

Étude du fonctionnement des nominalisations déverbiales dans un contexte de déspecialisation

Condamines, Anne¹ & Picton, Aurélie²

¹CLLE-ERSS, UMR 5263 (CNRS et U. Toulouse 2 Le Mirail)

²TIM, Faculté de Traduction et d'Interprétation, Université de Genève
anne.condamines@univ-tlse2.fr & aurelie.picton@unige.ch

1 Introduction

Dans cet article, nous nous intéressons au fonctionnement des nominalisations déverbiales dans un corpus spécialisé (textes scientifiques et techniques) par rapport à un corpus général (articles de la presse quotidienne). Le point de vue global que nous adoptons est celui de la « déterminologisation » et, plus précisément, de ce que nous appellerons ici la « déspecialisation » de la connaissance. Ce terme renvoie au phénomène de transfert de la connaissance spécialisée vers le grand public, mais de manière moins directe que dans la vulgarisation, ce qui se traduit par des manifestations linguistiques moins évidentes, que nous voulons mettre au jour. La déspecialisation ne concerne donc qu'en partie les phénomènes de vulgarisation qui eux, s'inscrivent dans une volonté presque didactique de mettre à portée du grand public une connaissance spécialisée. Dans ce contexte, nous cherchons à baliser la question de la déspecialisation à travers une approche de linguistique de corpus. Nous proposons la mise en place de jalons méthodologiques pour observer ce phénomène complexe et présentons des résultats préliminaires de nos observations, entre données quantitatives automatisées et analyses qualitatives. Pour ce faire, nous nous focalisons dans cet article sur les nominalisations déverbiales, dans la mesure où il est connu qu'elles apparaissent en plus grand nombre dans les corpus spécialisées que dans les corpus généraux (Kocourek, 1991). La dimension quantitative (comparée en corpus spécialisé et général) est la première que nous examinons ; du point de vue qualitatif, nous présentons les premiers résultats d'une analyse fine de certaines des nominalisations.

Dans un premier temps, nous précisons le positionnement de notre étude. Dans un second temps, nous présentons les corpus et la méthode (outils et ressources) que nous avons mise en place. En effet, nous pratiquons une linguistique de corpus outillée qui fait appel à différents types d'outils dont nous intégrons les résultats dans la progression de notre réflexion. Cette dimension méthodologique joue un rôle important dans cet article qui aborde un phénomène (la déspecialisation) qui a été très peu étudié à partir de corpus. Enfin, la dernière partie présente les résultats – quantitatifs et qualitatifs – que nous avons obtenus.

2 Problématique

Le phénomène de terminologisation, entendu comme le passage d'une partie du vocabulaire général vers un domaine spécialisé où il prend un sens adapté (restreint le plus souvent) a été relativement bien étudié (Calberg-Challot, 2007 ; Unguraenu, 2003 ; Sablayrolles, 2000). En revanche, la déterminologisation (terme proposé par Meyer et Macintosh, 1999), qui correspond au passage d'une partie de la terminologie d'un domaine spécialisé vers la langue générale a été beaucoup moins problématisée et décrite (voir cependant Dury, 2007 ; Nicolae et Gaudin, 2009). On retrouve la mention de ce phénomène chez différents auteurs où il prend des dénominations variées : « banalisation lexicale » chez Guilbert (1975) puis Galisson (1978), « déspecialisation » chez Gouadec (1990), « déterminologisation », « migration », ou même « dilution » (qui ne concerne que l'aspect sémantique) chez Meyer et Macintosh, (1999) ou encore « dédomanialisation » (Rastier et Valette, 2009). Il n'est pas certain que ces dénominations soient totalement équivalentes mais on retrouve dans toutes ces études l'idée que, lorsque la connaissance spécialisée « passe » dans le domaine général, ce phénomène s'accompagne de manifestations linguistiques particulières (fréquence, distribution différentes, etc.). C'est précisément

dans la mise au jour de ces traces de déspecialisation et de leur interprétation sur le fonctionnement des termes passés dans le langage courant que s'inscrit notre recherche.

Parallèlement, on a souvent caractérisé le fonctionnement des termes comme relevant d'un mode « déviant » (i.e. Kittredge et Lehrberger, 1983) ; par exemple:

« In describing SL [Sublanguage] and GL [General Language] one should make use of three complementary points-of-view or modes.
the restrictive mode: by excluding certain features of GL, SL can be described as a restricted form of language;
the deviant mode: SL can show specific features which are not found in GL and therefore can be considered a deviant form of GL;
the preferential mode: this approach of SL phenomena is complementary to the restrictive and deviant modes, and is expressed in terms of preferences. » (Mac Naught et al., 1991 : 3).

Dans (Condamines, 2003), nous avons proposé une première catégorisation des phénomènes « déviants » qui peuvent être repérés dans les corpus spécialisés, « déviant » devant être entendu comme « ne correspondant pas au fonctionnement attendu, i.e., le plus couramment rencontré par un locuteur/auditeur 'moyen' ». En voici une synthèse :

- mots ou groupes de mots inconnus,
- fréquence anormale,
- structures elliptiques, qui concernent
 - o soit un « mot vide » : déterminant ou préposition : *alarme système, service support, décommutation télémessure* ;
 - o soit un mot plein par exemple l'objet d'un verbe comme dans l'énoncé issu du domaine bancaire : *vous pouvez déposer librement sur votre compte*
- combinaisons « anormales » de mots ou groupes de mots,
- apparition d'un nouvel argument (anomalie syntaxique),
- anomalie sémantique d'un argument,
- coordination inattendue.

Une des questions qui sous-tend notre étude générale de la déterminologisation est celle de savoir si le passage d'une terminologie vers la langue « générale » s'accompagne d'une « normalisation » de la forme et de l'usage des termes, c'est-à-dire du rétablissement d'un fonctionnement « attendu ».

Afin d'apporter une première réponse à cette question, nous nous intéressons particulièrement dans cette contribution au cas des nominalisations déverbales dont on sait qu'elles sont significativement plus présentes en corpus spécialisé (voir ci-dessous). À partir de l'étude de deux corpus, l'un spécialisé (écrit par des scientifiques pour des scientifiques), l'autre général (articles de presse), nous comparons le fonctionnement des nominalisations déverbales qui, sur la base de critères statistiques, peuvent être considérées comme des termes du domaine spatial.

