



**HAL**  
open science

## Des communiqués de presse du Cnes à la presse généraliste Vers un observatoire de la diffusion des termes

Anne Condamines, Aurélie Picton

### ► To cite this version:

Anne Condamines, Aurélie Picton. Des communiqués de presse du Cnes à la presse généraliste Vers un observatoire de la diffusion des termes. P. Dury, J. Carlos de Hoyos, J. Makri-Morel, F. Maniez, V. Renner, M. Belen Villar Diaz. La néologie en langue de spécialité, CRTT, pp.165-188, 2014, 978-2-9533061-0-1. halshs-01108940

**HAL Id: halshs-01108940**

**<https://shs.hal.science/halshs-01108940>**

Submitted on 23 Jan 2015

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

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

# Des communiqués de presse du Cnes à la presse généraliste Vers un observatoire de la diffusion des termes

Anne Condamines (1) et Aurélie Picton (2)<sup>1</sup>

## Résumé

Cet article s'intéresse au phénomène de la dé-terminologisation, c'est-à-dire aussi bien au passage qu'à l'usage des termes dans la langue dite générale. La problématique s'est élaborée à partir d'une demande du Cnes (Centre National d'Etudes Spatiales) qui souhaitait voir comment le spatial imprègne la connaissance commune. Nous avons ré-interprété cette demande en mettant en place un observatoire permettant d'examiner, à partir des méthodes de terminologie outillée, les phénomènes de dé-terminologisation. Cet observatoire est composé d'un corpus d'articles de presse générale quotidienne et de communiqués de presse du CNES (disponibles sur internet). A partir des candidats-termes extraits grâce à Termostat, nous avons pu constituer une version « termes » du corpus de presse, constituée des articles comportant au moins 1 candidat-terme. L'article présente les questions de méthodologie de constitution du corpus et d'analyse outillée que pose l'étude de la dé-terminologisation. Il propose aussi quelques résultats qui ont déjà pu être obtenus grâce à l'analyse comparée (statistique, lexicographique et sémantique) dans les corpus de l'observatoire, du fonctionnement de certains termes-candidats (particulièrement ceux construits à partir de *satellite*) et il expose les perspectives d'analyses, nombreuses, qui pourraient être menées grâce à la constitution de cet observatoire, sur les phénomènes de glissement des termes vers la langue générale. A terme, cet observatoire permettra de caractériser systématiquement les éléments permettant de décrire la dé-terminologisation.

## Mots-clés

Constitution de corpus, Dé-terminologisation, Diffusion des termes, Domaine spatial, Observatoire de textes, Terminologie outillée.

## 1. Introduction

Les phénomènes de néologie par diffusion des termes des langues de spécialité vers la langue générale ont été peu étudiés de manière systématique et lorsqu'ils le sont, il s'agit le plus souvent d'un point de vue lexicographique. Les questions qui se posent alors concernent le choix des termes à sélectionner comme entrées des dictionnaires ainsi que la définition à retenir, alors que le dictionnaire ne s'adresse pas à des experts (Josselin-Leray, 2005; L'Homme, Vandaele, 2007; Maniez et al., 2008).

Dans cet article, nous abordons cette problématique d'une manière nouvelle et, à notre connaissance, inédite. En effet, notre réflexion est née d'une demande pratique de l'agence spatiale française, le Cnes (Centre National d'Études Spatiales). Cet organisme souhaitait savoir comment le spatial *imprègne* la langue commune, question, que, à travers notre regard de linguistes, nous avons interprétée en une problématique linguistique : comment les termes

---

<sup>1</sup>

(1) CLLE ERSS, UMR 5263, CNRS et Université Toulouse 2 Le Mirail, [anne.condamines@univ-tlse2.fr](mailto:anne.condamines@univ-tlse2.fr)  
(2) TIM, Faculté de Traduction et d'Interprétation, Université de Genève, [aurelie.picton@unige.ch](mailto:aurelie.picton@unige.ch)

du spatial se diffusent-ils dans la langue générale? Souhaitant mettre en œuvre des méthodes de linguistique de corpus et de terminologie textuelle, nous nous sommes, dans un premier temps, interrogées sur la méthode à définir pour procéder à une analyse du passage des termes dans la langue générale et, au-delà, à construire un véritable observatoire de ce phénomène, qui ne parte pas d'une liste prédéfinie de termes à observer, mais qui permette de faire émerger des textes les termes pertinents à observer sur une période donnée. Cet article relate donc la mise en place de cet observatoire ainsi que les premières études que nous avons pu mener, qui nous permettent d'évaluer les possibilités et les limites de cette démarche. La partie 2 de cet article précise le contexte de notre étude, à savoir la demande du Cnes, l'interprétation de la problématique du point de vue de la terminologie et la réflexion sur la méthode à mettre en place. La partie 3 concerne la mise en place de l'observatoire proprement dit (constitution des corpus et préparation des données). La partie 4 relate, quant à elle, les premières observations que nous avons pu mener. La partie 5 nous permet de tirer les premières conclusions sur l'étude en corpus du phénomène de diffusion des termes et sur la pertinence de l'observatoire construit pour l'observer.

## **2. Origine de l'étude et problématisation en terminologie**

Dans cette partie, nous situons notre étude, qui allie prise en compte d'une demande externe de l'agence spatiale française et inscription de cette demande dans une problématique de terminologie textuelle.

