

Calcul des cadres de sous-catégorisation des noms déverbaux français (le cas du génitif)

Ramadan Alfared Denis Béchet Alexandre Dikovsky

LINA, Université de Nantes, 2, rue de la Houssinière, 44000 Nantes

Ramadan.Alfared@etu.univ-nantes.fr

Denis.Bechet@univ-nantes.fr

Alexandre.Dikovsky@univ-nantes.fr

RÉSUMÉ

L'analyse syntaxique fine en dépendances nécessite la connaissance des cadres de sous-catégorisation des unités lexicales. Le cas des verbes étant bien étudié, nous nous intéressons dans cet article au cas des noms communs dérivés de verbes. Notre intérêt principal est de calculer le cadre de sous-catégorisation des noms déverbaux à partir de celui du verbe d'origine pour le français. Or, pour ce faire il faut disposer d'une liste représentative de noms déverbaux français. Pour calculer cette liste nous utilisons un algorithme simplifié de repérage des noms déverbaux, l'appliquons à un corpus et comparons la liste obtenue avec la liste Verbaction des déverbaux exprimant l'action ou l'activité du verbe. Pour les noms déverbaux ainsi obtenus et attestés ensuite par une expertise linguistique, nous analysons la provenance des groupes prépositionnels subordonnés des déverbaux dans des contextes différents en tenant compte du verbe d'origine. L'analyse est effectuée sur le corpus Paris 7 et est limitée au cas le plus fréquent du génitif, c'est-à-dire des groupes prépositionnels introduits par *de*, *des*, etc.

ABSTRACT

On Computing Subcategorization Frames of French Deverbal Nouns (Case of Genitive)

Fine dependency analysis needs exact information on the subcategorization frames of lexical units. These frames being well studied for the verbs, we are interested in this paper by the case of the noun deverbals. Our main goal is to calculate the subcategorization frame of deverbals in French from that of the source verb. However, this task needs a representative list of French deverbal nouns. To obtain such a list, we use a simplified algorithm detecting deverbal nouns in texts. The obtained list attested by linguists is compared with the existing list Verbaction of deverbals expressing the action/activity of French verbs. For these deverbal nouns, we analyse the origin of their subordinate prepositional phrases in different contexts relative to their source verbs. This analysis is carried out over the corpus Paris 7 and is limited to the most frequent cases of the genitive, i.e. to the prepositional phrases headed by the prepositions *de*, *des*, etc.

MOTS-CLÉS : nom déverbal, cadre de sous-catégorisation, groupe prépositionnel, analyse en dépendances.

KEYWORDS: Deverbal Noun, Subcategorization Frame, Prepositional Phrase, Dependency Tree.

1 Introduction

Comparons les trois phrases suivantes :

- (1) *Le service de livraison est fermé.*
- (2) *Le besoin de reconnaissance est un besoin fondamental de tous les êtres humains.*
- (3) *La livraison des commandes est assurée directement par le service de livraison du producteur.*
- (4) *On attend l'abrogation définitive de cette loi.*

Nous voyons que dans chaque phrase il y a au moins un groupe prépositionnel (GP) subordonné à un nom : e.g. **de** livraison est subordonné à *service*, **des** commandes est subordonné à *livraison*, **du** producteur est subordonné à une autre occurrence de *livraison*. Parmi ces dépendances entre les noms gouverneurs et les prépositions têtes des GP, lesquelles sont *attributives*, lesquelles sont *dépendances d'objet* ? On ne peut pas effectuer une analyse fine en dépendances sans savoir répondre à cette question. En effet, dans la grammaire de dépendances il faut au moins faire la distinction entre les dépendances obligatoires d'objet et les dépendances facultatives de modificateurs en tout genre. Cette différence entre les deux types de dépendances correspond aussi à une différence fondamentale sémantique : dans le cas de dépendances attributives la sémantique du GP définit la valeur d'un attribut de la sémantique du nom gouverneur, tandis que dans le cas de dépendances d'objet la sémantique du nom gouverneur est une fonction appliquée à la sémantique du GP. Mais comment peut-on savoir dans quel cas de figure on se trouve ?

On peut distinguer deux catégories de substantifs :

- (c_1) ceux dont au moins un GP subordonné est un complément d'objet (c'est-à-dire, correspond à un argument sémantique)¹,
- (c_2) et ceux qui n'en ont aucun.

Les compléments des substantifs de la catégorie c_1 ont le même statut (par rapport au substantif) que celui des éléments du cadre de sous-catégorisation des verbes (par rapport aux verbes). Aussi ces substantifs peuvent avoir des compléments d'objet dans tous les contextes (ils peuvent aussi avoir des GP attributifs). Parmi ces substantifs on trouve les nombres indéterminés exprimant une quantité limitée et indéterminée d'objets, tels *dizaine*, *douzaine*, *trentaine*, etc. On y trouve aussi les noms communs tels *besoin*, *peur*, *marre*, etc. qui avec un verbe léger (e.g. *avoir*) forment un phrasème et qui en héritent les compléments d'objet (au génitif). Mais surtout on y trouve de très nombreux noms déverbaux. On s'intéresse dans cet article surtout aux noms déverbaux parce qu'il existe un certain nombre de règles qui définissent leurs arguments à partir des arguments des verbes d'origine.

Les déverbaux héritent du cadre de sous-catégorisation du verbe d'origine et ils le transforment. E.g. dans les phrases (1)-(4) nous trouvons les déverbaux *reconnaissance*, *livraison*, *service*, *producteur*, *abrogation*. Comme on peut le voir dans les analyses² ci-dessous, *service* (catégorie (c_2)) a un GP attributif *de livraison*, tandis que *besoin* et *livraison* (catégorie (c_1)) ont un GP d'objet *de reconnaissance* dans la phrase (2), *des commandes* dans la phrase (3). Le déverbal *abrogation* a un GP d'objet *de cette loi* dans la phrase (4). En même temps, *besoin* peut aussi avoir un GP attributif, e.g. *de tous les êtres humains* dans la phrase (2).

1. Comme tous les arguments sémantiques, ces arguments peuvent être liés à des éléments du contexte, ce qui correspond à l'ellipse de surface mais ne les prive pas du statut d'arguments obligatoires.