3 Présentation de l'étude menée autour des nominalisations déverbales

L'origine de notre étude se situe dans une demande du Centre National d'Études Spatiales (désormais Cnes) qui, à l'occasion de son cinquantième anniversaire, voulait savoir si et comment le domaine du spatial « perflusait » le domaine général. Nous avons réinterprété cette demande en une interrogation sur le phénomène de la déterminologisation (Condamines et Picton, 2011). Pour mener cette réflexion, nous avons entamé la constitution d'un « observatoire » afin de tenter de saisir la dynamique des termes et de leur diffusion dans la langue générale (Condamines et Picton, à paraître). Cet « observatoire » est basé sur trois types de corpus : un corpus scientifique fourni par le Cnes, un corpus de communiqués de presse rédigés par le service « communication » du Cnes et un corpus constitué d'articles extraits de quotidiens

francophones. Pour la présente étude, nous n'utilisons que deux des corpus : le corpus scientifique et le corpus de presse. Ce choix nous permet, d'une part, de mieux cibler les fonctionnements des nominalisations déverbales entre langue de spécialité et langue générale dans de grands corpus et, d'autre part, de limiter le nombre de données à traiter. En effet, notre choix de travailler à partir de différents corpus d'assez grande taille, et à travers – à plus long terme – une dimension diachronique constitue un enjeu de taille d'un point de vue méthodologique, qui nous amène à chercher à équilibrer approches quantitatives et qualitatives. Les corpus ainsi que les outils que nous avons utilisés dans notre démarche sont présentés dans les sections suivantes.

3.1 Corpus

Le premier corpus sélectionné est un corpus de textes scientifiques du Cnes (désormais CScience). Il est composé de deux séries de documents :

1. des chapitres d'optique et d'optoélectronique spatiale du cours de Techniques et Technologie des Véhicules Spatiaux (désormais TTVS), édité tous les 4 ans depuis 1994 par le Cnes aux Éditions Cepadués (Cnes, 1994, 1998, 2002). Ce cours est rédigé par plus de 80 experts du Cnes, à l'attention de semi-experts (Bowker et Pearson, 2002) et contient une dizaine de chapitres qui englobent l'ensemble des domaines de compétence du Cnes ;
2. des rapports de spécification des première et troisième générations de balises Doris. Doris (Détermination d'Orbite et Radiopositionnement Intégrés par Satellite) est un système de positionnement de satellites par balises terrestres dont la première génération a été conçue et développée à la fin des années 80 et la troisième génération au début des années 2000. Ces textes sont donc rédigés par des experts du domaine pour des experts de ce même domaine.

Le second corpus que nous observons est constitué d'un ensemble d'articles issus de la presse quotidienne, du 1^{er} janvier 1998 au 31 décembre 2011, regroupés par année (désormais CPresse). La presse est reconnue comme le moyen privilégié de mise à disposition de la connaissance spécialisée auprès du grand public (Jacobi, 1986 ; Moirand, 2007 ; Roqueplo, 1974). Elle semble donc aussi un lieu privilégié pour examiner les phénomènes de déspecialisation.

Ces articles ont été collectés *via* la plateforme LexisNexis®ⁱⁱ qui propose des abonnements à l'ensemble de la presse nationale et internationale en version électronique. Pour cette étude, nous avons sélectionné les quotidiens nationaux français disponibles intégralement sur la plateforme entre 1998 et 2011, soit les 4 quotidiens Le Monde, Le Figaro, Les Échos et la Croix. Nous avons extrait les 20 premiers articles retournés par mois par la plateforme pour chaque journal, soit 3360 articles, qui contiennent au moins un terme du spatial. Les termes du spatial utilisés pour la sélection de ces articles ont été extraits automatiquement à l'aide du logiciel TermoStat (Drouin, 2003), à partir du corpus scientifique CScience décrit *supra*. Ce filtrage des textes à l'aide des termes permet de ne pas regrouper seulement les articles des rubriques spécialisées (par ex. sciences, astronomie, etc.), mais de rassembler tous les articles qui contiennent *a priori* au moins un terme du domaine spatial. En effet, les nominalisations que nous ciblons (qui sont aussi, nous le verrons, des termes du spatial) peuvent apparaître dans n'importe quelle rubrique (Une, politique, histoire, etc.); or, après de premières observations, nous nous sommes aperçues que le découpage en rubriques est assez peu prévisible et est très instable d'une édition à une autre ou d'un quotidien à l'autre.

	Corpus scientifique	Corpus de presse
Nombre d'occurrences	271 010	6 660 853

Tableau 1 : Nombre d'occurrences pour chaque corpus

3.2 Outils et ressources utilisés

Notre méthode d'analyse, quelle que soit les études engagées (Condamines *et al.*, 2012 ; Picton, 2009) est basée sur une linguistique de corpus outillée, qui met en œuvre autant que faire se peut des outils et des ressources existants. En effet, d'une part, et particulièrement dans le contexte de recherche du Cnes, nous sommes amenées à manipuler des données de taille conséquente et, d'autre part, nous cherchons à préparer et assister le travail d'analyse linguistique proprement dit grâce à des outils. Cette approche demande une réflexion sur l'articulation d'approches quantitatives/statistiques et d'approches qualitatives manuelles. Ainsi, pour accompagner notre démarche méthodologique, nous nous sommes appuyées sur différents outils et ressources existants.

Dans un premier temps, les corpus ont été lemmatisés et étiquetés avec TreeTagger (Schmid, 1995). Dans un second temps, afin de repérer les nominalisations, nous avons utilisé la ressource Verbactionⁱⁱⁱ qui recense une liste de près de 10000 couples verbe/nominalisation (Hathout *et al.*, 2002) et qui est disponible sur le site de CLLE-ERSS. Cette liste nous a permis d'annoter automatiquement les nominalisations dans les corpus.

Dans la mesure où nous nous intéressons spécifiquement ici aux termes du domaine spatial, nous avons extrait et validé des termes à l'aide de TermoStat (Drouin, 2003). Cette extraction de termes vise un double objectif : premièrement, elle nous a permis d'orienter la constitution des corpus de presse, puisque les articles retenus pour ce corpus doivent contenir au moins un terme de la liste extraite avec TermoStat pour être sélectionnés (cf. *supra*) ; deuxièmement, cette extraction nous permet de cibler les « nominalisations-termes » pertinentes à observer, évitant ainsi de choisir des unités à observer *a priori*, sur la base de l'intuition seule.