### **2.1. Point de départ de l'étude : la demande du Centre National d'Études Spatiales**

À l'occasion de ses cinquante ans, le Cnes a souhaité réunir différentes productions et points de vue traduisant l'analyse d'autres disciplines scientifiques de tous ordres sur la place du spatial auprès du grand public depuis ces 50 dernières années. À cette occasion, nous avons été sollicitées avec une demande assez inédite : le service de communication du centre voulait en effet savoir comment le spatial, en tant qu'ensemble de connaissances, réelles ou imaginaires, se diffuse dans le grand public, de façon plus ou moins consciente (Condamines, Picton, 2011). Pour reprendre la demande dans les termes mêmes qui nous ont été proposés, il s'agissait de savoir comment les connaissances sur le spatial « perfusent » le grand public. En tant que linguistes travaillant sur des domaines spécialisés, le défi était double : d'une part tenir compte de ce besoin, d'autre part, et surtout, lui donner une interprétation qui le rapproche de problèmes scientifiques connus et travaillés en linguistique. A partir de la métaphore de la perfusion, nous avons surtout travaillé la dimension « injonction, lente, régulière et prolongée » (TLFi) mais aussi l'aspect « non perçu » : un patient perfusé ne ressent pratiquement rien. À partir du point de vue de la linguistique de corpus, cadre de travail que nous adoptons ici, il s'agit alors d'essayer de travailler cette notion de « flux », encore difficile à saisir à l'aide de corpus. Pour tenter toutefois de prendre en compte la dynamique du discours, la méthode généralement utilisée consiste à construire un corpus ouvert composé de textes de différentes périodes, à les comparer et à les mettre à jour régulièrement pour obtenir ce que Renouf appelle des corpus dynamiques (Renouf, 1994). Mais cette mise à jour n'est possible qu'avec certains types de textes produits à intervalles réguliers (par exemple les dépêches de presse) et demande des outils spécifiques pour pouvoir traiter finement ces données linguistiques dynamiques. Pour ces raisons, nous privilégions, dans notre approche, des corpus comparables en diachronie courte. Compte tenu de tous ces paramètres, la question était pour nous de nous interroger d'une part sur les problématiques théoriques que nous devons convoquer pour éclairer cette question et, d'autre part, sur les corpus et la méthodologie que nous pouvions construire pour l'examiner. Dans la section

suiuante, nous présentons différents éléments théoriques en jeu dans cette problématique de diffusion.

## 2.2. Un point de vue terminologique

Bien que rarement systématiquement étudié, ce phénomène de *glissement* (pour reprendre l'idée de flux) de la terminologie vers le lexique général a revêtu plusieurs dénominations dont « banalisation lexicale » par Guilbert (1975) puis Galisson (1978), « dé-spécialisation » (considérée comme équivalente à vulgarisation) chez Gouadec (1990), « déterminologisation », « migration », ou même « dilution » (qui ne concerne que l'aspect sémantique) chez Meyer et al., (1999) ou encore « dédomanialisation » (Rastier, Valette, 2009).

Tout d'abord, dans cette étude, notre « entrée » dans les textes se fait par les termes. Il s'agit d'un choix méthodologique et pratique qui combine à la fois le fait que les termes sont des pivots entre langue et connaissances et que la recherche de chaînes de caractères ou de groupes de chaînes de caractères (des termes) est en adéquation avec les possibilités des outils. Sur la base de la problématique formulée par le Cnes, le point le plus facile à cerner pour nous était que nous allions être confrontées au passage de termes dans la langue générale et donc *a priori* à un phénomène de néologie, de forme ou de sens, dans la langue générale. Étonnamment si ce passage du spécialisé au général est très souvent cité comme un mode possible de néologie (Sablayrolles, 2000), il a été peu étudié de manière systématique. Les travaux existants portent plutôt sur le phénomène inverse, celui qui permet aux terminologies d'emprunter à la langue générale (Guilbert, 1975; Calberg-Challot, 2007); comme mentionné en introduction, une question qui se pose souvent alors est celle de savoir quels termes doivent être consignés dans un dictionnaire de langue générale.

Pour repérer les phénomènes de néologie dans la langue générale, nombreux sont les travaux qui, dans les années 1990, ont proposé de comparer un corpus de langue « générale » (la presse, le plus souvent) et une liste d'exclusion (les entrées d'un ou plusieurs dictionnaires) afin de repérer les formes non recensées dans le dictionnaire et donc candidates à une entrée et une définition dans un dictionnaire de langue générale (Cabré, De Yzaguirre, 1995; Mathieu, 1998). Si ces travaux présentent un intérêt certain, ils ne répondent pas tout à fait à notre problématique. En effet, notre objectif n'est pas lexicographique, puisqu'il ne s'agit pas de construire un dictionnaire des termes du spatial dans la langue générale, mais plutôt d'ordre méthodologique et théorique puisqu'il vise à comprendre comment, par quels mécanismes linguistiques, s'effectue la diffusion des termes du spatial dans le grand public. Il ne peut donc s'appuyer sur un dictionnaire pré-existant (en tout cas pas uniquement).

Deuxièmement, parmi les nombreuses questions que la problématique de ce glissement soulève, nous pouvons en retenir deux ici. La première est qu'il s'agit le plus souvent d'un phénomène non conscient à la fois des locuteurs et des interlocuteurs. Il se différencie alors de la vulgarisation par le fait qu'il ne relève pas d'un processus délibéré de mise à portée de connaissances scientifiques auprès du grand public (Roqueplo, 1974). Mais parallèlement, le fait qu'un terme se diffuse dans la langue générale est souvent lié à un intérêt accru du public/interlocuteur pour la thématique à laquelle est lié ce terme : la médecine, l'informatique, le nucléaire... le spatial. La néologie, de forme comme de sens, n'est donc possible que si elle est collectivement acceptée :

Une évolution [de la norme linguistique] ne saurait donc s'imposer que dans la mesure où elle résulte de causes collectives, autrement dit de facteurs affectant l'ensemble du groupe linguistique considéré (pas forcément la communauté entière). (Nyckees, 2000 : 37)

Une autre question qui se pose vient de la nécessité de s'interroger sur la nature du terme « intégré » à la langue générale. Ainsi par exemple, Ungureanu se demande à propos du terme diffusé « est-ce toujours un terme, est-ce devenu un mot ? » (Ungureanu, 2003 : 108). Pour le dire autrement, si un terme a intégré le lexique commun, est-il toujours, pour les locuteurs, chargé de son sens initial? La « conscience » même de son domaine d'origine est-elle toujours présente?