2. Ces analyses sont obtenues avec l'analyseur syntaxique et en utilisant la grammaire catégorielle de dépendance du français qui font partie du système CDG Lab (Alfred *et al.*, 2011).

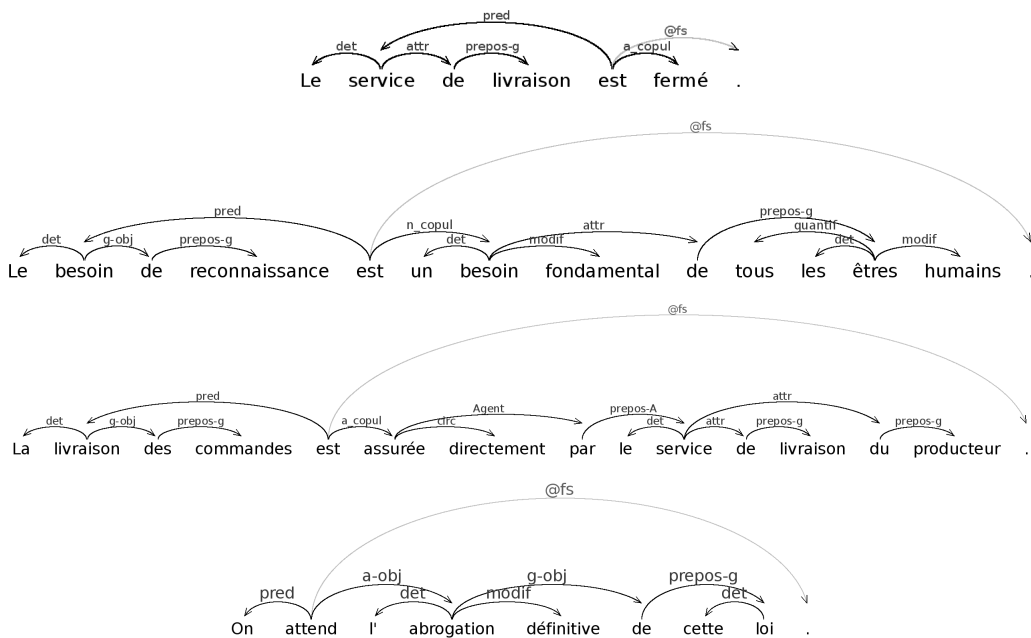


FIGURE 1 – Analyses des phrases de l'introduction

Les nombres indéterminés et les noms composant les phrasèmes ne posent pas de problème parce qu'ils sont peu nombreux. Par contre, la recherche des déverbaux n'est pas une tâche simple (surtout pour les verbes des deuxièmes et troisièmes groupes). Et les difficultés ne s'arrêtent pas là : le vrai problème est de déterminer les cadres de sous-catégorisation des déverbaux à partir des cadres de sous-catégorisation des verbes. Ce sont les deux problèmes abordés dans cet article. Par rapport au système CDG Lab mentionné ci-dessus, ces problèmes se traduisent par des algorithmes de complétion du lexique de la grammaire du français plaçant les déverbaux dans les classes lexicales qui correspondent à leurs cadres de sous-catégorisation. Cela explique notre intérêt pour cette problématique.

L'article est organisé comme suit : la section 2 présente un algorithme de repérage des noms déverbaux ainsi que son évaluation. La section suivante introduit le cadre théorique de l'article sur les liens entre les cadres de sous-catégorisation des noms déverbaux et des verbes d'origine. La section 4 présente nos expériences sur les règles de transformation des cadres (du verbe au déverbal). L'article se termine par une discussion et les perspectives de ce travail.

2 Recherche des déverbaux

En français les noms déverbaux sont formés des verbes selon le schéma :

$$\text{BASE} + \text{TERMINAISON} \Rightarrow \text{var}(\text{BASE}) + \text{SUFFIXE},$$

où var(BASE) est une variation de la base et SUFFIXE est un suffixe formateur (parfois vide : -0).

Le plus simple est d'obtenir la liste plus ou moins complète des suffixes formateurs cf. (Namer, 2009). En ce qui concerne les variations de la base, il s'avère difficile de trouver une source où elles sont toutes présentées systématiquement (en tout cas nous ne l'avons pas trouvée). Heureusement, pour les verbes du premier groupe, auquel appartient la majorité absolue des verbes français, les variations sont très rares et on peut les négliger (ou bien ne considérer que les plus simples). Cela rend possible d'appliquer aux verbes du premier groupe un algorithme simpliste (appelé ci-dessous **proto-déverb**) qui effectue la transformation :

BASE+TERMINAISON \Rightarrow BASE+SUFFIXE

selon 17 règles citées dans la table 3. Nous avons appliqué **proto-déverb** à deux corpus : un d'arbres syntaxiques : le corpus Paris 7 (Abeillé *et al.*, 2003)³, un autre textuel : le corpus GEOPO (du CLLE-ERSS(Ho-Dac, 2007)). **Proto-déverb** pose plusieurs problèmes :

Problème de faux déverbaux. L'algorithme **proto-déverb** peut construire des combinaisons BASE+SUFFIXE qui ne sont pas des déverbaux. Par exemple, pour le suffixe *-ence* : *scier* \Rightarrow *science*, pour le suffixe *-tion* : *roter* \Rightarrow *rotation*.

Problème de fausse origine. La combinaison BASE+SUFFIXE construit par **proto-déverb** peut être un déverbal, mais provenir d'une autre base. Par exemple, pour le suffixe *-ant* : *garer* \Rightarrow *garant*, tandis que *garant* est formé du verbe *garantir*.

Problème de fausse direction. Parfois le nom construit par **proto-déverb** n'est pas un déverbal, mais c'est le verbe d'origine qui est dérivé du nom. Par exemple, pour le suffixe *-euse* : *mitrailler* \Rightarrow *mitrailleuse* (en fait, *mitrailleuse* \Rightarrow *mitrailler*). Pour ce genre de problème il vaudrait mieux parler de co-déverbaux, mais cette situation est assez rare et on peut aussi la négliger.