Enfin, la description et l'analyse qualitative des données sont assistées avec le concordancier AntConc (Anthony, 2005), qui supporte également les textes étiquetés et annotés.

4 Etude des nominalisations déverbiales

L'étude des nominalisations déverbiales s'est faite en plusieurs phases, d'abord quantitative automatisée afin de repérer des fonctionnements nécessitant une interprétation, puis qualitative et focalisée sur l'analyse des nominalisations-termes.

4.1 Approche automatisée

Dans l'analyse des nominalisations-termes proprement dite, l'approche automatisée a porté sur deux aspects : tout d'abord, la fréquence d'utilisation des nominalisations déverbiales dans les deux corpus et ensuite, le repérage de nominalisations déverbiales fonctionnant sur un mode « déviant », c'est-à-dire pour cette étude, l'ellipse de la préposition et du déterminant lorsqu'elles apparaissent en syntagme (en tête ou en expansion).

4.1.1 Nominalisations déverbiales en corpus scientifique vs en corpus de presse

On prétend généralement que les nominalisations sont particulièrement présentes dans les corpus spécialisés. Ainsi, dans des études précédentes (Condamines, 1998 ; Condamines et Bourigault, 1998) nous avons montré une présence largement plus nette des nominalisations dans les corpus spécialisées que dans des corpus littéraires, comme en témoigne le tableau que nous reproduisons ci-dessous (extrait de Condamines et Bourigault, 1998). Ce tableau rend compte de la répartition de trois catégories d'unités : les nominalisations, les noms qui ne sont pas des nominalisations et les verbes, d'une part dans trois romans et d'autre part dans des manuels techniques, l'un provenant de la DDE (Direction Départementale de l'Équipement) de la Haute Garonne, les deux autres d'EDF.

	Nominalisations	Autres noms	Verbes
Sartre	8%	48%	44%
Balzac	7%	53%	40%
Châteaubriand	6%	50%	44%
DDE	25%	54%	22%
EDF 1	20%	63%	17%
EDF 2	24%	55%	21%

Tableau 2 : répartition nominalisations/noms/verbes en corpus spécialisé vs en corpus littéraire

Les résultats présentés dans ce tableau sont très parlants : la proportion des nominalisations dans les corpus techniques est entre 3 et 4 fois plus élevée que dans les corpus littéraires. Mais surtout, cette présence élevée de nominalisations fait baisser pratiquement d'autant le nombre de formes verbales du corpus. Cette sur-utilisation de la forme nominale dont on peut penser qu'elle se substitue à la forme verbale pour exprimer une action constitue une sorte de d'anomalie par rapport à la norme :

« Dans le cas du nom déverbal [...], l'action ne trouve pas dans le verbe son expression grammaticale privilégiée mais rencontre une catégorie qui lui est « étrangère » (Rémi-Giraud, 1996, 109) ».

On retrouve ici la notion d'anomalie, de « déviance ».

Afin de faire écho à ces données, nous avons effectué la même étude sur les corpus du spatial (corpus scientifique et corpus de presse) afin de voir si nous rencontrons un phénomène similaire. Nous avons utilisé la même ressource Verbaction. Nous avons toutefois augmenté le nombre de nominalisations par l'ajout de nominalisations qui ne sont pas dans la ressource mais qui apparaissent dans le corpus scientifique. Pour repérer ces nominalisations, nous avons mis en œuvre une méthode similaire à celle décrite dans (Namer, 2009) mais dans une version simplifiée. Nous avons utilisé la fonction d'Antconq qui permet de faire la liste des unités par ordre alphabétique inversé, ce qui permet de lister les mots qui ont le même suffixe. Nous nous sommes intéressées ici essentiellement aux termes se terminant par *-ion*, *-age* et *-ence*. Nous avons ainsi pu rajouter manuellement 48 nominalisations à Verbaction (par ex. *autocollimation*, *débullage*, *multiplexage*, *constringence*, etc.).

Sur la base de cette liste de nominalisations étendue, nous avons obtenus les résultats suivants :

	Nominalisations	Autres noms	Verbes
Corpus scientifique	19 %	50 %	31 %
Corpus de presse	10 %	54 %	36 %

Tableau 3 : Répartition des nominalisations vs. autres noms – Comparaison CScience et CPresse

Les résultats, bien que moins tranchés que dans l'étude précédente, montrent la même la double tendance : une diminution du nombre de nominalisations dans le corpus de presse et une augmentation du nombre de verbes.

Cette première analyse quantitative confirme donc notre hypothèse : la déspecialisation s'accompagne d'une diminution des nominalisations déverbiales au profit des formes verbales.

4.1.2 Sélection des nominalisations

Dans l'objectif, d'une part, de comparer le fonctionnement des termes dans un corpus spécialisé et dans un corpus de presse et, d'autre part, de mener une analyse qualitative fine des fonctionnements en jeu, nous avons ciblé la présente étude sur l'observation de nominalisations qui sont des termes du domaine spatial.

Afin de sélectionner les nominalisations à observer, nous avons utilisé deux caractéristiques de fonctionnement : l'un strictement quantitatif mis en œuvre par le logiciel TermoStat qui permet de repérer les termes, l'autre distributionnel qui concerne la possibilité pour les termes-syntagmes de fonctionner sans déterminants ni préposition (Collet, 1997) (par ex. *balise alimentation* au lieu de *balise d'alimentation* ou encore *gain laser* pour *gain du laser*)^{iv}.

L'application du premier critère nous a permis de sélectionner 149 termes-nominalisations dans la liste extraite par Termostat.

La mise en œuvre du second critère s'est faite de la façon suivante. En projetant la ressource Verbaction sur le corpus CScience, nous avons recherché l'ensemble des combinaisons [NN] attestées, dont un des N au moins est une nominalisation de Verbaction. À l'issue de cette étape, nous avons trouvé 1781 occurrences comportant une structure avec nominalisation. Les occurrences obtenues ont ensuite fait l'objet d'un nettoyage afin de ne conserver que les cas où la structure sans préposition comporte une nominalisation et peut permuter avec une structure avec préposition : par ex. *coffret alimentation/coffret d'alimentation* vs. *coffret balise* ou *défaut secteur* (qui ne comportent pas de nominalisation) ou *écart type* (qui ne présente pas de permutation possible avec une structure avec préposition).