Il faut aussi s'intéresser à l'impact qu'a cette mobilité vers la langue commune sur le sens lui-même. Dans un grand nombre de cas, on peut penser que le grand public ne connaît pas exactement le sens d'un terme, en tout cas si l'on entend par là qu'il en a une définition experte. En reprenant les propositions de Putnam (1990), et à la suite de Gentilhomme (1994) ou Nicolae et Delavigne (2009), on peut penser que le grand public laisse aux experts la responsabilité du sens exact d'un terme (d'où l'idée « d'usage par délégation », voire par « déférence » proposé par Putnam) mais ne s'empêchent pas d'utiliser ce même terme. C'est même le fait de cette autorisation qui signale que le terme ne relève plus du seul domaine spécialisé, de la seule compétence langagière des experts. On trouve une idée assez proche dans le distinguo qu'établit Wierzbicka (1985) entre concept maximal et concept minimal du terme, repris par Dury (1997 : 26) :

Le spécialiste détiendrait [...] le concept maximal d'un terme ou tout du moins serait celui qui accède aux informations conceptuelles les plus nombreuses [...] et le non-spécialiste, par contre, serait celui qui accéderait au concept minimal, minimum vital mais suffisant pour communiquer.

De fait les usages se trouvent en partie modifiés et, comme le fait la linguistique de corpus, on peut faire l'hypothèse que les contextes d'apparition du terme ainsi passé dans l'usage « courant » vont pouvoir être utilisés pour repérer les modifications sémantiques. En reprenant la proposition de Meillet (1921), également soutenue par Sablayrolles (2000), selon les cas, le sens va être soit étendu mais appauvri, soit restreint mais enrichi.

Enfin, la diffusion des termes dans la langue générale peut se faire *via* des figures de style, l'usage perdant parfois toute référence domaniale. Le phénomène de métaphore est de ce point de vue particulièrement important. Par exemple, on peut parler aujourd'hui d'une *pléiade d'artistes*, de *vouloir décrocher la lune*, ou encore de *tirer des plans sur la comète*. Notons que ces cas-là intéressent aussi nos interlocuteurs du Cnes, qui les considèrent (à juste raison) comme des indices de l'influence du spatial sur le grand public. Mais, faute d'en avoir rencontrés de manière significative à ce stade de notre étude (et sur l'intervalle temporel que nous ciblons), dans la suite de cet article, nous ne parlerons pas de ce type de phénomène.

Compte tenu de ce contexte de questionnement théorique et de la demande du Cnes, nous proposons donc dans cet article une réflexion sur la définition d'une méthode permettant d'examiner les glissements du domaine de l'espace vers la langue générale, d'une part pour identifier en corpus les termes concernés et d'autre part pour étudier les phénomènes linguistiques qui sont à l'œuvre. Nous expliquons cette démarche méthodologique dans les sections suivantes.

## 2.3. Construction d'une démarche méthodologique

### 2.3.1. Une approche basée sur la linguistique de corpus

Pour répondre à notre problématique de la diffusion des termes vers la langue générale, nous situons notre approche dans la linguistique de corpus et dans la terminologie textuelle (Pearson, 1998; Condamines, Dehaut, 2011). Nous nous appuyons ainsi sur les progrès des méthodes outillées qui peuvent être mises en œuvre sur les corpus spécialisés. Nous voyons ainsi comment, selon l'objectif de l'étude, nous pouvons faire appel aux trois différents indices - de formes, de distribution de ces formes et de quantification de ces formes - que nous fournissent les outils tels que les extracteurs de termes, les analyseurs syntaxiques ou les concordanciers. (Condamines, Dehaut, Picton, 2012). Enfin, nous inscrivons nos travaux dans des études de diachronie courte (Dury, 2006; Humbley, 2011; Tolédano, Candel, 2006), ce qui nous amène à mettre en place des méthodes dédiées et à nous interroger sur la nature et le rôle des corpus à examiner (Picton, 2009).

Concrètement, notre approche se met ici en place sur la base de la comparaison de corpus. Dans le cadre de l'étude de la néologie, encore peu de travaux d'exploration outillée se basent sur la comparaison directe de corpus de textes, censés permettre de repérer les termes qui sont apparus (ou sont en train de migrer) dans la langue générale; notons toutefois les travaux de l'IULA (Cabré, Nazar, 2011) et ceux de l'OLST (Drouin et al., 2006).

Mais dans le cadre de notre problématique de diffusion de termes, la volonté de comparer des corpus (et non pas un dictionnaire avec un corpus spécialisé), se heurte à la question du choix de corpus représentatifs (Biber, 1993) pour observer le passage des termes dans le temps.

