



Sous-catégorisation en pour et syntaxe lexicale

Benoît Sagot, Laurence Danlos, Margot Colinet

► **To cite this version:**

Benoît Sagot, Laurence Danlos, Margot Colinet. Sous-catégorisation en pour et syntaxe lexicale. Traitement Automatique du Langage Naturel 2014, Jul 2014, Marseille, France. hal-01022351

HAL Id: hal-01022351

<https://hal.inria.fr/hal-01022351>

Submitted on 11 Jul 2014

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.

Sous-catégorisation en *pour* et syntaxe lexicale

Benoît Sagot¹ Laurence Danlos¹ Margot Colinet^{1,2}
(1) Alpage, INRIA & Université Paris-Diderot, 75013 Paris
(2) LLF, CNRS & Université Paris-Diderot, 75013 Paris
{prenom.nom}@inria.fr

Résumé. La sous-catégorisation d'arguments introduits par la préposition *pour* a été sous-étudiée par le passé, comme en témoigne l'incomplétude des ressources lexico-syntaxiques existantes sur ce point. Dans cet article, nous présentons rapidement les différents types de sous-catégorisation en *pour*, qui contrastent avec les emplois de *pour* comme connecteur de discours. Nous décrivons l'intégration des arguments en *pour* au lexique syntaxique *Lefff*, enrichissant ainsi les informations de sous-catégorisation de nombreuses entrées verbales, nominales, adjectivales et adverbiales.

Abstract. Sub-categorized arguments introduced by the French preposition *pour* has been under-studied in previous work, as can be seen from the incompleteness of existing lexical-syntactic resources in that regard. In this paper, we briefly introduce the various types of sub-categorization in *pour*, which are to be distinguished from occurrences of *pour* as a discourse connective. We describe how we added arguments in *pour* within the syntactic lexicon *Lefff*, thus refining sub-categorization information for many verbal, nominal, adjectival and adverbial entries.

Mots-clés : Sous-catégorisation, Arguments en *pour*, Lexiques syntaxiques.

Keywords: Sub-categorization, Arguments in *pour*, Syntactic lexicons.

1 Introduction

Le développement de ressources lexicales syntaxiques pour le français est encore aujourd'hui un champ de recherche ouvert, comme en attestent les nombreux efforts réalisés dans cette direction par diverses équipes, parfois depuis plusieurs dizaines d'années. De telles ressources constituent des bases de données linguistiques descriptives, mais sont également utiles pour différentes tâches de TAL, et notamment en analyse syntaxique symbolique (Riezler *et al.*, 2002; Villemonte De La Clergerie, 2013) et même statistique (Collins, 1997; Mirroshandel *et al.*, 2013). Elles sont enfin au cœur de nombreux travaux à l'interface entre la syntaxe et d'autres niveaux d'analyse linguistique, et notamment la sémantique et le discours.

Nous nous intéressons dans ce travail aux syntagmes prépositionnels en *pour*, en mettant l'accent sur les infinitives. Ces syntagmes sont intéressants pour plusieurs raisons. Tout d'abord, et bien que dans certains cas au moins ils constituent clairement des arguments d'éléments prédicatifs (verbes, noms, adjectifs, voire adverbes), ils ont été relativement peu étudiés et sont assez mal représentés au sein des différentes ressources actuelles : tables du lexique-grammaire (Gross, 1975; Tolone, 2011), Dicovalence (van den Eynde & Mertens, 2006), le Lexique des Verbes Français (Dubois & Dubois-Charlier, 1997) ou le *Lefff* (Sagot, 2010). Ceci vient au moins en partie de ce que les tests les plus élémentaires pour distinguer arguments et modificateurs reposent sur la cliticisation, opération qui n'est jamais possible directement à partir d'un argument réalisé sous la forme d'un syntagme prépositionnel en *pour*.