2.1 Évaluation de l'algorithme *proto-déverb*

On voit que l'application de l'algorithme **proto-déverb** nécessite une expertise du résultat de son application. Notre démarche d'évaluation consiste ici à réaliser deux expérimentations indépendantes ainsi qu'une comparaison directe avec la liste VerbaCTION de déverbaux dénotant l'action ou l'activité exprimée par le verbe. La première expérimentation compare la liste des déverbaux produit par **proto-déverb** avec une liste produite par un expert depuis une partie des textes du corpus GEOPO(Ho-Dac, 2007). La seconde expérimentation, effectuée sur le corpus arboré de Paris 7 (Abeillé *et al.*, 2003), qui sera longuement présentée et commentée dans la seconde partie de l'article consacré aux GP des déverbaux, permet d'obtenir, en la restreignant à la notion d'être ou non un déverbal, une évaluation des déverbaux proposés par l'algorithme **proto-déverb**. Finalement, nous avons utilisé la liste des déverbaux de VerbaCTION (Hathout *et al.*, 2002; Tanguy et Hathout, 2002) développée à l'INaLF/ATILF et à l'ERSS pour la comparer avec la liste produite par **proto-déverb**.

2.2 Évaluation sur des textes du corpus GEOPO

La première expérience vérifie la pertinence de l'algorithme **proto-déverb** pour la construction des couples déverbal/verbe sur des textes du corpus GEOPO. Ce corpus est constitué d'environ 270 000 mots. Il regroupe 32 textes longs qui sont des articles expositifs (informatifs et argumentatifs) proposant des réflexions relatives à notre monde d'aujourd'hui (la crise pétrolière, la guerre contre "l'axe du mal", le terrorisme, l'explosion chinoise, la paix au moyen-orient, etc.) La méthodologie

3. Parmi les 17 règles seulement 12 de la table 2 ont été appliquées au moins une fois dans ce corpus.

utilisée pour l'évaluation consiste à repérer automatiquement les séquences constituées d'un nom commun⁴ suivi d'une des formes de la préposition *de* (*de*, *d'*, *du* ou *des*). Ensuite, les séquences repérées sont annotées à la fois par l'algorithme **proto-déverb** et par un expert. L'annotation de **proto-déverb** indique pour un nom commun s'il est un déverbal. Dans ce cas l'algorithme fournit le verbe⁵ dont il est dérivé. L'expert indique de même si les séquences repérées correspondent à un déverbal (en particulier, le mot doit être un nom commun dans la phrase). Dans ce cas, le verbe d'origine est fourni par l'expert. Dans le cas contraire, l'annotateur indique la raison pour laquelle le mot n'est pas un déverbal. Cela peut être :

1. le mot n'est pas un nom commun dans le contexte, comme par exemple *parler* dans *plutôt que de parler de centralisation, [...]*,
2. le nom fait partie d'une locution comme *point* dans *ainsi, d'un point de vue législatif [...]*,
3. le nom ne correspond pas (dans le sens de la phrase) à un déverbal comme *priorité* dans *Il s'agit au contraire d'une priorité des pouvoirs publics*,
4. l'expert annotateur n'est pas sûr que le nom soit un déverbal comme le nom *poids* pour le verbe *peser*.

L'algorithme **proto-déverb** ne pouvant pas détecter les situations 1 et 2, nous avons éliminé les séquences correspondantes des statistiques. Les cas 4 étant incertains, nous avons aussi préféré les supprimer. Nous n'avons donc gardé que les séquences du point 3 et celles qui ont été annotées par les experts comme des déverbaux.

Sur les 286 séquences de phrases traitées, 220 ont été retenues par les experts (14 de statut incertain (point 4) et 52 correspondant aux points 1 ou 2).

	Déverbal (proto-déverb)	Non déverbal (proto-déverb)
Déverbal (expert)	28 (12,7%)	63 (28,6%)
Non déverbal (expert)	4 (1,8%)	125 (56,8%)

TABLE 1 – Comparaison sur GEOPO entre **proto-déverb** et les experts annotateurs

Ces résultats suggèrent que les règles utilisées par l'algorithme **proto-déverb** donnent pour la grande majorité des occurrences de vrais déverbaux (28 contre 4) dans ce type de corpus. Par contre, la couverture des règles pourrait être améliorée. En fait, l'algorithme simpliste **proto-déverb** se contente de traiter les verbes du premier groupe. De plus, certaines règles trop peu précises n'ont pas été incluses dans cette version de l'algorithme comme la règle qui supprime le *r* final *BASE+er* ⇒ *BASE+e* et qui permet par exemple de générer le couple *contrôle/contrôler*. L'ajout de ce type de règles aurait conduit à une production d'un trop grand nombre de faux couples déverbaux/verbes qui aurait complexifié la tâche des experts sur le corpus Paris 7.

2.3 Évaluation sur le corpus arboré Paris 7

Pour compléter ces statistiques, nous pouvons aussi utiliser une partie des résultats de la seconde expérience pour juger de la qualité de l'algorithme **proto-déverb**. Lors de cette seconde

4. En fait, il s'agit des mots qui sont des noms communs dans le lexique *Lefff* (Sagot, 2010).

5. Exceptionnellement, l'algorithme peut fournir plusieurs verbes correspondant au mot. Par exemple, *parade* correspond à *parer* et à *parader*.

expérience nous avons repéré, dans le corpus d'arbres syntaxiques de Paris 7, les constructions où un nom commun provenant d'une règle de l'algorithme **proto-déverb** possède un groupe prépositionnel (GP) introduit par la préposition *de*. Sur ces éléments, nous avons demandé à des experts annotateurs de repérer les séquences qui contiennent effectivement le déverbal du verbe donné par l'algorithme **proto-déverb**. Nous verrons dans la seconde partie de l'article sur les compléments au génitif des déverbaux que, dans le cas où l'expert pense que le nom est un déverbal, il doit fournir l'éventuel lien entre le GP et un des arguments du verbe d'origine. Par exemple dans la phrase *Il ne faut pas secondariser l'enseignement supérieur ni inciter à un démembrement des universités*, le GP *des universités* du déverbal *démembrement* correspond au complément à l'accusatif (complément d'objet direct) du verbe *démembrer*. En faisant abstraction de ce résultat sur le GP pour l'instant, nous pouvons savoir si la proposition de l'algorithme **proto-déverb** correspond bien à un déverbal dans le contexte d'une phrase. Les résultats en fonction des règles de l'algorithme **proto-déverb** sont les suivants :

Règle	Déverbal (expert)	Non déverbal (expert)	Nombre de couples
er ⇒ ade	2	1	3
er ⇒ age	63	10	73
er ⇒ aison	5	0	5
er ⇒ ant	26	5	31
er ⇒ ation	274	7	281
er ⇒ ement	132	6	138
er ⇒ ence	24	7	31
er ⇒ eur	45	17	62
er ⇒ euse	1	0	1
er ⇒ oir	1	4	5
er ⇒ rice	1	1	2
er ⇒ ure	10	4	14
Total	584 (90,4%)	62 (9,6%)	646

TABLE 2 – Évaluation de **proto-déverb** sur le corpus Paris 7 en fonction des règles

Ce tableau indique pour chaque règle utilisée pour l'analyse du corpus le nombre de couples nom/verbe qui correspondent ou non à un couple déverbal/verbe pour l'expert ainsi que le total. Globalement, le résultat est assez bon mais n'est pas uniforme en fonction de la règle de passage du verbe au nom. La précision globale est de 90,4%.