### 2.3.2. Choix du type de corpus

Le choix de baser notre approche sur une comparaison de corpus impose une réflexion importante sur le genre des textes à prendre en compte pour des données représentatives de ce passage entre langue de spécialité et langue générale. De ce point de vue, le rôle de mise à disposition de la connaissance experte (et donc de diffusion de la terminologie) est généralement attribué aux médias (Jacobi, 1986; Moirand, 2007). Dans la mesure où les corpus journalistiques sont parmi les plus accessibles et ce, en grande quantité, nous avons considéré que la presse était un genre adapté pour observer les termes utilisés dans la langue générale. Toutefois, nous n'avons pas opté pour la presse ayant une vocation de vulgarisation mais pour la presse générale. En effet, il nous a semblé que l'usage y était moins lié à une volonté de vulgarisation que celui que l'on pouvait trouver, par exemple dans la presse spécialisée (La Recherche, Sciences et Vie, etc.) et qu'il pouvait donc mieux refléter une utilisation banalisée, en tout cas supposé « compréhensible » par un lecteur de presse « moyen ».

Du point de vue de la langue de spécialité, le recours à des textes spécialisés du domaine spatial a été écartée pour deux raisons principales : premièrement, l'hétérogénéité des spécialités présentes au Cnes (optique, mécanique, astronomie, etc.) rend difficile la circonscription d'un corpus exploitable et représentatif du domaine. L'accès aux données et à des spécialistes de ces domaines à l'agence spatiale est également limité et ardu, d'autant plus que nous nous inscrivons dans une perspective diachronique (Picton, 2009; Dury, 2004). Pour cette raison, nous avons choisi plutôt de nous intéresser à un genre de textes « à l'interface » entre spécialistes et grand public : les communiqués de presse du Cnes. En effet, dans une perspective de diffusion des connaissances, au moins une autre communauté de locuteurs peut

intervenir entre les experts et les journalistes de presse quotidienne : les chargés de communication des organismes publics et privés. Il est bien évident que ces chargés de communication n'ont pas seulement une intention de diffusion de la connaissance mais aussi de mise en valeur des compétences des sociétés qu'ils représentent. Toutefois, du point de vue de l'usage des termes, les productions de cette communauté discursive nous ont paru intéressantes (hypothèse que l'on retrouve d'ailleurs également chez Nicolae et Delavigne (2009) par exemple). En effet, cet usage et sa diffusion relève pour une grande part d'un processus non-conscient et n'est donc sans doute que peu affecté par une volonté commerciale. Il nous est apparu ainsi que le basculement de l'usage expert vers l'usage « général » pouvait se repérer dans le passage du discours des chargés de communication à celui des journalistes de presse générale qui sont les principaux destinataires de leurs communiqués. Ces derniers sont de plus intégralement disponibles sur le site du Cnes, ce qui nous a permis de les récupérer de manière automatisée, assez aisément.

### 3. Construction de l'observatoire

#### 3.1. Constitution des corpus

Dans cette partie, nous expliquons comment, à partir des extraits de la presse quotidienne, nous avons construit les premières bases d'un observatoire de la diffusion des termes en corpus.

##### 3.1.1. Communiqués de presse (CP)

Le premier corpus constitué regroupe l'ensemble des communiqués de presse mis en ligne par le Cnes sur son site<sup>2</sup>, du 1<sup>er</sup> janvier 2003 au 31 décembre 2011, soit 535 communiqués. Ces textes ont été récupérés automatiquement pour chaque mois de cet intervalle et regroupés par année (Tableau 1). Ce corpus nous sert tout à la fois de base de comparaison pour observer la diffusion des termes, et de base à l'extraction des termes pertinents à observer : nous faisons en effet l'hypothèse que les termes présents dans ce corpus « intermédiaire », à l'intention de la presse, doivent être les cibles privilégiées d'une diffusion dans la langue générale.

Années	Nombre d'occurrences
CP_2003	37 274
CP_2004	30 240
CP_2005	44 645
CP_2006	35 232
CP_2007	20 720
CP_2008	24 692
CP_2009	31 672
CP_2010	28 397
CP_2011	44 875
<b>TOTAL</b>	<b>297 747</b>

Tableau 1 : Corpus de communiqués de presse

<sup>2</sup> <http://www.cnes.fr> (consultée le 31.10.2012)

### 3.1.2. Presse générale (VT et VG)

Le deuxième corpus de l'étude est composé d'articles de presse générale, du 1<sup>er</sup> janvier 1998 au 31 décembre 2011, regroupés également par année. Ces articles ont été récupérés *via* la plateforme LexisNexis®<sup>3</sup> qui propose des abonnements à l'ensemble de la presse nationale et internationale en version électronique. Pour cette étude, nous n'avons sélectionné que les quotidiens nationaux français disponibles intégralement sur la plateforme du 1<sup>er</sup> janvier 1998 au 31 décembre 2011, soit 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. Néanmoins, sur cette base, nous proposons de travailler sur deux versions de ce corpus de presse générale :

- une « version termes » (désormais VT), qui regroupe les articles des quotidiens qui utilisent au moins un terme du spatial (Tableau 2). Cette sélection permet de ne pas regrouper seulement les articles des rubriques spécialisées (par ex. sciences, astronomie, etc.), mais de rassembler tous les articles qui contiennent au moins un terme candidat. En effet, non seulement les usages que nous ciblons peuvent apparaître dans n'importe quelle rubrique (Une, politique, histoire, etc.), mais, après de premières observations, nous nous sommes aperçues que le découpage en rubriques était très instable d'une édition à une autre ou d'un quotidien à l'autre. Cette sélection des articles à partir d'une liste de termes est précisée dans la section suivante.

<b>Années</b>	<b>Nombre d'occurrences</b>
<b>1998</b>	497 938
<b>1999</b>	516 584
<b>2000</b>	492 032
<b>2001</b>	469 738
<b>2002</b>	446 198
<b>2003</b>	491 712
<b>2004</b>	459 526
<b>2005</b>	482 481
<b>2006</b>	428 297
<b>2007</b>	523 817
<b>2008</b>	456 197
<b>2009</b>	504 827
<b>2010</b>	460 671
<b>2011</b>	430 835
<b>TOTAL</b>	6 660 853

Tableau 2 : Corpus de presse « version termes »

- une « version globale » (désormais VG), qui regroupe les articles des quotidiens, sélectionnés aléatoirement (Tableau 3).