Par ailleurs, l'étude des compléments en *pour*, et singulièrement des infinitives en *pour*, est indispensable à l'interface entre syntaxe et discours. Ainsi, dans le cadre du développement du French Discourse TreeBank (FDTB) (Danlos *et al.*, 2012), la première étape consiste à repérer toutes les occurrences de connecteurs du discours. Ceci suppose, pour les unités lexicales qui peuvent avoir un rôle de connecteur, de savoir faire le départ entre les occurrences où elles sont connecteurs et les occurrences où elles ne le sont pas. Colinet *et al.* (2014) ont étudié les emplois de *pour* comme connecteur de discours lorsqu'il introduit une infinitive, en mettant au point un ensemble de critères formels permettant de les distinguer des emplois comme introducteur d'infinitives sous-catégorisées. Outre les emplois comme connecteurs (1a) et les emplois sous-catégorisés (1b), cette étude, sur laquelle repose l'amélioration du *Lefff* présentée ici, a identifié deux autres types d'emplois plus spécifiques : ceux introduisant une « relative sans mot QU » (Huddleston & Pullum, 2002) (1c) et les emplois métadiscursifs, dont trois exemples sont donnés en (1d).

- (1) a. Luc a fait une bouillabaisse pour faire plaisir à Marie.
- b. Il a suffi d'un café à Marie pour se remettre d'aplomb.
- c. Un pont pour franchir l'Amazone a été construit en 1745.
- d. Pour conclure. Pour ne citer que lui. Pour le dire autrement.

L'objectif du travail présenté ici est double : (i) réaliser une étude comparative du traitement des différents types d'arguments en *pour* dans le lexique-grammaire, Dicovalence, le Lexique des Verbes Français et le *Lefff* ; et (ii) intégrer le résultat dans le *Lefff*, en s'aidant notamment des résultats de l'annotation manuelle des occurrences de *pour* (Colinet *et al.*, 2014) dans le French TreeBank (Abeillé *et al.*, 2003).

Cet article est organisé comme suit. Dans un premier temps, nous passons en revue rapidement certains travaux antérieurs sur la sous-catégorisation en *pour* (section 2). Ensuite, nous présentons de façon succincte mais systématique les différents types de sous-catégorisation en *pour*, et nous les différencions des emplois de *pour* comme connecteur de discours (section 3). Nous présentons enfin la façon dont nous avons intégré ces résultats au lexique *Lefff* (section 4).

2 Travaux antérieurs

Outre les travaux de Cadiot (1990a; 1990b) et la grammaire de Wyler (2014), les ressources que nous avons étudiées sont les tables du lexique-grammaire (Gross, 1975; Tolone, 2011), Dicovalence (van den Eynde & Mertens, 2006), le Lexique des Verbes Français (Dubois & Dubois-Charlier, 1997) et le *Lefff* (Sagot, 2010). Mais dans toutes ces ressources, la sous-catégorisation (verbale, nominale, adjectivale, adverbiale) en *pour* y est mal décrite, notamment dans les nombreux cas où *pour* alterne avec *à* ou avec *de*, où souvent seul *à* ou *de* est mentionné. Il en est ainsi en (2) et (3), peut-être sous l'influence du procédé de cliticisation disponible, qui est le procédé habituel pour les objets indirects en *à* ou *de*.

(2) Luc a préposé Max à/pour garder l'entrepôt. / Luc l'y a préposé.

(3) Luc complimente Max de/pour avoir fini à temps. / Luc l'en a chaleureusement complimenté.

Dans le Lexique des Verbes Français (LVF) (Dubois & Dubois-Charlier, 1997), la préposition *pour* est indiquée dans les codes de sous-catégorisation par deux lettres distinctes : *k* lorsqu'elle alterne avec *contre*, et *q* sinon. Seuls 34 verbes sont répertoriés comme sous-catégorisant un argument en *pour/contre*. Tous font partie de la classe C1 des verbes de communication à sujet humain ou animal, avec l'opérateur « loq PR/CT » signifiant « parler pour (PR) ou contre (CR) ». De manière générale, on peut du reste noter que si *pour* permet (généralement) d'introduire un argument nominal, infinitif ou phrastique, *contre* n'introduit que des arguments nominaux. On peut également remarquer que, dans ces cas d'alternance *pour/contre*, ces prépositions peuvent généralement être employées sans objet (*J'ai voté pour*). À côté de ces 34 verbes, 280 autres entrées verbales sous-catégorisant un argument en *pour* sont répertoriés, où *pour* n'alterne pas avec *contre*.

