus en trois étapes, qui est une simplification de la méthodologie décrite dans (Sagot & Danlos, 2008) :

- conversion des ressources de départ au format *Lefff*,
- regroupement des entrées des différentes ressources ainsi obtenues et de celles du *Lefff*,
- fusion des regroupements d'entrées.

2.1. Conversion au format *Lefff*

La structure des informations présentes dans le *Lefff* n'est pas très éloignée de celle de DICOVALENCE, en partie parce que le *Lefff* a repris à son compte des choix faits par DICOVALENCE. De façon générale, la conversion au format *Lefff* d'une entrée de DICOVALENCE se fait selon les principes suivants, qui sont des *approximations* :

- chaque *paradigme* de DICOVALENCE est converti en une *fonction syntaxique* du *Lefff*⁷ ;
- chaque pronom présent dans un paradigme est converti en une réalisation au sens du *Lefff* (ainsi, la présence du pronom *te* dans un paradigme *PI* conduit à l'ajout d'une réalisation *cla* — clitique accusatif — à la fonction *Obj*, perdant ainsi l'information sur la possibilité d'un objet humain) ;
- chaque reformulation de DICOVALENCE est convertie en redistribution équivalente du *Lefff* 3.

Le codage des constructions pronominales dans le *Lefff* suit les mêmes principes que DICOVALENCE, comme vu à la section précédente. Les principes généraux de conversion s'appliquent donc à toutes les entrées de DICOVALENCE qui sont pronominales ou qui admettent des réalisations pronominales.

L'extraction automatique d'entrées pronominales à partir du lexique-grammaire est plus difficile. D'une part, certaines tables ne sont pas disponibles librement. D'autre part, les constructions pronominales sont peu prises en comptes dans les tables, comme nous l'avons vu précédemment.

Nous nous sommes donc contentés d'extraire les informations suivantes :

- des lemmes verbaux intrinsèquement pronominaux ou admettant une construction pronominale autonome, en regardant les entrées des tables disponibles comportant une particule pré-verbale *se*, ainsi que leur cadre fonctionnel (les fonctions syntaxiques, mais pas leurs réalisations⁸),
- des lemmes verbaux admettant une redistribution *se*-moyenne (ainsi que leur cadre fonctionnel de base) à partir des tables 10 et 16, les tables 6, 9 et 12 n'étant pas distribuées,
- des lemmes verbaux de la table 32C constituant des entrées de type *se*-neutre.

Dans la suite de l'article, nous noterons E^i une entrée de la ressource *i* convertie au format *Lefff* (*i* étant donc DICOVALENCE, LG pour le lexique-grammaire ou *Lefff*).

7. Rappelons que ceci n'est qu'une approximation, et que DICOVALENCE n'utilise pas la notion de fonction syntaxique, mais celle de paradigme pronominal.

8. Extraire automatiquement les réalisations possibles pour chaque fonction est délicat en l'absence d'une « table des tables » complète et formalisée (cf. les travaux de (Gardent *et al.*, 2005)).

2.2. Regroupement des entrées provenant de différentes ressources

Pour un lemme donné, chaque ressource est susceptible de comporter plusieurs entrées. Il faut donc déterminer le nombre d'entrées que l'on souhaite obtenir au final. Chacune d'entre elles sera construite par fusion d'une ou de plusieurs entrées issues des ressources de départ, selon une méthode décrite plus bas. En effet, il faut au préalable construire ces ensembles d'entrées E^{Lefff} , $E^{\text{Dicovaleance}}$ et/ou E^{LG} , que nous appelons *regroupements*.

Ces regroupements sont définis à partir d'une relation d'*inclusion*, définie comme suit. Pour un lemme verbal donné, une entrée E^1 est *incluse dans* ou *plus spécifique* qu'une entrée E^2 ssi l'ensemble des énoncés décrits par E^1 est inclus dans l'ensemble des énoncés décrits par E^2 .

La mise en œuvre dans ce travail de cette définition générale repose sur deux approximations.