À l'issue de ce tri, il n'est resté que 325 occurrences faisant intervenir 78 nominalisations différentes.

Nous avons alors fait l'intersection de la liste de ces 78 nominalisations et celle des 149 fournies par TermoStat.

Nous avons retenu ainsi 31 nominalisations qui sont des termes et qui entrent dans des constructions de type [NN] permutable avec une construction de type [N prep N]. Le fonctionnement de ces 31 nominalisations combine donc deux indices de « terminologisation » : un indice statistique (le score attribué par TermoStat pour l'extraction) et un indice de fonctionnement distributionnel (l'absence de préposition).

À court terme, nous examinerons le fonctionnement de ces 31 termes ; cependant, pour cette étude, nous avons focalisé notre attention sur 10 de ces nominalisation termes, choisies au hasard soit : *absorption*, *acquisition*, *alimentation*, *application*, *conception*, *émission*, *gain*, *sortie*, *télémétrie*, *verrouillage*. Ceci constitue selon nous un premier point d'entrée riche pour analyser finement les fonctionnements sémantico-syntaxiques de ce type de nominalisation dans nos deux corpus.

4.2 Description sémantico-syntaxique du fonctionnement de dix nominalisations : analyse qualitative

Après avoir sélectionné les dix nominalisations-termes à observer, chaque occurrence de chacun de ces dix cas a été étudiée dans chaque corpus afin de procéder à une caractérisation de ses contextes d'apparition.

4.2.1 Analyse dans le corpus scientifique

Les tableaux ci-dessous rendent compte de la répartition de ces nominalisations (en tête ou en expansion, avec ou sans préposition) dans les deux corpus.

	En tête Nomi + Prep + (det) N	En tête Nomi + (det) N	En expansion N + Prep + (det) Nomi	En expansion N + Prep + (det) +Nomi
<i>absorption</i>	8	0	41	0
<i>acquisition</i>	14	0	45	4
<i>alimentation</i>	13	22	68	0
<i>application</i>	16	7	0	0
<i>conception</i>	38	1	27	1
<i>émission</i>	1	3	3	12
<i>gain</i>	24	1	26	1
<i>sortie</i>	48	0	88	0
<i>télémesure</i>	0	1	5	0
<i>verrouillage</i>	2	1	2	0
Total	164	36	305	18

Tableau 4 : Corpus CScience : répartition des structures comportant une nominalisation, avec ou sans préposition

	En tête Nomi + Prep + (det) N	En tête Nomi + (det) N	En expansion N + Prep + (det) Nomi	En expansion N + Prep + (det) Nomi
<i>absorption</i>	5	0	0	1
<i>acquisition</i>	16	0	2	2
<i>alimentation</i>	26	1	22	0
<i>application</i>	45	21	24	5
<i>conception</i>	45	0	55	2
<i>émission</i>	224	8	22	0
<i>gain</i>	66	1	6	0
<i>sortie</i>	118	0	74	0
<i>télémesure</i>	0	1	9	0
<i>verrouillage</i>	3	0	1	0
Total	548	32	215	10

Tableau 5 : Corpus CPresse : répartition des structures comportant une nominalisation, avec ou sans préposition

Ces deux tableaux permettent de mettre au jour plusieurs constats préliminaires. Tout d'abord, dans le corpus scientifique :

- 54/523 des occurrences ne font pas intervenir de préposition (soit 10,3 % du total)
- 323/523 des occurrences font apparaître la nominalisation en expansion (61,76 % du total)

Dans le corpus de presse :

- 42/805 des occurrences ne font pas intervenir de préposition (soit 5,2 % du total).
- 225/805 des occurrences font apparaître la nominalisation en expansion (27,95 %).

Ces résultats montrent donc deux tendances nettes :

1. une plus grande présence des structures sans préposition dans le corpus scientifique que dans le corpus de presse (près du double). Ce fonctionnement correspond à ce qui était attendu : en situation de déspecialisation, les nominalisations en syntagmes nominal ont tendance à retrouver un fonctionnement « normal » : les prépositions ont tendance à être rétablies.
2. une plus grande présence des nominalisations en position d'expansion dans le corpus scientifique par rapport au corpus général (plus du double). Cet aspect-là n'a jamais été repéré dans les travaux précédents, à notre connaissance. Dans le corpus science, il concerne des exemples comme *courbe de gain*, *bande d'absorption*. Ce phénomène n'est pas l'objet d'étude de cet article mais il méritera d'être étudié par la suite.

4.2.2 Analyse dans le corpus de presse

Sur la base d'une analyse comparée dans les deux corpus des contextes d'apparition des dix nominalisations (*absorption*, *acquisition*, *alimentation*, *application conception*, *émission*, *gain*, *sortie*, *télémesure*, *verrouillage*) soit en expansion (si la nominalisation est en tête), soit en tête (si la nominalisation est en expansion), nous avons pu identifier quatre catégories de fonctionnement :

1. La nominalisation est utilisée avec des termes techniques dans les deux corpus
2. La nominalisation est utilisée avec des termes relevant du domaine spatial mais aussi avec des noms plus généraux, dans CPresse uniquement
3. La nominalisation est polysémique dans CScience et monosémique dans CPresse (mais avec un sens spécialisé)
4. La nominalisation est polysémique dans CScience (avec un sens spécialisé : processus ou objet) et est polysémique dans CPresse (avec un sens général et un sens spécialisé).

Ces fonctionnements permettent de mettre en évidence la richesse des phénomènes à prendre en compte dans une perspective d'analyse de la détermination.

Cas 1. Les nominalisations sont utilisées avec des termes techniques dans les deux corpus : dans ce cas, les nominalisations sont utilisées de façon très proche dans les deux corpus, c'est-à-dire avec des termes qui relèvent du domaine spatial dans le corpus scientifique et d'un domaine technique (généralement le spatial) dans le corpus de presse. C'est le cas de *absorption*, *acquisition*, *alimentation*, *gain*. Dans tous les cas, ces nominalisations renvoient à des processus.