Dans les Tables verbales du Lexique-Grammaire, des arguments en *pour* sont indiqués de deux façons :

- ils peuvent faire l'objet d'une propriété dédiée dont la validité est associée à chaque entrée d'une des classes. C'est par exemple le cas de la propriété $N_0 V N_1 \text{ à } N_2$ *pour* $D_{num} N_{monnaie}$ codée dans la table 36DT¹, qui est associée par exemple aux verbes *acheter*, *brader*, etc. (cf. *Pierre a bradé ce livre (à Marie) pour 2 euros*). On retrouve également les propriétés $Prép_2 = \textit{pour}$ et propriété $Prép_3 = \textit{pour}$ dans la table 38RR².
- ils peuvent être directement indiqués dans les tables, sous la forme d'un codage explicite de la préposition *pour*. Par exemple, dans la table 1 (verbes à un argument post-verbal prépositionnel qui peut être phrastique), la préposition introduisant l'argument post-verbal est indiquée explicitement. Ainsi, dans cette table, l'entrée du verbe *insister* indique que cette préposition est *pour* (cf. *Pierre a insisté pour partir / pour que Marie parte*).

Au total, 176 entrées verbales dans les tables font usage d'une façon ou d'une autre de la préposition *pour*.

Dans Dicovalence, les arguments en *pour* sont indiqués dans le « paradigme prépositionnel ». Seules 103 entrées incluent un argument en *pour*, dont 9 en alternance avec *contre*. Parmi les 94 autres, 37 sont indiqués comme pouvant être réalisés par (la pronominalisation d')une infinitive.

Enfin, dans le *Lefff* (version 3.3), 116 entrées verbales sous-catégorisent un argument dont une réalisation en *pour* est possible. Il s'agit pour la plupart d'entrées obtenues par fusion avec Dicovalence ou le LVF (Sagot & Danlos, 2012). On trouve également 46 adjectifs, 5 noms et 4 adverbes qui sous-catégorisent un argument permettant une réalisation en *pour*.

La sous-catégorisation non-verbale d'arguments en *pour*, c'est-à-dire les arguments en *pour* de prédicats nominaux, adjectivaux ou adverbiaux a été quant à elle encore moins étudiée.

1. Classe des verbes « datifs » (structure de base $N_0 V N_1 \text{ à } N_2$, avec N_2 cliticisable en *lui* ; cf. *Zoé donne un livre à Max / Zoé lui donne un livre*).

2. Classe des verbes « résiduels doubles » (structure de base $N_0 V N_1 Prép N_2 Prép N_3$ et qui ne rentrent pas dans une autre classe plus spécifique).

3 La sous-catégorisation en *pour*

3.1 Critères et annotation des données du French TreeBank

Comme indiqué précédemment, l'un des principaux obstacles à l'étude de la sous-catégorisation en *pour* est la difficulté qu'il y a à les identifier formellement. En effet, si les « relatives sans mot QU » et les *pour* introducteurs d'expressions métadiscursives sont faciles à identifier, la distinction entre *pour* connecteurs de discours et *pour* introduisant un argument sous-catégorisé n'est pas aisée. Il s'agit en effet d'une instance particulièrement délicate du problème général de la distinction entre arguments et modificateurs.

Colinet *et al.* (2014) rappellent deux critères classiques pour distinguer les arguments des modificateurs (critères 1 et 2 ci-dessous), et en proposent d'autres qui s'appliquent spécifiquement au cas de la préposition *pour*. Certains de ces critères s'appliquent correctement à toutes les catégories prédicatives, d'autres ne concernent que la sous-catégorisation verbale.