1. Nous avons considéré que les entrées les plus spécifiques sont les entrées $E^{\text{Dicovaleance}}$. Les entrées E^{Lefff} le sont un peu moins, et les entrées E^{LG} le sont encore moins⁹. Nous avons donc cherché, pour chaque lemme verbal, les relations d'inclusion entre entrées de la forme $E^{\text{Dicovaleance}} \subset E^{\text{Lefff}}$ et $E^{\text{Lefff}} \subset E^{\text{LG}}$.
2. Nous avons choisi l'heuristique suivante : une entrée E^1 est incluse dans une autre entrée E^2 si elle a exactement le même inventaire de fonctions syntaxiques de base (Suj, Obj, Objde, Objà) et si l'inventaire des autres fonctions syntaxiques dans E^2 est inclus dans celui de E^1 .

C'est ainsi que pour le lemme verbal *voler*, l'entrée $E_{86440}^{\text{Dicovaleance}}$ (issue de l'entrée 86440 de DICOVALENCE), à savoir

```
voler86450Dicovaleance Lemma;v;<arg0:Suj:sn|cln,arg1:Obj:(sn|cla),
arg2:Objà:(â-sn|cld)>;;%actif,%passif,%se_moyen
```

est considérée comme incluse dans l'entrée du *Lefff*

```
voler2Lefff Lemma;v;<arg0:Suj:cln|scompl|sinf|sn,arg1:Obj:(cla|sn),
arg2:Objà:(cld|â-sn)>;;%actif,%passif
```

On notera que la fonction Suj a dans l'entrée du *Lefff* des réalisations erronées *scompl* et *sinf* (complétive et infinitive).

Les regroupements d'entrées sont alors obtenus comme suit : nous partons de chaque entrée qui n'inclut aucune autre entrée (souvent, une entrée issue de DICOVALENCE) puis nous suivons les relations d'inclusion jusqu'à trouver une entrée qui n'est incluse dans aucune autre. Chaque chaîne d'inclusion ainsi construite forme un regroupement.

9. C'est parce que nous n'extrayons des tables que des informations très partielles que la granularité est si réduite. Les tables du lexique-grammaire fournissent en réalité des informations syntaxiques très riches — sauf précisément pour les constructions pronominales.

2.3. Fusion des entrées

Les entrées qui forment un regroupement sont fusionnées de la façon suivante :

- l’ensemble des fonctions syntaxiques est construit comme l’union des ensembles des fonctions syntaxiques présentes dans les entrées sources,
- pour chaque fonction, l’ensemble des réalisations est également obtenu par union, en indiquant quelle source propose quelle réalisation (aucune indication sur une réalisation particulière indique qu’elle était répertoriée dans toutes les entrées sources),
- une fonction syntaxique n’est à réalisation obligatoire que si c’est le cas dans toutes les ressources sources,
- l’ensemble des redistributions possibles est construit comme l’union des ensembles de redistributions possibles présentes dans les entrées sources.

La fusion des deux entrées ci-dessus pour le lemme verbal *voler* donne ainsi le résultat suivant :

```
voler2 v 100;Lemma;v;<arg0:Suj:c|n|scomplLeff|sinfLeff|sn,
      arg1:Obj:(sn|cla),arg2:Objà:(cld|à-sn)>;
      ;%actif,%passif,%se_moyen
```

La redistribution *se-moyenne* a donc été ajoutée à l’entrée initiale du *Leff*. De plus, les indications de provenance permettront d’améliorer la qualité et l’efficacité de la validation manuelle, qui reste indispensable pour obtenir une ressource de bonne qualité. C’est ainsi que les réalisations *scompl* et *sinf* pour la fonction *Suj*, que seul le *Leff* propose, seront plus facilement identifiées comme erronées.