2.4 Comparaison avec la liste de déverbaux Verbaaction

Une troisième expérience consiste à comparer la liste de 5537 couples déverbal/verbe produits par l'algorithme **proto-déverb** à partir du lexique *Lefff* avec les 8432 couples du lexique *Verbaaction* qui comprend une liste portant sur des noms morphologiquement apparentés au verbe et qui peuvent être utilisés pour dénoter l'action ou l'activité exprimée par le verbe. Pour cette comparaison nous avons trois possibilités suivant qu'un couple nom/verbe correspond à une ou aux deux listes. La table 3 présente les résultats pour les verbes du premier groupe qui sont produits par l'algorithme **proto-déverb**.

Règle	Déverbal (Verbaction)	Non déverbal (Verbaction)	Nombre de couples
er ⇒ ade	36	33	69
er ⇒ age	994	60	1054
er ⇒ aison	22	15	37
er ⇒ ant		312	312
er ⇒ ante	1	127	128
er ⇒ ation	918	85	1003
er ⇒ ement	778	61	839
er ⇒ ence	19	18	37
er ⇒ ette	6	187	193
er ⇒ eur	1	938	939
er ⇒ euse		538	538
er ⇒ oir		145	145
er ⇒ oire		20	20
er ⇒ rice		17	17
er ⇒ trice		5	5
er ⇒ ure	33	166	199
er ⇒ xion	1	1	2
Total	2809 (50,7%)	2728 (49,2%)	5537

TABLE 3 – Comparaisons entre **proto-déverb** et le lexique Verbaction en fonction des règles

Le nombre de couples avec un verbe du premier groupe du lexique Verbaction qui ne sont pas produits par l’algorithme est de 5623 (sur 8432). En fait, comme nous l’avons déjà vu précédemment, l’algorithme **proto-déverb** ne couvre pas la totalité des verbes du premier groupe. Par contre, il fournit des couples qui ne sont pas dans Verbaction comme les noms de suffixe *ant*, *ante*, *ette*, *eur*, etc qui ne désignent pas directement l’action ou l’activité exprimée par le verbe d’où l’intérêt de cet algorithme dans cette étude.

Pour être complet, nous avons aussi comparé les annotations de nos experts sur le corpus GEOPO avec la liste de Verbaction ce qui a donné les résultats de la table 4. L’adéquation entre nos experts et le lexique Verbaction est assez bonne (85,5%) mais il reste quelques différences comme les couples *structure/structurer*, *position/positionner*, *craintes/craindre* ou *résistant/résister* qui s’explique par le choix de Verbaction de ne s’intéresser qu’aux déverbaux dénotant l’action ou l’activité exprimée par le verbe.

En conclusion, l’algorithme **proto-déverb** est simple et assez précis. Il fournit un grand nombre de déverbaux ayant des liens variés avec le verbe (contrairement à Verbaction qui est spécialisé).

	Déverbal (Verbaction)	Non déverbal (Verbaction)
Déverbal (expert)	69 (31,4%)	22 (10%)
Non déverbal (expert)	10 (4,5%)	119 (54,1%)

TABLE 4 – Comparaison sur GEOPO entre le lexique Verbaction et les experts annotateurs

3 Cadres de sous-catégorisation des déverbaux

Passons maintenant au problème principal, celui de déterminer les cadres de sous-catégorisation des déverbaux à partir des cadres de sous-catégorisation des verbes d'origine. Avant tout il faut expliquer comment nous représentons ces cadres. Notre représentation est basée sur une affectation de cas aux compléments d'objet proposée dans (Dikovskiy, 2011). Cette affectation est faite en fonction des clitiques selon la règle suivante (simplifiée ici) :

Soit π un complément d'objet d'un verbe V dans une phrase. Alors π est au cas C si dans la phrase il peut être pronominalisé et transformé en un clitique P au cas C ancré sur V . Enfin, si π ne peut pas être pronominalisé, alors il est au cas oblique (o)⁶.

En français il n'y a que quatre cas de clitiques : *accusatif* (a), e.g. *me, la, en*, *génitif* (g) : *en*, *datif* (d), e.g. *te, lui, y*, et *locatif* (l) : *y*. Respectivement, selon cette règle chaque complément d'un verbe a un des cinq cas : a, g, d, l ou o ⁷. Aussi le cadre de sous-catégorisation du verbe est défini par les fonctions syntaxiques de ses arguments obligatoires (*actants*) : *sujet* ($subj$), *objet direct* (od), *objet indirect* (oi)), réalisés par les cas qui leur correspondent, e.g. *chanter*($subj/n$) (intransitif), ou *chanter*($subj/n, od/a$), *donner*($subj/n, od/a, oi/d$), *émerger*($subj/n, oi/g$), *parler*($subj/n, od/a$) ou *parler*($subj/n, oi1/g, oi2/d$), *placer*($subj/n, od/a, oi/l$), *échanger*($subj/n, od/a, oi/o$), etc. Parfois, pour préciser la réalisation du complément direct par une phrase dont la tête est un verbe à l'infinif, nous utilisons l'étiquette *inf* : *faire*($subj/n, od/a, oi/inf$). Le cadre s'étend sur les GP en position d'argument circonstanciel : *circ/C*, quand C est le cas du GP