1. Un argument est *sémantiquement obligatoire*, pas un ajout (on notera que le caractère *sémantiquement obligatoire* n'implique pas que sa réalisation en syntaxe soit obligatoire). On peut tester ce critère en tentant de supprimer l'argument en *pour* et en vérifiant l'acceptabilité et le changement éventuel de sens. Ainsi, dans *Luc a fait une bouillabaisse pour faire plaisir à Marie*, l'infinitive est effaçable sans changement d'acceptabilité. En revanche, dans *Luc s'est dépêché pour finir*, l'effacement de l'infinitive conduit à un énoncé dont l'interprétation nécessite de récupérer l'argument effacé dans le contexte discursif gauche.
2. Un ajout est plus facilement antéposable qu'un argument. Il s'agit d'une tendance et non d'un critère strict.
3. Commutation avec *à, de, contre*. Comme évoqué précédemment, *pour* commute dans certains cas avec l'une de ces prépositions introduisant un complément considéré par les ressources existantes comme sous-catégorisé. Nous y revenons plus en détail ci-dessous.
4. En général, lorsque le connecteur de discours *pour* introduit une infinitive, cette infinitive est contrôlée par le sujet (le sujet de l'infinitive est coréférent au sujet de la principale). Si donc on est face à un cas de contrôle d'une infinitive en *pour* par un argument autre que le sujet, il y a de fortes chances qu'il s'agisse d'un cas de sous-catégorisation.
5. De manière générale, on ne peut rajouter la séquence « *et ce,* » avant *pour* que dans les cas où *pour* est connecteur (cf. *Luc a fait une bouillabaisse et ce, pour faire plaisir à Marie* vs. **Luc s'est dépêché et ce, pour finir*). Certains cas sont toutefois moins clairs (?*Luc a pris un taxi et ce, pour aller à l'aéroport*).
6. Les possibilités de réfutation diffèrent selon que *pour* introduit un argument ou un modificateur. Ainsi, si l'on répond « *c'est faux* » à une assertion comme *Luc s'est dépêché pour finir*, on affirme que Luc ne s'est pas dépêché. À rebours, si l'on répond « *c'est faux* » à une assertion telle que *Luc a fait une bouillabaisse pour faire plaisir à Marie*, on affirme soit que Luc n'a pas fait de bouillabaisse, soit qu'il en a faite une mais dans un autre but que celui de faire plaisir à Marie. Nous ne rentrerons pas dans les détails, mais d'autres questions sémantiques liées aux interrogatives et aux propriétés de factualité vont dans le même sens.
7. Si l'on dispose d'un exemple où deux arguments en *pour* sont rattachés au même verbe sans être coordonnés, alors le premier est susceptible d'être sous-catégorisé, et le second est nécessairement modificateur (autrement dit, le second *pour* est un connecteur de discours).

Grâce à ces critères, Colinet *et al.* (2014) ont annoté manuellement toutes les occurrences de *pour* introduisant une infinitive dans le French TreeBank. Les syntagmes prépositionnels nominaux ont été laissés de côté, l'annotation posant un problème supplémentaire, à savoir le fait de déterminer si le GN introduit par *pour* réfère à une éventualité ou non. Sur les 1161 occurrences ainsi annotées, 558 introduisent des compléments sous-catégorisés, 518 sont des connecteurs de discours, 52 introduisent des « relatives sans mot QU », 33 introduisent des expressions métadiscursives.

3.2 Sous-catégorisation verbale en *pour*

Nous avons confronté les résultats de cette annotation aux données fournies par les différentes ressources lexicales mentionnées ci-dessus, en ayant parfois eu recours de façon complémentaire à l'introspection et à des recherches sur internet. Nous en sommes arrivés à l'inventaire suivant d'arguments sous-catégorisés en *pour*, qui reprend le découpage des emplois verbaux selon les tables du lexique-grammaire.³ Dans la suite, la notation V^i -*inf* dénote une infinitive dont le sujet est coréférent au i -ième argument du verbe (0 pour le sujet, par exemple), et W dénote toute séquence de dépendants (par exemple, V^0 -*inf* W dénote une infinitive complète dont le sujet coréférent au sujet du verbe de la principale).