3. Résultats, conclusion et perspectives

Les résultats préliminaires de ce travail sont les suivants. Nous avons extrait de DICOVALENCE 5 273 entrées pronominales ou ayant des réalisations pronominales, parmi les 8 214 entrées que comporte la ressource. Le lexique-grammaire a permis la construction de 550 telles entrées, mais celles-ci ne comportent aucune information sur les réalisations des fonctions syntaxiques. La ressource construite à partir de ces entrées et du *Leff* comporte 5 464 entrées pronominales ou ayant des réalisations pronominales, qu’il nous faudra valider manuellement.

Les constructions pronominales sont un exemple de données linguistiques moins bien couvertes que d’autres par les tables du lexique-grammaire. À l’inverse, une ressource comme DICOVALENCE, de par ses fondements linguistiques, s’est fortement intéressée à ces constructions. Naturellement, ce n’est pas toujours le cas : les constructions symétriques, par exemple, sont bien étudiées dans le lexique-grammaire, mais difficilement modélisables dans le cadre de l’approche pronominale et donc dans DICOVALENCE. C’est donc bien par la mise en commun des données linguistiques présentes dans différentes ressources qu’il sera possible de parvenir à une description syntaxique lexicale satisfaisante du français.

Œuvres citées

- Abeillé A. 2002. *Une grammaire électronique du français*. Paris, France : CNRS Editions.
- Boons J.-P., Guillet A. & Leclère C. 1976a. *La structure des phrases simples en français, Classes de constructions transitives*. Rapport interne, LADL, CNRS, Paris 7.
- Boons J.-P., Guillet A. & Leclère C. 1976b. *La structure des phrases simples en français, Constructions intransitives*. Genève : Droz.
- Danlos L. & Sagot B. 2007. Comparaison du lexique-grammaire des verbes pleins et de dicovalence : vers une intégration dans le lefff. In *Actes de TALN'2007*, Toulouse, France.
- Dubois J. & Dubois-Charlier F. 1997. *Les verbes français*. Paris, France : Larousse-Bordas.
- Gardent C., Guillaume B., Perrier G. & Falk I. 2005. Maurice Gross' grammar lexicon and natural language processing. In *Proc. of the 2nd LTC*, Poznań, Poland.
- Gross M. 1975. *Méthodes en syntaxe*. Paris, France : Hermann.
- Leclère C. 2002. Organisation of the lexicon-grammar of french verbs. *Linguisticae Investigationes*, **25**(1), 29–48.
- Ruwet N. 1972. *Théorie syntaxique et syntaxe du français*. Paris, France : Le Seuil.
- Sagot B., Clément L., Villemonte de La Clergerie E. & Boullier P. 2006. The Lefff 2 syntactic lexicon for French : architecture, acquisition, use. In *Proc. of LREC'06*.
- Sagot B. & Danlos L. 2007. Améliorer un lexique syntaxique à l'aide des tables du lexique-grammaire – Constructions impersonnelles. *Cahiers du Cental*. (à paraître).
- Sagot B. & Danlos L. 2008. Méthodologie lexicographique de constitution d'un lexique syntaxique de référence pour le français. In *Actes du colloque Lexicographie et informatique : bilan et perspectives*, Nancy, France.
- Van den Eynde K. & Mertens P. 2006. Le dictionnaire de valence DICOVALENCE : manuel d'utilisation. http://bach.arts.kuleuven.be/dicovalence/manuel_061117.pdf.

Résumé

Dans cet article, nous décrivons la façon dont les constructions pronominales sont représentées dans DICOVALENCE et dans le lexique-grammaire. Nous présentons une méthode pour l'extraction et la fusion d'informations lexicales syntaxiques sur ces constructions, et pour leur intégration dans le lexique TAL Lefff.

Summary

Pronominal constructions in Dicovalence and in the Lexicon-Grammar — Integration in the Lefff

In this paper, we describe how pronominal constructions are represented in DICOVALENCE and in the lexicon-grammar. We introduce a method for extracting and merging lexical syntactic information about these constructions, and integrating it in the Lefff NLP lexicon.

Adresse de l'auteur:

Laurence Danlos, Benoît Sagot
Alpage, INRIA & Université Paris 7
30 rue du Ch. des rentiers
75013 Paris, France