<b>Années</b>	<b>Nombre d'occurrences</b>
<b>1998</b>	295 905
<b>1999</b>	336 359
<b>2000</b>	292 501
<b>2001</b>	281 455
<b>2002</b>	367 494
<b>2003</b>	264 701
<b>2004</b>	237 167
<b>2005</b>	346 810
<b>2006</b>	351 970
<b>2007</b>	416 091
<b>2008</b>	373 714
<b>2009</b>	386 459
<b>2010</b>	256 280
<b>2011</b>	303 894
<b>TOTAL</b>	4 512 800

Tableau 3 : Corpus de presse « version globale »

Nous rediscutons plus loin l'impact que le choix de l'une ou l'autre de ces versions peut avoir sur l'analyse (§ 4.1. et § 5.).

### **3.2. Préparation des données**

Pour cette étude outillée, l'ensemble des corpus a été lemmatisé et étiqueté avec TreeTagger (Schmidt, 1995).

Comme mentionné plus haut, les communiqués de presse ont été choisis comme genre de textes à l'*interface* entre les langues de spécialité et la langue générale. Suivant ce choix, nous faisons l'hypothèse que les termes du spatial présents dans ces communiqués sont potentiellement diffusés de manière privilégiée dans la presse et donc, dans la langue générale. Pour cette raison, les candidats-termes du domaine présents dans les communiqués sont extraits automatiquement avec l'extracteur TermoStat (Drouin, 2003). À ce stade de notre recherche, nous arrêtons notre sélection aux 100 premiers candidats-termes proposés par TermoStat. Sur ces 100 candidats, 87 ont été validés, par nous-mêmes et après consultation des experts (par ex. *Cnes, satellite, espace, lanceur, agence spatiale, observatoire, orbite, observation, agence spatiale française, mission, fusée, lancement*).

Cette extraction de termes vise donc un double objectif : premièrement, elle nous permet de cibler directement les termes pertinents à observer dans les corpus de presse, évitant ainsi de choisir des unités à observer *a priori*, sur la base de l'intuition seule ; deuxièmement, cette extraction permet d'orienter la constitution des corpus de presse « version termes » (VT), puisque les articles retenus pour ce corpus doivent contenir au moins un terme de la liste extraite avec TermoStat pour être sélectionnés (cf. § 3.1.1.).

## **4. Premières observations**

Ces différents éléments méthodologiques offrent une base très riche pour l'étude de la diffusion des termes à travers plusieurs types d'observations croisées. Premièrement, le fait de

disposer d'une grande quantité de données supposées représentatives permet d'obtenir des résultats quantitatifs sur la diffusion des termes, éléments encore peu disponibles actuellement en terminologie. Deuxièmement, d'un point de vue qualitatif, des observations linguistiques fines sont également possibles grâce aux outils d'exploration des textes pour l'analyse des différents sous-corpus. Ce regard croisé est garant en parallèle d'une réponse double à la problématique posée dans cette recherche : premièrement, il offre des éléments de réponse à la demande pratique du Cnes (quels termes du spatial sont diffusés et comment le sont-ils auprès du grand public?); deuxièmement, il permet de proposer de nouvelles pistes d'investigation sur des questions théoriques et empiriques en terminologie. Pour montrer cela, nous proposons dans cet article de présenter plusieurs cas d'observation. Afin de guider cette présentation, nous organiserons nos observations autour du terme *satellite* (et de ses composés), qui est le deuxième terme repéré par TermoStat. Après un aperçu global des tendances de diffusion dans nos corpus (§ 4.1.), nous explorerons respectivement quatre cas de figure :

- le suivi de l'évolution de concepts reliés au cours du temps (§ 4.2.)
- le repérage de fonctionnalités qui intéressent le grand public (§ 4.3.)
- la comparaison du mode de construction des termes complexes (§ 4.4.)
- l'observation de changements sémantiques (§ 4.5.).

Ces différentes pistes examinées vont nous permettre d'illustrer la richesse des phénomènes en jeu, tels que l'on peut les saisir à travers l'observatoire que nous proposons. Nous montrons comment peuvent se mettre en place les analyses sur la diffusion des termes en corpus comparables, par des allers-retours entre hypothèses, basées sur des connaissances linguistiques *a priori*, et prise en compte des occurrences des termes et de leurs contextes.

#### 4.1. Observations quantitatives globales

Le fait de disposer de corpus comparables *a priori* représentatifs et de grande taille offre un regard rarement adopté, car difficilement saisissable, sur la « quantification » de certains aspects de la diffusion des termes. Notamment, à partir de l'extraction des 100 premiers termes extraits par TermoStat sur le corpus CP, nous pouvons mesurer la proportion de termes présents dans la langue générale dans nos deux corpus. Ces données permettent de mesurer la diffusion dans les corpus, mais également de s'assurer de leur représentativité.

Ainsi, premièrement, dans le corpus VT, 100% des termes extraits sont présents. Si ce résultat pouvait être attendu dans la mesure où la constitution-même du corpus VT repose sur cette liste de termes, elle permet de nous assurer que tous sont bien représentés dans la presse<sup>4</sup>. Ces termes représentent 254 505 occurrences dans ce corpus, soit 3.9% des occurrences totales. Ces premiers chiffres montrent donc une proportion tout à fait remarquable de termes dans ce corpus de langue générale<sup>5</sup>.

---

<sup>4</sup> Il aurait été possible en effet que certains termes fréquents soient surreprésentés et bloquent la présence des autres termes dans les 200 textes sélectionnés par mois.