L'idée principale de notre calcul des cadres de sous-catégorisation des noms déverbaux est que dans le cas où on sait de quel verbe est formé le nom déverbal, on peut déduire le cadre du second à partir de celui du premier. Il s'agit donc de définir les règles de transformation des cadres⁸ des verbes les cadres des noms dérivés. Ce calcul des cadres par réduction rend possible de contourner le problème difficile de définition des cadres de sous-catégorisation des verbes (cf. (Fillmore, 1968, 1977; Grimshaw, 1990; Dowty, 1991; Van Valin, 1997; Mel'čuk, 2004)). Les cadres de sous-catégorisation des verbes sont alors donnés a priori et sont représentés par les listes des réalisations des actants : fonction syntaxique/cas définies ci-dessus : $subj/n, od/a, oi/d, oi/g, oi/l, oi/o, oi/inf$. Les cadres des noms déverbaux sont représentés par les réalisations des objets (sauf od/a), mais aussi des attributs : $attr/C$. Quand un argument circonstanciel d'un verbe ou un arguments attributif d'un nom n'est pas réalisé par un cas nous allons noter cette réalisation par $circ/\perp$ ($attr/\perp$). Sur cette signature étendue on peut définir les transformations diathétiques de nominalisation des verbes par les séquences de règles du genre ($subj/n \rightarrow 0$)⁹ ou ($od/a \rightarrow oi/g$), ($subj/n \rightarrow attr/g$) ou ($oi/l \rightarrow oi/l$), etc.

Notre hypothèse initiale était que les transformations diathétiques de nominalisation en français sont définies par quatre règles simples :

- le sujet du verbe disparaît et est converti en un attribut du nom dérivé : ($subj/n \rightarrow attr/C$),
- son objet direct ($C = a$) devient l'objet indirect au génitif ($od/a \rightarrow oi/g$),
- tous les autres objets indirects sont tout simplement hérités sous réserve de leur présence dans le cadre (e.g. ($oi/d \rightarrow oi/d$) ou ($oi/o \rightarrow oi/o$),

6. Certes, il faut définir plus précisément cette règle pour les clitiques ambigus : y au datif vs. y au locatif, *en* au génitif vs. *en* à l'accusatif, etc. et la compléter par le cas n (*nominatif*) du sujet.

7. Cf. (van den Eynde et Mertens, 2003)) où on tient compte de tous les pronoms.

8. Diathetic shift rules (ang.).

9. $A \rightarrow 0$ veut dire que A est supprimé.

- les GP circonstanciels deviennent GP attributifs ($circ/C_1 \rightarrow attr/C_2$).

Il s'avère que le tableau réel est bien plus complexe. Pour commencer, certains arguments des déverbaux hérités des cadres des verbes d'origine ont tendance à perdre leur statut obligatoire. Il s'agit surtout des compléments locatifs et obliques. Par exemple, le cadre du verbe pronominal *se déplacer* est soit (oi/l) soit (oi/o) (un des deux est obligatoire : cf. *Il se déplace à Bordeaux* ; *Il se déplace en voiture* mais **Il se déplace*). Encore pire est la situation où le verbe d'origine est di-transitif avec le cadre ($subj/n, od/a, oi/g$), (e.g. *saupoudrer(subj/n, od/a, oi/g)*). Dans ce cas un des compléments est supprimé. En fonction du contexte le complément d'objet au génitif du déverbal *saupoudrage(oi/g)* peut provenir soit du complément d'objet direct du verbe (cf. *saupoudrage des champs (subj/n \rightarrow 0, od/a \rightarrow oi/g, oi/g \rightarrow 0)*), soit de son complément d'objet indirect au génitif (cf. *saupoudrage de réformettes (subj/n \rightarrow 0, od/a \rightarrow 0, oi/g \rightarrow oi/g)*). Et enfin, très souvent le GP subordonné à un nom déverbal peut provenir du sujet et même d'un des arguments circonstanciels du verbe d'origine. Par exemple, dans la phrase *L'argumentation du clan pro-inflation consiste à ...* le GP *du clan pro-inflation* est un attribut du nom déverbal *argumentation(0)*. Cet attribut provient du sujet du verbe *argumenter(subj/n, od/a)*. Ainsi dans ce contexte le cadre de sous-catégorisation du verbe est transformé selon la règle ($subj/n \rightarrow attr/g, od/a \rightarrow 0$). D'un autre côté, dans la phrase *La débandade de 1982-1983 qui vit le pays...* le verbe *débander(subj/n)* est intransitif. Aussi l'attribut *de 1982-1983* du déverbal *débandade* provient d'un argument circonstanciel du verbe (*débander en 1982-1983*) : ($circ/o \rightarrow attr/g$).

Sans argument supplémentaire ces exemples peuvent faire croire qu'il faut faire face à un nombre important de transformations de cadres de sous-catégorisation. Heureusement, les actants des noms déverbaux obéissent au postulat suivant (visiblement remontant à la présomption d'invariabilité des actants des mots de (Mel'čuk et Holodovič, 1970))¹⁰ :

postulat de provenance : les actants des noms déverbaux proviennent d'un des actants du verbe d'origine.

Une conséquence directe de ce postulat est que l'actant du déverbal ne peut pas provenir d'un argument circonstanciel du verbe d'origine mais seulement de son sujet ou d'un des objets présent dans son cadre. Dû à ce postulat notre objectif devient plus précis : on peut établir la provenance des actants des noms déverbaux réalisés par des GN non attributifs seulement à partir des GN non circonstanciels du verbe d'origine. Quoique très restrictif, le postulat de provenance laisse beaucoup de variantes : différents suffixes formateurs, différents rôles syntaxiques des noms déverbaux dans différents contextes et différents actants correspondants du verbe d'origine. Pour simplifier la tâche nous nous limitons dans cet article aux GN au génitif qui réalisent les actants des noms déverbaux. Dans les expériences présentées dans la section suivante nous étudions leur provenance et les règles de conversion qui les concernent.