3. Faute de place, nous ne discutons pas ici des locutions verbales figées sous-catégorisant un argument en *pour*.

Table 1 : N_0 V pour V^0 -inf W La majorité des cas d’alternance *pour/contre* rentrent dans ce cadre là (*batailler, se battre, manifester...*). On peut également citer un exemple d’alternance *à/pour* (*œuvrer*) et un exemple d’alternance *de/pour* (*se dépêcher*). On trouve également dans le FTB des cas sans alternance (*foncer, se précipiter, s’endetter, intervenir...*), y compris des verbes symétriques dans leur emploi à sujet conjoint (N_0 plur se retrouvent pour V^0 -inf W / N_0 se retrouve avec N_1 pour V^{0+1} -inf W), une entrée lexicale excluant l’assertif positif (*ne pas se gêner*) et quelques cas un peu particuliers (*aller (pour)* et *être (pour)* au sens de *être sur le point (de), être parti (pour)*).

Table 2 : N_0 V Loc N_1 pour V^0 -inf W Seuls trois exemples trouvés dans le FTB : *venir, se rendre* et *sortir* (ce dernier imposant que la préposition *Loc* soit *de*). L’infinitive peut être réalisée avec ou sans la préposition *pour*.

Table 3 : N_0 V N_1 Loc N_2 pour V^1 -inf W Seul un exemple trouvé : *envoyer*. Ici aussi, l’infinitive peut être réalisée avec ou sans la préposition *pour*.

Table 9 : N_0 V à N_1 (qu P / pour V^1 -inf W) Quelques verbes de communication (*écrire, s’adresser...*) et quelques cas particuliers (*recourir, s’en remettre, se substituer*).

Table 11 : N_0 V (à ce qu P / à/pour V^i -inf W) Les deux types de contrôle sont attestés dès lors que l’argument en *pour* est une infinitive : $i = 0$ (contrôle par le sujet) et $i = 1$ (contrôle par l’objet). Exemples de contrôle par le sujet : *employer, nommer, utiliser*; exemples de contrôle par l’objet : *utiliser* (ambigu avec un contrôle par le sujet). Dans certains cas, seule la préposition *pour* est possible, et *à* est impossible : il en est ainsi par exemple de *choisir, désigner, retenir* (contrôle par le sujet) ou de *choisir* (qui est donc ambigu), *délaisser, invoquer* (contrôle par l’objet). Un cas spécifique est celui de verbes dont l’objet direct non humain est introduit par un déterminant possessif référant obligatoirement au sujet (*Max a dévoué sa/*ma vie à/pour soigner les bêtes*).

Tables 12 et 13 : N_0 V N_1 de/pour V^1 -inf W Seuls des verbes sémantiquement proches de *complimenter, punir, récompenser* sont attestés.

Table 16 : N_0 V Prép N_1 pour V^i -inf W Les deux types de contrôle sont attestés dès lors que l’argument en *pour* est une infinitive : $i = 0$ (contrôle par le sujet) et $i = 1$ (contrôle par l’objet). Plusieurs cas selon la valeur de *Prép* : avec *Prép = sur* on trouve par exemple *s’appuyer, compter*; avec *Prép = de* on a notamment *profiter*; avec *Prép = auprès de* on peut citer *s’activer*; avec *Prép = avec* on retrouve notamment les verbes symétriques dont la construction à sujet conjoint a été mentionnée au titre de la table 1.

Table 18 : N_0 V Prép N_1 Prép N_2 (de/pour V^1 -inf W / que Psubj) Plusieurs exemples hétérogènes, comme *enrôler* (N_1 Loc N_2 pour V^1 -inf W) ou *prétexter* (*de* N_1 auprès de N_2 pour V^1 -inf W).