<sup>5</sup> Notons toutefois que ces chiffres s'appuient sur des occurrences non désambiguïsées. Certaines occurrences d'unités telles que *satellite*, *étoile* ou *atmosphère* ne sont donc pas (ou ne sont plus aujourd'hui) liées au domaine spatial (par ex. « une danseuse étoile »). Néanmoins, nous reviendrons sur le choix de désambigüiser ou non dans la dernière partie, dans la mesure où en diachronie, cette étape peut entraîner un certain biais dans l'analyse des résultats.

Dans le corpus VG par contre, seuls 71% des termes extraits sont présents. Mais l'élément le plus significatif réside dans le fait que ces termes ne représentent plus que 9 263 occurrences, soit 0.9% des occurrences totales du corpus. De plus, s'il n'est peut-être pas très surprenant de ne pas y retrouver d'occurrences de termes très spécifiques (par ex. *aéronomie* ou *lidar*), il est beaucoup plus surprenant de constater l'absence de projets spatiaux-phares (*emergesat* ou *egnos*), ainsi que des termes peu spécifiques et très fréquents dans les CP, tels que *observation spatiale*, *recherche spatiale* et *système spatial*.

Ces premiers chiffres permettent donc de souligner un comportement très différent entre les deux corpus. En effet, si l'on s'attendait à cibler des articles contenant des termes du spatial de manière privilégiée dans VT, nous ne nous attendions pas à une telle absence de termes dans VG. De fait, ces résultats obtenus dans le corpus VG montrent peut-être plus la manière dont l'information peut se perdre dans la masse de données des médias. En effet, VG et VT ayant été constitués à partir de la même base LexisNexis®, il semblerait qu'avec la sélection totalement aléatoire de textes dans la masse d'articles de presse, l'on perde des occurrences pertinentes. À titre d'exemple en effet, alors que le terme *Cnes* apparaît plus de 4070 fois dans VT, celui-ci n'apparaît plus que 32 fois dans VG. Cette perte de données limite donc considérablement les observations à ce stade. Il apparaît alors clairement que d'autres questions doivent être réglées concernant la constitution d'un corpus de presse générale : ce corpus est-il vraiment représentatif de la masse des termes dans la langue générale? Sinon, comment échantillonner les corpus de presse de manière représentative pour une problématique diachronique telle que la nôtre? Ou encore, comment manipuler autant de données avec les outils actuels, tout en permettant une analyse linguistique fine? En attendant de pouvoir offrir des réponses satisfaisantes à ces questions, dans le cadre de cet article nous mettons ce corpus de côté dans la suite des observations et concentrons nos analyses sur la comparaison des corpus CP et VT.

#### 4.2. Suivi de l'évolution de concepts liés

À travers la comparaison croisée de corpus diachroniques en langue générale et en langue de spécialité, il est intéressant d'arriver à saisir la manière dont les technologies spatiales se diffusent auprès du grand public. Parmi ces technologies, les satellites artificiels sont des concepts centraux bien connus du public français (notamment à travers la filière SPOT<sup>6</sup>) qui, depuis les années 80, connaissent un grand nombre d'évolutions techniques. Parmi ces évolutions, nos corpus font ressortir la question de la miniaturisation de ces objets. On trouve en effet aujourd'hui des mini-satellites (de 100 à 1000kg), des micro-satellites (de 10 à 100kg), des nano-satellites (de 1 à 10kg) et des pico-satellites (moins de 1kg). À titre d'illustration, en regardant leur traitement dans les ressources dictionnaires générales (Tableau 4), on remarque très bien la différence de statut de ces termes entre langue générale et langue de spécialité : si ces technologies sont bien définies et répertoriées dans le Grand Dictionnaire Terminologique ou France Terme, elles ne le sont que partiellement dans le Petit Robert ou le Larousse. L'intérêt alors de travailler à partir de corpus est que cela peut permettre de mieux saisir dans quelle mesure la langue générale a été imprégnée de ces termes.

---

<sup>6</sup> <http://www.cnes.fr/web/CNES-fr/9183-gp-spot-25-ans-de-carriere-25-millions-d-images.php> (consultée le 31.10.2012)

	Larousse ligne	en	PR2012	GDT	FranceTerme
<i>satellite</i>	oui		oui	oui	oui
<i>mini-satellite</i>	oui		non	oui	oui
<i>micro-satellite</i>	oui		non	oui	oui
<i>nano-satellite</i>	non		non	oui	oui
<i>pico-satellite</i>	non		non	oui	oui

Tableau 4 : Traitement des termes de la famille « satellite » dans les ressources terminographiques et lexicographiques (2012)

L'observation de ces termes en corpus permet d'offrir une vision beaucoup plus riche de leur attestation en langue générale, et notamment de saisir l'évolution depuis 1998, en parallèle avec leur évolution dans les communiqués de presse. La Figure 1 présente la répartition de ces termes dans le corpus CP, depuis 2003; la Figure 2 présente la répartition de ces mêmes termes dans le corpus VT depuis 1998.