Si l'on rentre dans le détail de l'analyse, on peut voir un usage plus spécifique de ces noms dans le corpus scientifique ; cet usage spécifique se manifeste par plusieurs fonctionnements. Tout d'abord, lorsqu'elles sont en tête dans le corpus de science, ces nominalisations n'ont à peu près aucun complément en commun avec leurs homologues dans le corpus de presse. Ainsi, parmi les compléments d'absorption dans le corpus science, on trouve : *absorption* + de + (*lumière*, *photon*, *rayonnement*), comme en (1) :

1. *La profondeur d'absorption du photon (endroit du détecteur où le photon va générer un photoélectron de signal) dépend, en particulier, de sa longueur d'onde* (CScience)

Parmi les compléments dans le corpus presse, on trouve : *absorption* + prep + (*carbone, matière, dioxyde de carbone*).

2. *En mesurant à quelques jours d'intervalle la teneur en carbone dans une zone donnée , on peut en déduire le solde entre l' émission et l' absorption de carbone et construire un modèle de calcul.*(CPresse).

De la même façon, pour *gain*, on a d'un côté (Corpus Science) : *gain* + prep + (*encombrement, fiabilité, performances, sécurité*),

3. *Les avantages principaux des circuits hybrides par rapport aux composants discrets portent , d' une part , sur le gain en encombrement et en masse [...].*(CScience).

de l'autre (Corpus Presse): *gain* + prep + (*consommation, temps, masse...*)(voir (4)). Notons aussi l'utilisation, toujours dans CPresse seulement de la structure quasiment figée : *avoir gain de cause*

4. *le successeur de l' Airbus A320 , qui devra prolonger le succès de l' actuelle vache à lait de l' aéronautique européenne , en offrant un gain de consommation [...].* (CPresse).

Deuxièmement, on peut observer parfois l'utilisation exclusive de la nominalisation en position-expansion. Par exemple, pour *absorption*, on trouve des structures (*raie, bande, mesure, spectroscopie*) + de + *absorption* dans le corpus science mais aucune occurrence correspondant à cette structure dans le corpus presse. De même pour *acquisition*, on trouve dans le corpus science : (*balise, capteur, chaîne, faisceau, etc.*) + de *acquisition* (5), mais aucune occurrence de cette structure dans le corpus presse

5. *A partir de l'instant t1, chaque terminal éclaire le contre-terminal à l'aide de sa balise d'acquisition, dont le faisceau couvre tout le cône d'incertitude* (CScience).

Cas 2. La nominalisation est utilisée avec des termes relevant du domaine spatial mais aussi avec des noms plus généraux, dans CPresse uniquement : par exemple, *conception* et *émission*. Ces cas ne sont pas très éloignés des cas précédents et renvoient également à des processus. Cependant, les noms associés à ces nominalisations dans le corpus de presse (en tête ou en expansion) ne relèvent pas toujours d'un sens spatial ni même d'un sens technique mais d'un sens général. Citons par exemple pour *conception*, dans CPresse, la coexistence de syntagmes tels que *conception de fusées* (spatial), *conception de composants électroniques* (technique), mais également *conception de lieux de loisir*. À l'inverse, dans le corpus scientifique, on trouve uniquement : *conception de* + (*baffle, de circuit, ASICS*).

6. *La conception d'un système électronique se fait en associant plusieurs dizaines de composants discrets sur un support appelé circuit imprimé* (CScience).
7. *Les universitaires , qui profitent d' un recul de dix ans , doivent maintenant la formaliser plus rigoureusement et ambitionnent à terme de fournir aux industriels un outil pour prévoir le temps et le coût de conception de ces grands systèmes* (CPresse).

Parallèlement, *émission* a ceci de particulier qu'il est entré dans la langue générale avec un sens spécifique qui concerne la télévision ou la radio. Ce sens est très fréquent dans le corpus presse (8) mais le sens technique est aussi présent : *émission* + de + (*dioxyde carbone, lumière, rayons*) (9).

8. *Concevoir, réaliser et diffuser ou faire diffuser par satellite ou tout autre moyen, des émissions de télévision ayant un caractère culturel et international au sens large* (CPresse).

9. *Mais le système imaginé à Kyoto en 1997, qui permet aux pays de commercialiser leurs émissions de dioxyde de carbone, ouvrira une ère nouvelle dans la coopération internationale (CPresse).*

Dans ces cas, semblent cohabiter dans la presse un sens « terminologique » (pas forcément spatial) et un sens « général ».

Le cas de *application*

Pour cette nominalisation, on retrouve le fait que, dans le corpus de presse, cohabitent un sens spécialisé (spatial) et un sens général. Cependant, dans ce cas, le nombre de la nominalisation, singulier ou pluriel, joue un rôle discriminant très net. Dans CScience, qu'elle soit au pluriel ou au singulier, cette nominalisation, en tête, ne concerne que les utilisations possibles d'un satellite : *application* + de + (*métrologie, prévision météorologique, radiométrie infrarouge, etc.*), *application* + (*télécom, etc.*). Dans le corpus de presse, ce sens est très présent aussi mais la nominalisation est alors très majoritairement au pluriel : *applications* + de + (*défense, géolocalisation, navigation, télécommunication, métrologie, optique, etc.*). Lorsqu'elle est au singulier, cette nominalisation peut être très rarement utilisée pour désigner les utilisations possibles d'un satellite (*application de navigation + messagerie électronique*) (10) mais l'on trouve en majorité des mots très généraux : *application* + de + (*téflon, matériaux*) (11), ou relevant du droit : *application* + de + (*article, loi, normes, réforme*) (12).