Falloir et suffire On trouve un certain nombre d’attestations relevant des types suivants :

- *falloir* N_1 durée (à N_1) pour V^1 -inf W (*Il a fallu deux heures à Luc pour finir*; cf. également le point suivant),
- *falloir* (V -inf W / qu P / N_1) pour V -inf W (*Il faut manger pour vivre*),
- *suffire* (à N_1) (*de* V^1 -inf W / qu P / *de* N_2) pour V^1 -inf W (*Cela suffit à Marie pour être heureuse*),

Pour attributifs *passer pour, prendre N_1 pour*, etc.

Cas avec restrictions fortes sur le N_1 Quelques cas sont fortement liés à des classes de noms en position d’objet :

- N_0 argent/attribution : *avoir, déboursier, dépenser, emprunter*, et d’autres ; parallèlement, on trouve *donner* (avec un à N_2 et un contrôle par le N_2), *obtenir* et *recevoir* (avec un *de* N_2 et un contrôle par le N_2) et *manquer* (avec un à N_1 et pas de contrôle prédéfini).
- N_0 durée/date : *mettre, passer, se donner, prendre* une durée pour faire quelque chose, *donner* une durée à quelqu’un pour faire quelque chose, *disposer* d’une durée, *avoir jusqu’à, attendre (jusqu’à)* une date, *attendre* une durée pour (avant de) faire quelque chose.

3.3 Sous-catégorisation nominale, adjectivale et adverbiale en *pour*

La plupart des **noms** qui sous-catégorisent un argument en *pour* peuvent se regrouper en quelques classes sémantiques :

- Noms de type *méthode*, comme *moyen, mesure, solution*. L’argument en *pour* (nominal, infinitif ou phrastique) alterne avec un argument nominal en *de* ; pour *moyen*, il peut être infinitif ou phrastique (en *que* et non *de ce que*).
- Noms de type *combat*, comme *lutte, bataille, négociations, pression, surenchère*. Ici *pour* alterne le plus souvent avec *contre* (cf. cependant *concurrence*). Dans certains cas, on peut identifier un autre argument en *avec*.
- Noms de type *qualité*, comme *atout(s), imagination*.
- Noms de type *effort*, comme *intervention, tentative, démarche(s)*
- Noms de type *réunion*, comme *rassemblement, commission, partenariat, accord*. Dans certains cas *pour* alterne avec *contre*. Dans certains cas également, on peut identifier un autre argument en *de* (et parfois *entre*) voire un troisième argument en *avec* (cf. *l’accord des parties / entre les parties / de X avec Y pour V-inf W*).

- Noms de type *aide*, comme *assistance*, *contribution*. La préposition *pour* alterne souvent avec *à*. On peut identifier un autre argument en *de* (cf. *l'aide de Marie à/pour la bonne marche de la cérémonie*).
- Noms de type argent ou temps (cf. ci-dessus) : *dépenses*, *dette*, *échéance*. Autres compléments sont identifiables.
- Autres classes : noms de type *candidat*, *volontaire* ; noms de type *mandat*, *mission* (avec autre argument en *de*) ; noms de type *appel*, *ultimatum* (avec autre argument en *de*). Dans certains cas, *pour* alterne avec *à*.
- Autres cas : *argument*, *problème*, *condition*...

Un certain nombre d'**adjectifs** ont également été identifiés, dont les plus fréquents sont *nécessaire*, *indispensable*, *idéal*, *suffisant* et *insuffisant*. Enfin, quelques **adverbes** sous-catégorisent également un argument en *pour* : *suffisamment*, *insuffisamment*, *trop* et *assez*.

4 Enrichissement du *Lefff*

4.1 Le *Lefff*

Le *Lefff* (Lexique des Formes Fléchies du Français) (Sagot, 2010) est un lexique morphologique et syntaxique librement disponible. Il est développé dans le cadre du formalisme Alexina, en partie au moyen de techniques automatiques ou semi-automatiques, mais avec une forte composante manuelle. Le *Lefff* est conçu pour être linguistiquement pertinent tout en pouvant être directement utilisé dans des systèmes de TAL. Il repose sur la notion de redistribution : le lexique *intensionnel*, édité par les développeurs du lexique, associe à chaque entrée un cadre de sous-catégorisation initial 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 et pour chacune des formes fléchies compatibles.