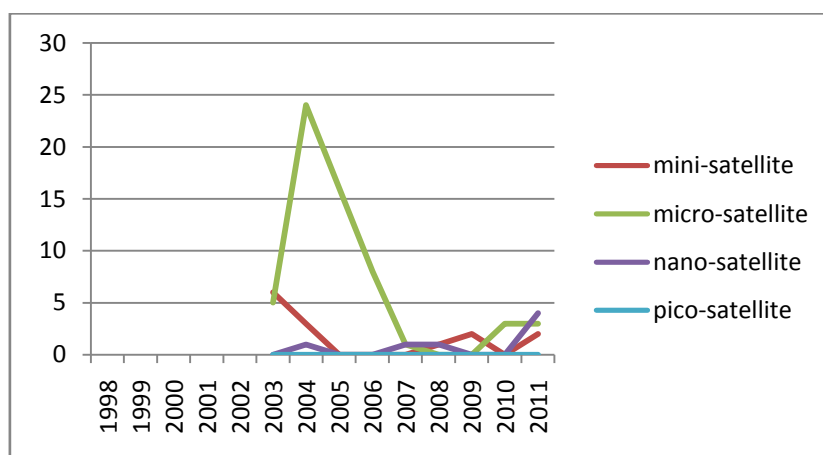


Figure 1 : Évolution de fréquence des nouveaux types de « satellite » - corpus CP

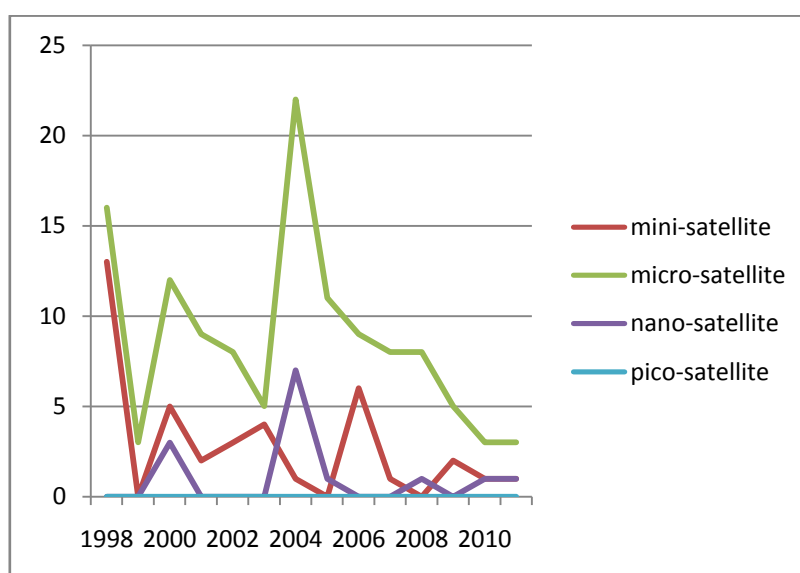


Figure 2 : Évolution de fréquence des nouveaux types de « satellite » - corpus VT

De ces graphiques émergent plusieurs informations, qui permettent de mesurer, du moins partiellement, le parallèle entre l'actualité du Cnes et la langue générale. Ces données offrent donc un premier angle de réponse à la demande pratique. Parmi les éléments saillants :

- Premièrement, tous ces termes (à l'exception de *pico-satellite*) sont attestés dans nos corpus depuis 1998 (et 1999 pour *nano-satellite*) : toutes ces technologies sont développées depuis les années 80 et ont été diffusées dans la presse avant 2000. Nous ne pouvons donc pas observer dans nos corpus de phénomène de néologie formelle ou totale qui serait liée aux activités du Cnes dans ces exemples;
- Deuxièmement, la répartition des termes varie grandement dans le temps et les courbes présentent plusieurs pics de fréquence, notamment pour *micro-satellite*, en 1998, 2000 et 2004. Si les données ne permettent pas de parler de néologie, elles permettent toutefois de comprendre que ces pics sont dus à des actualités fortes dans le paysage spatial. Ainsi, depuis fin 1998, le Cnes développe sa première filière de micro-satellites, appelée Myriade<sup>7</sup>, qui s'accompagne notamment d'appels à projets et de rassemblements scientifiques importants jusqu'en 2000. Ces actualités se répercutent alors immédiatement dans les corpus de presse;
- Enfin et en lien avec le point précédent, ces pics montrent qu'il y a donc bien un lien de diffusion, au moins partiel, entre presse générale et communiqués de presse. En effet, des parallèles importants peuvent être relevés entre les deux graphiques, qui prouvent l'influence – au moins partielle, mais nous y reviendrons – que le Cnes a sur la diffusion de ces termes. Par exemple, les deux graphiques présentent un pic commun en 2004, année du lancement du premier micro-satellite du Cnes, appelé Demeter<sup>8</sup>. Cette information est passée par les communiqués de presse et dans la foulée, dans la presse elle-même.

Ces différents éléments illustrent donc la manière dont des corpus diachroniques comparables permettent de visualiser plusieurs aspects de la diffusion des termes dans la langue générale. Ils montrent notamment le parallèle entre l'actualité spatiale française et la diffusion des termes spatiaux dans la presse, entre les communiqués de presse et la presse, mais également les limites de l'impact des communiqués de presse du Cnes sur la diffusion des termes : s'il est clair que le Cnes est un acteur majeur de l'actualité spatiale en France (comme le prouvent les différents pics de fréquence associables aux projets du Cnes), il est également clair que les informations spatiales, et donc les termes, peuvent être diffusées dans la presse à partir d'autres médias : autres communiqués d'agences scientifiques majeures (CNRS, ESA, etc.), reprises d'autres nouvelles, etc. Il serait donc pertinent de retravailler le corpus CP pour l'ouvrir à d'autres types de textes et améliorer la représentativité de ces données.

#### 4.3. Repérer les fonctionnalités qui intéressent le grand public

Toujours en focalisant nos observations sur le terme *satellite* mais avec un point de vue qualitatif, nous avons pu mettre rapidement en évidence que, parmi les contextes les plus fréquents de *satellite*, on note la présence de structures évoquant les fonctionnalités et les applications possibles liées à l'utilisation des satellites. Il s'agit principalement de la structure [GN par *satellite*] comme dans *observation de la terre par satellite*, le GN étant le plus souvent une nominalisation déverbale<sup>9</sup> et la préposition *par* introduisant le *satellite* comme le

<sup>7</sup> <http://smc.cnes.fr/MYRIADE/Fr/> (consultée le 31.10.2012).