10. *Si ce temps n' affecte pas l' utilisation d' **une application de messagerie électronique** ou de mise a jour d' une base de données , il se revele penalisant pour des applications multimedias ou de voix sur IP (CPresse).*
11. *La toute première **application du Teflon** sur une poêle à frire remonte à 1951 aux Etats-Unis (CPresse).*
12. *Une assemblée générale susceptible de révoquer ses dirigeants, convoquée à l'initiative d'associations de petits porteurs en **application de la loi** sur les nouvelles régulations économiques (CPresse).*

Il est évident, mais c'est tout particulièrement le cas avec *application*, que l'analyse des nominalisations et de leurs contextes en syntagme devrait être complétée par l'analyse des structures verbales correspondantes. Une telle analyse pourrait confirmer les interprétations sémantiques réalisées sur la seule base de l'analyse des contextes des nominalisations. Ainsi, pour *application*, la transformation syntactico-sémantique des structures verbales serait sans doute nettement différente pour chacun des sens identifiés. Pour le sens « spatial », la structure verbale serait sans doute [appliquer X **à/sur** Y] avec X relevant de la catégorie « satellite » et Y relevant de la catégorie « fonctionnalité » et la transformation en forme nominale donnerait [application **de** Y] (*application de météorologie*) avec suppression de X. Dans les autres cas, la structure verbale serait sans doute assez similaire [appliquer X **à/sur** Y] mais la transformation en nom aboutirait à [application **de** X] (*application de téflon, application de la loi*), ce qui d'ailleurs permettrait à la position du complément d'objet indirect (Y) de rester libre et autoriserait des structures avec complément indirect comme dans (11) *application du téflon sur une poêle à frire*, type de construction qui est tout à fait impossible avec *application de (météorologie + observation + radiotéléphonie...)*. Cette hypothèse devrait bien sûr être vérifiée par l'étude de toutes les occurrences du verbe *appliquer*.

Cette analyse comparée des structures argumentales des formes verbales et des formes nominales sera l'objet d'une prochaine étape de l'analyse.

Cas 3. La nominalisation est polysémique dans CScience et monosémique dans CPresse (mais avec un sens spécialisé, résultat) : dans cette catégorie de fonctionnement, la nominalisation semble fonctionner avec deux sens dans le corpus spécialisés et un seul dans le corpus général. Les nominalisations concernées sont *télémesure* et *verrouillage*. Dans le corpus spécialisé, ces deux nominalisations semblent pouvoir renvoyer soit au processus (*débit de télémesure, fréquence de verrouillage*) (exemple 13) soit au résultat du processus, qui se manifeste sous une forme tangible (*système de télémesure, connecteur circulaire à verrouillage*) (exemple 14).

13. *La bande passante et le débit de télémesure sont donc donnés par les relations : #formule# et débit de télémesure (bit s-1) par bande = #formule# avec p = nombre de bit par échantillon (CScience)*
14. *Entrée secteur : Cordon 2P+T équipé d'une fiche secteur 'Européenne ' de type démontable (250 V – 10/ 16 A) d'un côté et d'un connecteur circulaire à verrouillage de l' autre (CScience).*

Dans le corpus de presse, seule une des deux valeurs semble possible, qui concerne l'aspect « résultat d'un processus » : la valeur strictement résultative (exemple 15) pour *télémesure* et la valeur objet pour *verrouillage* (*données de télémesures, poignée de verrouillage*).

15. *Elle a obtenu en mars un contrat de 7,15 millions d' euros du Conseil régional de La Réunion pour construire une station de réception d' images satellite de haute résolution et fournir des données de télémesure pendant trois ans (CPresse).*

Cas 4. La nominalisation est polysémique dans CScience (avec un sens spécialisé : processus ou objet) et est polysémique dans CPresse (avec un sens général et un sens spécialisé) : dans cette catégorie, nous allons examiner le cas de *sortie*. Il semble avoir un fonctionnement assez proche de *télémesure* et *verrouillage* puisque, comme eux, il est utilisé de deux façons dans le corpus science : processus : *sortie* de + (*initialisation, fabrication*), (*cadence, fréquence*) + de + *sortie* (exemple 16) et résultat-objet tangible : *sortie* + de (det) + (*détecteur, laser, carte*), (*face, grandeur, pupille*) + de + *sortie* (exemple 17). Comme pour *télémesure* et *verrouillage*, on peut donc considérer que *sortie* à deux sens techniques dans le corpus spécialisé et un seul (processus) dans le corpus de presse (18, 19). En revanche, *sortie* a aussi un sens général dans le corpus de presse, dont on peut faire l'hypothèse qu'il a précédé le sens technique : *sortie* + de + (*piste, route, prison*), *sortie* + en + (*salle, librairie, mer*), (*année, permission, classement, tentative, etc.*) + de + *sortie*.

16. *Les composants élémentaires sont en général contrôlés en sortie de fabrication (CScience).*
17. *Ceci requiert un traitement anti-reflet sur la face de sortie de la diode laser (CScience)*
18. *Pas de sortie de navette en septembre pour la Nasa (CPresse)*
19. *Il faut au moins deux jours rien que pour vérifier le scaphandrier de sortie, précise Claudie André-Deshaies (CPresse).*

Sur la base de l'observation de ces 10 nominalisations, nous avons donc à ce stade identifié quatre types de fonctionnements sémantiques. Ces fonctionnements sont organisés d'une part en fonction du domaine : spatial/technique vs. général, d'autre part en fonction de la valeur de la nominalisation : processus ou résultat. Les tableaux suivants rendent compte de ces phénomènes :

		Corpus science	Corpus presse
Contexte Technique/spatial	processus	X	X
	résultat		
Contexte Général	processus		
	résultat		

Tableau 6 : fonctionnement sémantique de *absorption, acquisition, alimentation* et *gain*

		Corpus science	Corpus presse
Contexte Technique/spatial	processus	X	X
	résultat		
Contexte Général	processus		X
	résultat		

Tableau 7 : fonctionnement sémantique de *conception, émission* et *application*

		Corpus science	Corpus presse
Contexte Technique/spatial	processus	X	X
	résultat	X	
Contexte Général	processus		
	résultat		X

Tableau 8 : fonctionnement sémantique de *télémessure* et *verrouillage*

		Corpus science	Corpus presse
Contexte Technique/spatial	processus	X	X
	résultat	X	
Contexte Général	processus		
	résultat		X

Tableau 9 : fonctionnement sémantique de *sortie*

4.2.3 Synthèse de l'analyse des dix nominalisations

Nous ne pouvons bien sûr pas tirer des conclusions générales à partir de l'étude de ces dix termes, mais ces descriptions permettent d'ores et déjà de mettre au jour la diversité et la finesse des aspects à prendre en compte pour l'observation des phénomènes de déterminologisation et désécialisation.