4 Expériences et évaluation des résultats

Ici, nous nous intéressons à découvrir le rôle des GP des déverbaux pour le verbe qui leur correspond. Nous avons identifié dans un premier temps plusieurs possibilités puis nous avons cherché à comprendre sur des exemples d'un corpus les liens entre les compléments du verbe et les GP introduit par la préposition *de* du déverbal. Notre étude a consisté sur le corpus d'arbres

10. Nos expériences avec le corpus Paris 7 confirment ce postulat (au moins pour les noms déverbaux repérés dans le corpus).

syntaxiques de phrases de Paris 7 (Abeillé *et al.*, 2003), à repérer toutes les occurrences de noms communs qui correspondent à l'algorithme **proto-déverb** et qui possèdent un (ou plusieurs) GP introduit par la préposition *de*. Sur ces occurrences, nous avons demandé à des experts d'indiquer pour ces occurrences si les couples nom/verbe produit par l'algorithme **proto-déverb** correspondent effectivement (dans le contexte des occurrences) à un couple déverbal/verbe. Dans un deuxième temps, pour les occurrences validées comme des déverbaux, les experts devaient indiquer la fonction que pouvaient prendre les GP par rapport au verbe. Voici quelques exemples qui illustrent les différents cas de figure rencontrés par les experts annotateurs :

1. *Le langage de la formation, trop souvent fait d'approximations, [...] avec le couple langage/langer* : en fait, *langage* n'a rien à voir avec *langer*.
2. *L'alliance des vrais-faux contraires suscite la parade des supposés amis avec le couple parade/parer* : le nom *parade* peut être associé à deux verbes suivant son sens. Soit *parer*, soit *parader*. Dans cette phrase, le sens de *parade* correspond à *parader*. Ces deux premiers exemples sont étiquetés **non D** (non déverbal).
3. *Avec une croissance de 3,3% à la même date, les départements d'outre-mer bénéficieront d'un rattrapage d'un point supplémentaire avec le couple rattrapage/rattraper* : ici nous avons clairement un déverbal. Le GP *d'un point supplémentaire* joue le rôle du complément à l'accusatif (complément d'objet direct) du verbe *rattraper* car nous pourrions dire *ceci rattrape un point supplémentaire*. Cet exemple est étiqueté **C=a** par les experts.
4. *Reports ou renvois qui n'empêchent pas M. Bérégofoy de juger "injuste" l'accusation d'immobilisme avec le couple accusation/accuser* : le GP du déverbal correspond à un complément au génitif du verbe *accuser* (complément d'objet indirect introduit par *de*) car on peut dire *on l'accuse d'immobilisme*. Cet exemple est étiqueté **C=g** par les experts.
5. *Dans l'intervention humanitaire, il n'y a pas d'obligation de résultats avec le couple obligation/obliger* : cette fois le GP correspond à un complément au datif (complément d'objet indirect introduit par *à*) car nous pouvons dire *cela nous oblige à des résultats*. Cet exemple est étiqueté **C=d** par les experts.
6. *De plus, la Russie, mastodonte pétrolier empêtré dans des problèmes infinis, paraît incapable de stopper la dégringolade de sa production, [...] avec le couple dégringolade/dégringoler* : ici, la *production* pourrait être le sujet de *dégringoler* car nous pouvons dire *la production dégringole*. Cet exemple est étiqueté **subj** par les experts.
7. *Au plan agricole, la délégation de Gdansk a pris contact avec les professionnels de la race bovine limousine, dont les performances zootechniques, notamment en élevage extensif de plein air, leur ont semblé particulièrement adaptées aux conditions polonaises avec le couple élevage/élever* : cette fois, le GP correspond à un argument circonstanciel (optionnel) du verbe *élever* puisque nous pourrions dire *on les élève en plein air*. Cet exemple est étiqueté **circ** par les experts.
8. *Début 1991, autour de Joachim Trautwein, les six "élus" de l'ancien encadrement suivent trois séminaires de deux jours, en présence d'un sociologue est-allemand et d'un chercheur français du CNRS avec le couple chercheur/chercher* : dans cette phrase, il est difficile de rattacher le CNRS au verbe *chercher* car le GP désigne l'appartenance à une organisation de la personne qui cherche. Nous pouvons dire dans ce cas que bien que nous ayons un déverbal, le GP de ce déverbal ne provient pas de la structure des compléments du verbe d'origine. Cet exemple est étiqueté **V/S** (un déverbal dont le GP ne peut pas être subordonné au verbe : (0→attr/g)).

Les annotations ont été classées par la règle de l'algorithme **proto-déverb** utilisée pour faire la correspondance entre le déverbal et le verbe. Les séquences ont été extraites d'un sous ensemble du corpus (Abeillé *et al.*, 2003) constitué de 31 fichiers sur 45 (les autres fichiers nous permettant de vérifier nos hypothèses) et comportant 15106 phrases (sur 21776). Le programme d'extraction a identifié 647 séquences *nom + GP* avec *nom*, un nom commun produit par l'algorithme **proto-déverb** (le nom peut être séparé de son GP par d'autres éléments du groupe nominal). Sur ces 647 séquences les experts ont supprimé une phrase pour laquelle le lien entre le GP et le verbe n'était pas clair¹¹. Les experts ayant divergé sur l'interprétation de la fonction de certains GP entre **C=a** et les autres cas, nous avons dans un premier temps conservé l'annotation qui n'était pas **C=a**. Toutefois, nous indiquons, dans la colonne **C=a**, avec le symbol *+* suivi d'un nombre, le nombre de cas **C=a** divergents. En fait, sur les 646 séquences, 18 annotations sont restées différentes même après concertation des annotateurs. Comme nous l'avons vu précédemment, les experts ont repéré 62 séquences ne correspondant pas à un déverbal et 584 correspondant à un déverbal. Voici la table des résultats classés par règle :

Règle	non D	C=a	C=d	C=g	subj	circ	V/S	Total D	Total
er ⇒ ade	1				1	1		2	3
er ⇒ age	10	44 + 3		4	11	4		63	73
er ⇒ aison		5						5	5
er ⇒ ant	5	17 + 2	1		2	5	1	26	31
er ⇒ ation	7	210 + 8	3	3	39	17	2	274	281
er ⇒ ement	6	101 + 4	3	2	11	12	3	132	138
er ⇒ ence	7	5	1		14	4		24	31
er ⇒ eur	17	24 + 1		1	2	10	8	45	62
er ⇒ euse						1		1	1
er ⇒ oir	4					1		1	5
er ⇒ rice	1	1						1	2
er ⇒ ure	4	6	1			2	1	10	14
Total	62	413 + 18	9	10	80	57	15	584	646

TABLE 5 – Analyse sur le corpus Paris 7 de la fonction des GP des déverbaux

Nous voyons que très majoritairement (413 sur 584 soit 70,7%), le GP correspond au complément d'objet direct du verbe. Les deux autres types de compléments d'objet indirect (C=d et C=g) sont bien moins nombreux car les verbes du premier groupe avec ce type de complément sont peu nombreux. Les annotateurs en ont repéré 19 soit 3,3%.