Considérons par exemple l'entrée intensionnelle simplifiée suivante, tirée de la version 3.3 du *Lefff* :

```
délaisser1 Lemma:v;<Suj:clnlsn,Obj:(clalséréclnlséréfl),Obl:(pour-snlpour-sinflpour-scompl)>;@CtrlSujObl;
%actif,%passif,%se_moyen,%ppp_employé_comme_adj
# Lien externe: Dicovalence(délaisser25480) # Exemple: Paul a délaissé sa famille pour aller se battre
```

Cette entrée décrit un verbe ayant trois arguments : un argument dont la fonction syntaxique est la fonction sujet, qui peut être réalisé par un clitique ou un syntagme nominal ; un objet direct dont la réalisation est facultative (liste entre parenthèses) et qui peut prendre la forme d'un clitique, d'un syntagme nominal ou d'un *se* (réfléchi ou réciproque) ; et enfin un argument oblique en *pour*, nominal ou phrastique (infinitive ou complétive). L'information selon laquelle l'argument oblique, s'il est infinitif, est contrôlé par le sujet est donnée également. Enfin, la liste des redistributions admissibles (actif, passif, *se* moyen, participe passé employé comme adjectif) est indiquée. Une équivalence avec l'entrée correspondante de Dicovalence est donnée, ainsi qu'un exemple (tiré, pour cette entrée, ici de Dicovalence).

4.2 Intégration des arguments en *pour*

À partir de l'étude évoquée à la section précédente, nous avons complété manuellement les entrées lexicales du *Lefff* avec les arguments réalisables en *pour*, ou avons rajouté à des arguments existants les réalisations en *pour* lorsqu'elles étaient manquantes. Dans les cas d'alternance en *à* ou en *de*, il a fallu à chaque fois déterminer la fonction syntaxique de l'argument en présence : *Objà/Objde* (objet indirect) ou *Obl* (oblique). Par ailleurs, dans un certain nombre de cas où un même lemme a plusieurs sens et donc plusieurs entrées dans le *Lefff*, il nous a fallu identifier celle(s) des entrées qui étaient concernée(s). Pour certains verbes, un changement de l'inventaire d'entrées s'est avéré nécessaire.

Après ce travail, et outre les nombreuses améliorations et corrections réalisées en cours de route, la couverture du *Lefff* en ce qui concerne les arguments en *pour* a considérablement augmenté. Ainsi, le nombre d'entrées verbales admettant un argument en *pour* est passé de 116 à 269. Pour les noms, l'augmentation est de 5 à 84, et de 46 à 53 pour les adjectifs. Les 4 adverbes concernés étaient déjà correctement renseignés dans le *Lefff*. Outre l'ajout de ces réalisations en *pour*, les alternances en *contre*, *à*, *de* et autres prépositions (y compris la préposition vide) ont été renseignées, ainsi que les informations de contrôle. La version du *Lefff* ainsi obtenue est numérotée 3.4.

5 Conclusion et perspectives

Nous avons présenté une étude de la sous-catégorisation en *pour* en français, question insuffisamment étudiée jusqu'à présent, notamment en raison de la difficulté qu'il y a à trouver des critères pertinents pour distinguer arguments et modifieurs introduits par cette préposition. La perspective discursive nous a été ici utile, les occurrences de *pour* introduisant des infinitives modifieurs étant des connecteurs de discours. Nous avons montré brièvement comment nous avons complété le *Lefff* grâce au résultat de ce travail, qui s'appuyait sur les ressources lexicales existantes et sur une annotation manuelle de tous les *pour* introduisant une infinitive dans le French TreeBank.