De plus, à ce stade même préliminaire, il semble bien que des fonctionnements commencent à se dégager concernant la situation de désécialisation. Premièrement, du point de vue strictement quantitatif, nos résultats confirment que le nombre de nominalisations diminue dans les corpus de langue générale par rapport aux corpus spécialisés. Du point de vue qualitatif, à travers l'observation de la combinatoire en

syntagmes, une tendance se dessine en faveur du rétablissement des prépositions. D'autre part, la position préférée des nominalisations dans le corpus de presse semble être en tête de syntagme alors que c'est la position en expansion qui domine dans les corpus spécialisés. Du point sémantique, si certaines nominalisations semblent fonctionner de manière à peu près similaire dans les deux corpus, d'autres montrent un fonctionnement sémantique plus élargi dans le corpus de presse (un sens général cohabite avec le sens spécialisé). Plus étonnant, dans certains cas, le terme en corpus spécialisé semble être plus polysémique que dans le corpus de presse.

Remarquons que la grille sémantique qui permet l'organisation de ces fonctionnements (spatial/technique vs. général ; processus vs. résultat) pourra servir pour l'analyse de toutes les nominalisations, puis de l'ensemble des termes.

5 Conclusion

Dans la perspective de commencer à baliser l'étude de la déterminologisation avec une approche de linguistique de corpus, l'étude présentée dans cet article présente deux facettes, complémentaires. L'une cible la mise en place d'une méthode, l'autre la caractérisation des premiers résultats obtenus. L'analyse présentée a pour point de départ un certains nombres de fonctionnements avérés dans les corpus spécialisés, par rapport aux corpus généraux : une fréquence élevée de nominalisations déverbiales et une ellipse fréquente des prépositions. Nous avons donc souhaité examiner particulièrement les nominalisations obéissant à ces deux critères, dans un corpus scientifique et dans un corpus de presse, tous les deux concernant le domaine spatial. La méthode, outillée, nous a permis, dans un premier temps d'identifier les nominalisations-termes i.e. possédant les deux caractéristiques recherchées afin que nous focalisions notre analyse sur elles. L'analyse des structures comportant une nominalisation, soit en tête, soit en expansion et de leurs contextes syntagmatiques a été faite à l'aide d'un concordancier mis en œuvre sur des corpus étiquetés. Les premiers résultats constituent un double apport à l'analyse du phénomène de la déterminologisation :

- un apport méthodologique, quant à la réflexion sur les corpus à examiner, l'intégration d'outils de différentes natures dans la mise en place des observables et la manière d'articuler approches quantitatives, qualitatives et descriptions fines dans des données complexes, de grande taille ;
- un apport à la caractérisation sémantique des phénomènes à travers la constitution d'une grille d'interprétation qui a permis de mettre au jour, pour l'instant, quatre catégories de fonctionnement mais aussi repérage de la complexité des fonctionnements sémantiques entre corpus scientifique et corpus de presse (dont un cas où la couverture sémantique du terme semble plus large dans le corpus de science que dans celui de presse).

De nombreuses pistes s'offrent à nous pour continuer à analyser les phénomènes de déterminologisation et de déspecialisation. Tout d'abord, nous allons devoir sans doute étendre le corpus spécialisé. En effet, comme nous l'avons précisé, le domaine du spatial s'est constitué par conjonction de différentes connaissances techniques. Bien qu'il soit assez difficilement envisageable de chercher à représenter l'ensemble des sous-disciplines et techniques en jeu, notre corpus scientifique devrait toutefois être en partie complété dans cette perspective. Nous allons aussi mieux prendre en compte la dimension de ces corpus, entre les années 1990 et 2011 (ajustement des corpus et comparaisons diachroniques fines des phénomènes observés). Du point de vue de la caractérisation sémantique, plusieurs études sont prévues, à court et moyen termes. Tout d'abord, nous prévoyons d'étudier l'ensemble les 31 nominalisations-termes que nous n'avons pas encore examinées. Cette étude nous permettra d'évaluer la pertinence de la grille d'analyse définie, en particulier pour ce qui concerne l'existence d'une couverture plus large pour certains termes dans le CS que dans le CPresse. En effet, les deux cas repérés dans cette catégorie (*télémesure* et *verrouillage*) concernent assez peu d'occurrences (une dizaine), malgré la taille de CPresse. Nous voulons aussi étudier le fonctionnement des formes verbales correspondant aux nominalisations afin de vérifier si, dans le CPresse, elles viennent remplacer, au moins dans certains cas, l'utilisation des nominalisations, ce qui constitue une de nos hypothèses de départ. Il faudra enfin que nous étendions l'analyse des contextes

des nominalisations et des termes en général. Il est clair en effet que n'étudier que le contexte syntagmatique des nominalisations constitue un cadre d'étude qui s'est avéré pertinent pour cette première étude mais qui est trop restrictif pour une analyse totale. Par exemple, il est évident que la prise en compte de contextes plus larges nous aiderait à discriminer les interprétations processuelles vs. résultative/objet (par exemple : au moment de la sortie (processus) de vs la sortie présente un défaut (objet) (exemples forgés)).

Le terrain d'une analyse de linguistique outillée des phénomènes de déterminologisation a été bien déblayé et les premières bases (méthodologiques et interprétatives) de l'édifice instaurées. Mais le chantier est encore en cours.