À la lecture des exemples annotés C=d, on s'aperçoit que les déverbaux autorisent de manière générale qu'un GP normalement introduit par la préposition *à* soit introduit par la préposition *de* avec le déverbal. Ainsi, on peut dire *Dans l'intervention humanitaire, il n'y a pas d'obligation de résultats* à la place de *Dans l'intervention humanitaire, il n'y a pas d'obligation à des résultats*.

Par contre, les compléments du verbe au génitif (C=g) se traduisent naturellement en un GP du déverbal introduit par *de*. Ainsi, dans la phrase *reports ou renvois qui n'empêchent pas M. Bérégovoy*

11. "Le déplacement en milieu psychiatrique du traitement de l'acte - qui ne serait parlé que là, médicalement - exclut l'inculpé de la société : non seulement il est dispensé, mais il est empêché de répondre de son acte devant les parties directement concernées alors que les deux démarches devraient être complémentaires" avec le couple *déplacement/déplacer*.

de juger injuste l'accusation d'immobilisme, la séquence *accusation d'immobilisme* correspond à *accuser d'immobilisme*.

Le cas des arguments circonstanciels est aussi assez simple puisque les 57 exemples montrent que ces compléments du verbe introduit par des prépositions comme *en*, *pendant*, *vers*, *pour*, *dans*, etc, peuvent se transformer en GP introduit par *de* d'un déverbal. Dans les 57 cas rencontrés par les annotateurs, le GP était toujours optionnel et représentait un attribut plutôt qu'un actant du déverbal.

Le cas de la transformation du sujet du verbe en un GP introduit par *de* du déverbal est plus problématique. Les annotateurs en ont repéré un grand nombre (80 sur 584 soit 13,7%). Toutefois, ils ne sont pas toujours tombés d'accord sur ces séquences car une partie (18 sur 80) a été annotée comme **C=a** par au moins un annotateur. En fait, cela s'explique assez bien car la majorité de ces cas problématiques correspondent soit à des verbes pouvant être pronominaux, soit à des verbes dont le complément d'objet direct peut devenir sujet. Par exemple, nous pouvons dire que *les achats de biens durables redémarrent* mais aussi qu'*on redémarre les achats de biens durables*. Dans ce cas, le GP du déverbal *redémarrage* peut effectivement correspondre au sujet ou à l'objet de *redémarrer*. Dans le même ordre d'idée, si l'on parle de *la précarisation du marché du travail*, on peut considérer soit que *le marché du travail précarise les travailleurs*, soit que *la société précarise le marché du travail*.

Mis-à-part ces problèmes d'ambiguïté, une analyse fine des exemples annotés **subj** rejoint et complète l'analyse que nous avons faite à propos du cas **C=d**. Alors que la forme normale de la transformation du sujet dans le déverbal devrait être un GP introduit par *par*, il semble courant que cette préposition *par* soit remplacée par la préposition *de*. Cela est particulièrement vrai lorsque le verbe est intransitif comme *la dégringolade de sa production* dans le phrase *De plus, la Russie [...] paraît incapable de stopper la dégringolade de sa production [...]*. Dans ce cas très précis, il est même impossible d'utiliser un GP introduit par *par*. Pour les verbes transitifs dont un des compléments est introduit par *de*, les exemples du corpus montrent que le nom déverbal provenant du verbe perd son actant-sujet.

La conclusion de cette étude semble bien indiquer que notre hypothèse d'origine sur les transformations des cadres de sous-catégorisation des verbes vers ceux des déverbaux était globalement fondée. Nous avons toutefois dû revoir le cas du sujet des verbes qui peuvent sous certaines conditions devenir de vrais arguments du déverbal sous la forme d'un GP introduit par *de* et ajouter le cas de la transformation des compléments au datif en un GP au génitif :

- soit le sujet du verbe disparaît et est converti en un attribut optionnel du nom dérivé (en général un GP introduit par la préposition *par*) : (*subj/n*→*attr/⊥*), soit il est le complément unique du déverbal (en général un GP introduit par la préposition *de*), en particulier si le verbe est intransitif ou bien lorsque le verbe comporte normalement un complément qui est alors élidé : (*subj/n*→*oi/g*),
 - l'objet direct (*C = a*), s'il n'est pas élidé, devient un GP introduit par *de* (*od/a*→*oi/g*),
 - tout autre objet indirect, s'il n'est pas élidé, est tout simplement hérité sous réserve de sa présence dans le cadre (e.g *oi/d*→*oi/d*) ou (*oi/o*→*oi/o*) avec la possibilité pour le datif de se retrouver transformé en un GP introduit par la préposition *de* à la place de la préposition à : (*oi/d*→*oi/g*),
 - les GP circonstanciels au génitif sont transformés en GP attributifs au génitif.
- Cette analyse complète celle de (Benveniste, 1966) qui ne prévoit que deux alternatives : génitif subjectif / génitif objectif.

Pour terminer ce tableau, nous nous sommes aperçus qu'il y a très peu de cas de déverbaux comportant deux GP introduits par *de* qui correspondent à deux compléments du verbe (contrairement au cas où l'un des deux GP introduit par *de* correspond à un argument circonstanciel du verbe). Un exemple intéressant provient du déverbal *radiation* et de son verbe d'origine *radier* qui comporte deux compléments, un complément d'objet direct $C = a$ et un complément d'objet indirect introduit par *de* $C = g$. *Et semble pencher, en privé, pour une radiation de son groupe du second marché boursier* : les deux GP *de son groupe* et *du second marché boursier* correspondent à *on a radié le groupe du second marché*.

Par contre, nous n'avons trouvé aucun cas de déverbal comportant un GP au génitif provenant du sujet du verbe et un autre GP du déverbal (pas forcément au génitif) provenant d'un complément de ce verbe. Il semble donc que les cadres de sous-catégorisation des déverbaux avec deux actants dont un provient du sujet du verbe transitif, s'ils existent, sont très restreints.