Les perspectives de ce travail sont doubles. Tout d'abord, étant donnée l'homogénéité sémantique d'une grande partie des lexèmes concernés, nous comptons nous appuyer sur le wordnet WOLF pour extraire de nouvelles entrées susceptibles de sous-catégoriser des arguments en *pour*, à partir de leur proximité lexico-sémantique par rapport à des lexèmes déjà identifiés comme tels. Ensuite, nous souhaitons obtenir des résultats comparatifs en analyse syntaxique symbolique avant et après intégration au *Lefff* des résultats de cette étude, en nous appuyant sur l'analyseur syntaxique FRMG (Villemonte De La Clergerie, 2013). Il est possible toutefois que ce travail nécessite de modifier les corpus annotés impliqués, dans la mesure où l'annotation des occurrences de *pour* n'est pas toujours cohérente avec les résultats du travail présenté ici : de nombreux arguments en *pour* sont annotés comme modifieurs. Une incorporation de nos annotations manuelles sur les *pour V-inf* et une extension de ces annotations aux autres occurrences de *pour* pourrait s'avérer un préalable nécessaire.

Références

- ABEILLÉ A., CLÉMENT L. & TOUSSENEL F. (2003). Building a treebank for French. In A. ABEILLÉ, Ed., *Treebanks*. Kluwer, Dordrecht.
- CADIOT P. (1990a). Contrôle anaphorique et prépositions. *Langages*, **97**, 8–23.
- CADIOT P. (1990b). À propos du complément circonstanciel de but. *Langue française*, **86**, 51–64.
- COLINET M., DANLOS L., DARGNAT M. & WINTERSTEIN G. (2014). Les emplois de la préposition *pour* suivie d'une infinitive : description, critères formels et annotation en corpus. In *Actes du 4ème Congrès Mondial de Linguistique Française*, Berlin, Allemagne. À paraître.
- COLLINS M. (1997). Three generative, lexicalised models for statistical parsing. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics*, p. 16–23, Madrid, Spain.
- DANLOS L., ANTOLINOS-BASSO D., BRAUD C. & ROZE C. (2012). Vers le FDTB : French Discourse Tree Bank. In *TALN 2012 : 19ème conférence sur le Traitement Automatique des Langues Naturelles*, p. 471–478, Grenoble, France.
- DUBOIS J. & DUBOIS-CHARLIER F. (1997). *Les verbes français*. Paris, France : Larousse-Bordas.
- GROSS M. (1975). *Méthodes en syntaxe : Régimes des constructions complétives*. Paris, France : Hermann.
- HUDDLESTON R. & PULLUM G. (2002). *The Cambridge Grammar of the English Language*. Cambridge : Cambridge University Press.
- MIRROSHANDEL S. A., NASR A. & SAGOT B. (2013). Enforcing Subcategorization Constraints in a Parser Using Sub-parses Recombining. In *Proceedings of NAACL 2013*, Atlanta, États-Unis.
- RIEZLER S., KING T. H., KAPLAN R. M., CROUCH R., MAXWELL III J. T. & JOHNSON M. (2002). Parsing the Wall Street Journal using a Lexical-Functional Grammar and discriminative estimation techniques. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, p. 271–278, Philadelphie, Pennsylvanie, États-Unis.
- SAGOT B. (2010). The *Lefff*, a freely available, accurate and large-coverage lexicon for French. In *Proc. of the 7th Language Resource and Evaluation Conference*, La Valette, Malte.
- SAGOT B. & DANLOS L. (2012). Merging syntactic lexica : the case for French verbs. In *LREC'12 Workshop on Merging Language Resources*, Istanbul, Turquie.
- TOLONE E. (2011). *Analyse syntaxique à l'aide des tables du Lexique-Grammaire du français*. PhD thesis, LIGM, Université Paris-Est, France, Laboratoire d'Informatique Gaspard-Monge, Université Paris-Est Marne-la-Vallée, France.
- VAN DEN EYNDE K. & MERTENS P. (2006). Valency dictionary - DICOVALENCY : user's guide. <http://bach.arts.kuleuven.be/dicovalence/>.
- VILLEMONTÉ DE LA CLERGERIE É. (2013). Improving a symbolic parser through partially supervised learning. In *Proceedings of the 13th International Conference on Parsing Technologies (IWPT)*, Naria, Japon.
- WYLER G. (2014). Manuel de la grammaire française. <http://gabrielwyler.com/accueil.html>.