Références bibliographiques

- Anthony, L. (2005). AntConc: design and development of a freeware corpus analysis toolkit for the technical writing classroom. Professional Communication Conference. *IPCC 2005. Proceedings*. pp. 729-737.
- Bowker, L. et Pearson, J. (2002). *Working with Specialized Language: A Practical Guide to Using Corpora*. London/New York: Routledge.
- Calberg-Challot, M. (2007). Quand une langue de spécialité emprunte au langage courant : le nucléaire, étude de cas. In Humbley, J. (dir.), *Aspects de la recherche en langues de spécialité, Cahier du CIEL 2007-2008*, Paris, Université Paris VII, UFR EILA. pp. 71-85.
- Collet, T. (1997). La réduction des unités terminologiques complexes de type syntagmatique. *Meta : journal des traducteurs*, Vol. 42(1). pp. 193-206.
- Condamines, A. (1998). «Analyses des nominalisations dans un corpus spécialisé : comparaison avec le fonctionnement en corpus "général". Clas A., Mejri S., Baccouche T. (eds.) : *La Mémoire des mots*, Actes des Journées Scientifiques du réseau Lexicologie, Terminologie, Traduction, Tunis, 25-27 septembre, Montréal : Aupelf. pp.351-368.
- Condamines, A. (2003). «Sémantique et Corpus Spécialisés : Constitution de bases de connaissances terminologiques. Mémoire d'Habilitation à diriger les Recherches. Université Toulouse2 Le Mirail.
- Condamines, A. et Bourigault D. (1999). Alternance nom/verbe : explorations en corpus spécialisés. B.Victorri et François J. (eds) : *Sémantique du lexique verbal, Actes de l'atelier de Caen, 22-23 janvier 1999, Cahiers de l'Elsap*. pp.41-48.
- Condamines A. et Picton A. (2011). Les mots du spatial chez le grand public, une histoire entre imaginaire et réalité. In Azoulay G., Pestre D. (dirs.) *Polyphonies spatiales*, Paris : Gallimard. pp.190-192.
- Condamines, A. et Picton, A. (à paraître). « Des communiqués de presse du Cnes à la presse généraliste. Vers un observatoire de la diffusion des termes ». P. Dury, F. Maniez (eds).
- Condamines, A. Dehaut, N. et Picton, A. (2012). Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Études outillées en corpus. C. Gérard et J. Kabatek (eds.). *Cahiers de Lexicologie* n°101, Néologie sémantique et analyse de corpus. pp.161-184.
- Drouin, P. (2003). Term Extraction Using non-Technical Corpora as a Point of Leverage. *Terminology*, Vol. 9(1). pp. 99-117.
- Dury, P. (2007). La déterminologisation du formant eco- et la terminologisation de l'adjectif carbon neutral : un aperçu diachronique de la migration des termes. *Actes de la 7ème Conférence Terminologie et Intelligence Artificielle (TIA)*. Sophia-Antipolis, France, 8-10 octobre. pp. 61-70.
- Galisson, R. (1978). *Recherches de lexicologie descriptive, la banalisation lexicale*. Paris : Nathan.
- Gouadec, D. (1990). *Terminologie : Constitution des données*. Paris : AFNOR.
- Guilbert, L. (1975). *La créativité lexicale*. Paris : Larousse.
- Hathout, N., Namer, F. et Dal, G. (2002). An Experimental Constructional Database: The MorTAL Project. In Boucher, P. (éd.), *Many Morphologies*. Cascadilla, Somerville, Mass. pp.178-209.

- Jacobi, D. (1986). *Diffusion et vulgarisation, itinéraires du texte scientifique*, Besançon : Presses Universitaires de Franche-Comté.
- Kittredge, R. et Lehrberger J. (dirs.) (1983). *Sublanguage: Studies of Language in Restricted Semantic Domains*, Berlin/New York: W. de Gruyter.
- Kocourek, R. (1991). *La langue française de la technique et de la science*. 2e édition. Wiesbaden : Oscar Brandtetter.
- Mc Naught, J., Nkwenti-Azeh, B., Martin, W. et Ten Pas, E., (1991). Eurotra 7-1: Feasibility of Standards for Terminological Description of Lexical Items. Luxembourg : *Commission of the European Communities, Directorate-General, Telecommunications, Information Industries and Innovation*.
- Meyer, I. et Macintosh, K. (1999). *L'étirement' du sens terminologique : aperçu du phénomène de la déterminologisation*. In Béjoint P., Thoiron P. (dirs.), *Le sens en terminologie*. Lyon : Presses Universitaires de Lyon. pp. 198-217.
- Moirand, S. (2007). *Les discours de la presse quotidienne. Observer, analyser, comprendre*. Paris : Presses universitaires de France, Linguistique nouvelle.
- Namer, F. (2009). Nominalisation et composition en français: d'où viennent les verbes composés?, *Lexique* 20. pp. 169-201.
- Nicolae, C. et Delavigne V. (2009). Naissance et circulation d'un terme : Une histoire d'exoplanètes. *Textes et corpus*, Vol. 4, Actes de sixièmes journées de Linguistique de Corpus, Lorient, septembre 2009. pp. 143-155. En ligne http://www.licorn-ubs.com/jlc6/ACTES/Nicolae_JLC09.pdf (consultée le 31.10.2013).
- Noailly, M. (1990). *Le substantif épithète*. Paris : Paris : Presses universitaires de France.
- Picton, A. (2009). *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*. Thèse de doctorat en Sciences du Langage, Université Toulouse 2.
- Rastier, F. et Valette, M. (2009). De la polysémie à la néosémie. *Le français moderne*, Meiri S. (dir.), *La problématique du mot*, Vol. 77. pp. 97-116.
- Rémi-Giraud, S. (1996). Pour une approche notionnelle de la nominalisation. In *Les noms abstraits, histoire et théorie*. Lille : Presses Universitaires du Septentrion. pp.105-116.
- Roqueplo, P. (1974). *Le partage du savoir, science, culture, vulgarisation*. Paris : Le seuil.
- Sablayrolles, J.-F. (2000). *La néologie en français contemporain*. Paris : Honoré Champion.
- Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. In Actes de l'atelier *ACL SIGDAT-Workshop*, Dublin, Irlande. En ligne, <ftp://ftp.ims.uni-stuttgart.de/pub/corpora/tree-tagger2.pdf>
- Unguraneu, L. (2003). *L'interpénétration langue générale-langue spécialisée dans le discours d'internet*, Thèse de doctorat, Université Paris 13/Université Technique de Moldova.

ⁱ Nous remercions le CNES qui nous a confié cette étude et qui, pour ce faire, a mis à notre disposition ces corpus scientifiques.

ⁱⁱ <http://www.lexisnexis.com/> (consultée le 31.10.2013)

ⁱⁱⁱ Cette ressource est disponible sur le site de CLLE-ERSS : <http://redac.univ-tlse2.fr/lexiques/verbaction.html> (consultée le 31.10.2013)

^{iv} Ces structures NN sont bien étudiées dans l'ouvrage de M. Noailly intitulé « le substantif épithète » (Noailly, 1990).