5 Discussion et perspectives

Notre calcul de la provenance des compléments au génitif des noms déverbaux ne donne pas encore une solution complète et définitive au problème du calcul des cadres de sous-catégorisation des déverbaux même s'il fournit une approximation tout-à-fait satisfaisante. En effet :

- 1) dans les contextes où le nom déverbal a deux compléments d'objet, nous n'avons pas de règle contextuelle définissant la provenance de chacun (nos règles sont individuelles) ;
- 2) et même dans le cas le plus fréquent où le nom déverbal n'a qu'un seul actant, nos règles ne résolvent pas le choix précis de l'actant du verbe d'origine dont il provient.

Pour arriver à ces règles il faut raffiner notre analyse préliminaire en considérant les noms déverbaux dans les *contextes d'actant*. Il s'agit des contextes où le nom déverbal $N[V_0]$ provenant du verbe V_0 et gouvernant son complément $GP(\mathbf{de})$ est lui-même subordonné à un verbe V_g , c'est-à-dire :

$$V_g \xrightarrow{C-obj} N[V_0] \longrightarrow GP(\mathbf{de}).$$

Par exemple dans la phrase *Ce prince de la pensée vêtu en clochard ne semblait pas voir l'admiration [...] de ses élèves*, le déverbal $N[V_0]=admiration$ du verbe $V_0=admirer$ comporte un GP $GP(\mathbf{de})=de\ ses\ élèves$ et est subordonné au verbe $V_g=voir$. Dans cet exemple, c'est le verbe *voir* qui fournit le complément d'objet direct de *admirer*. Aussi ce complément sort du cadre de sous-catégorisation du déverbal *admiration* tandis que le GP *de ses élèves* qui est le sujet du verbe d'origine *admirer* devient son attribut au génitif. Il s'agit donc de la transformation : $((subj/n \rightarrow attr/g); (od/a \rightarrow 0))$.

E. Paducheva, dans son travail classique (Paducheva, 1977), décrit pour le russe le calcul complet des réductions des choix de l'actant de V_0 dont provient $GN(\mathbf{de})$ en fonction des actants différents de C dans le cadre de sous-catégorisation du verbe gouverneur V_g (c'est-à-dire ceux qui ne sont pas remplis par le déverbal). Pour élaborer les règles similaires pour les noms déverbaux français il nous faudra plus de données et une expertise linguistique additionnelle très fine qui devra regarder au-delà des simples groupes nominaux et de leur GP.

Références

- ABEILLÉ, A., CLÉMENT, L. et TOUSSENEL, F. (2003). Building a treebank for French. In ABEILLÉ, A., éditeur : *Treebanks : Building and Using Parsed Corpora*, pages 165–188. Kluwer, Dordrecht.
- ALFARÉD, R., BÉCHET, D. et DIKOVSKY, A. (2011). CDG Lab : a toolbox for dependency grammars and dependency treebanks development. In GERDES, K., HAJICOVA, E. et WANNER, L., éditeurs : *Proc. of the 1st Intern. Conf. on Dependency Linguistics (Depling 2011)*, Barcelona, Spain. <http://depling.org/proceedingsDepling2011/>.
- BENVENISTE, E. (1966). *Problèmes de linguistique générale*. Gallimard, Paris.
- DIKOVSKY, A. (2011). Categorical dependency grammars : from theory to large scale grammars. In GERDES, K., HAJICOVA, E. et WANNER, L., éditeurs : *Proc. of the 1st Intern. Conf. on Dependency Linguistics (Depling 2011)*, Barcelona, Spain. <http://depling.org/proceedingsDepling2011/>.
- DOWTY, D. (1991). Thematic proto-roles and argument selection. *Language*, 67(3):547–619.
- FILLMORE, C. J. (1968). The case for case. In BACH, E. et HARMS, R. T., éditeurs : *Universals in Linguistic Theory*, pages 1–88. Holt, Rinehart and Winston.
- FILLMORE, C. J. (1977). The case for case reopened. In COLE, P. et SADOCK, J. M., éditeurs : *Syntax and Semantics 8 : Grammatical Relations*, pages 59–81. New York : Academic Press.
- GRIMSHAW, J. (1990). *Argument Structure*. MIT Press, Cambridge, Mass.
- HATHOUT, N., NAMER, F. et DAL, G. (2002). An Experimental Constructional Database : The MorTAL Project. In BOUCHER, P., éditeur : *Many Morphologies*, pages 178–209. Cascadilla, Somerville, Mass.
- HO-DAC, L.-M. (2007). *La position initiale dans l'organisation du discours, une exploration en corpus*. Thèse de doctorat, Université Toulouse-le Mirail, France.
- MEL'ČUK, I. et HOLODOVIČ, A. (1970). To the theory of grammatical voice : (definition, calculus). In *Problemy lingvisticheskoj tipologii i struktury jazyka*. Nauka, Leningrad. (Rus.).
- MEL'ČUK, I. A. (2004). Actants in semantics and syntax 1,2. *Linguistics*, 42(1,2).
- NAMER, F. (2009). *Morphologie, lexique et traitement automatique des langues*. Hermes Science, Lavoisier.
- PADUCHEVA, E. (1977). O proizvodnyh diaitezah otpredikatnyh imen v russkom jazyke. In *Problemy lingvisticheskoj tipologii i struktury jazyka*. Nauka, Leningrad. (Rus.).
- SAGOT, B. (2010). The lefff, a freely available and large-coverage morphological and syntactic lexicon for French. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*.
- TANGUY, L. et HATHOUT, N. (2002). Webaffix : un outil d'acquisition morphologique dérivationnelle à partir du web. In PIERREL, J.-M., éditeur : *Actes de la 9^e Conférence Annuelle sur le Traitement Automatique des Langues Naturelles (TALN-2002)*, pages 245–254. Nancy. ATALA.
- van den EYNDE, K. et MERTENS, P. (2003). La valence : l'approche pronominale et son application au lexique verbal. *Journal of French Language Studies*, 13:63–104.
- VAN VALIN, Jr., R. D. (1997). Generalized semantic roles and the syntax-semantics interface. In CORBLIN, F., DOBROVIE-SORIN, C. et MARANDIN, J.-M., éditeurs : *Empirical Issues in Formal Syntax and Semantics 2 : Selected papers from the Colloque de Syntaxe et Semantique à Paris*, pages 373–388. Peter Lang.