<sup>8</sup> <http://smc.cnes.fr/DEMETER/Fr/> (consultée le 31.10.2012).

<sup>9</sup> Bien que ce ne soit pas toujours le cas, par exemple pour *bourse* ou *chaîne*, cas sur lequel nous revenons plus loin.

moyen par lequel est mise en œuvre la fonctionnalité indiquée par le GN. Cet usage fréquent constitue déjà un indice sur le fait que les lecteurs sont supposés être intéressés par les possibilités offertes par l'utilisation des satellites.

L'examen de cette structure, de la nature et de la répartition des GN dans chacun des corpus va nous permettre d'aborder la question des liens entre phénomènes langagiers et intérêt du public dans le phénomène de la néologie. Le tableau 5 rend compte des 6 premiers GN (par nombre d'occurrences) que nous avons trouvés selon les sous-corpus (VT), dont la majorité semble liée aux concepts de navigation et de télécommunications par satellite. Notons que cette structure apparaît 760 fois dans le corpus de presse VT et 69 fois dans le corpus des communiqués de presse (soit, environ deux fois plus dans les communiqués de presse (2,3 pour mille mots) que dans le corpus VT (1,1 pour mille mots)).

	<b>Le Monde</b>	<b>Les Échos</b>	<b>La Croix</b>	<b>Le Figaro</b>	<b>VT</b>	<b>CP CNES</b>
<b>1</b>	<i>navigation</i>	<i>radionavigation</i>	<i>navigation</i>	<i>navigation</i>	<i>navigation</i> 25,7%	<i>navigation</i> 16,7%
<b>2</b>	<i>radionavigation</i>	<i>navigation</i>	<i>positionnement</i>	<i>positionnement</i>	<i>radio-navigation</i> 20,3%	<i>télécommunication</i> 16,7%
<b>3</b>	<i>positionnement</i>	<i>télécommunication</i>	<i>télévision</i>	<i>localisation</i>	<i>positionnement</i> 10,5%	<i>observation de la terre</i> 7,8%
<b>4</b>	<i>observation</i>	<i>positionnement</i>	<i>localisation</i>	<i>télévision</i>	<i>télévision</i> 7%	<i>localisation</i> 7,8%
<b>5</b>	<i>télécommunication</i>	<i>télévision</i>	<i>radionavigation</i>	<i>télécommunication</i>	<i>télécommunication</i> 6,8%	<i>internet</i> 7,8%
<b>6</b>	<i>localisation</i>	<i>téléphone</i>	<i>guidage</i>	<i>téléphone</i>	<i>localisation</i> 4,6%	<i>positionnement</i> 6%

Tableau 5 : Groupes nominaux (GN) les plus fréquents dans la structure [GN par satellite]

Ce tableau montre que, dans tous les cas, *navigation* ou *radionavigation* apparaissent comme la fonctionnalité la plus présente dans les corpus. Ce constat peut permettre deux conclusions : d'une part, on remarque que cette fonctionnalité, très présente dans les communiqués est aussi parfaitement repérée dans la presse, et même plus : si l'on regarde la répartition d'un point de vue statistique, on s'aperçoit que 46% des occurrences de [GN *par satellite*] concernent la navigation ou la radionavigation dans la presse quotidienne et 20,1% dans les communiqués de presse. Nous sommes donc dans une sorte d'usage « en miroir » quant au premier rôle accordé à cette fonctionnalité, qui est certainement attribuable au développement du GPS. Cette hypothèse est d'autant plus probable que les GN *localisation* et *positionnement* sont également liés à cette technologie. D'autre part, cette homogénéité pourrait être un argument en faveur d'une interprétation de diffusion : le grand public semble être capable d'utiliser ces termes, et plus généralement, cette structure, en tout cas de les comprendre. Il y aurait donc une sorte de continuité dans le passage de l'usage scientifique à l'usage banalisé, *via* l'utilisation de ces termes par les communiqués de presse.

Si l'on s'intéresse à présent de manière globale aux 53 GN qui peuvent apparaître dans la structure [GN *par satellite*] (corpus CP et VT confondus), on se rend compte que seulement 22 sont utilisés dans les CP. Pourtant, parmi les 31 qui ne sont pas dans ces CP, 6 sont utilisés dans au moins trois des quotidiens de VT dont *bouquet*, *chaîne*, *détection*, *guidage*, *téléphonie*, *téléphone*. Si plusieurs hypothèses peuvent être proposées pour expliquer ce décalage, l'observation étendue de ces cas montre que ces termes semblent correspondre à différents degrés d'évolution et d'intégration des applications satellitaires auprès du grand public dans le temps : nous présentons trois cas d'évolution différents à travers les exemples

de *télévision par satellite*, *navigation par satellite* et les termes liés au thème de la médecine par satellite.

#### 4.3.1. Terme diffusé : *télévision par satellite*

Comme nous venons de le voir, les GN les plus fréquents dans les corpus sont tous liés à différents types de télécommunications, la télévision en tête. Cet exemple est particulièrement intéressant dans la mesure où il illustre un exemple d'intégration très clair d'une application du spatial auprès du grand public et d'appropriation du terme par le locuteur non spécialiste depuis quelques années déjà (et bien avant 1998, date de nos premiers corpus de presse). Cet aspect se manifeste linguistiquement par une succession d'ellipses à partir du terme *bouquet de chaînes de télévision par satellite*, qui aujourd'hui ne se retrouve repris dans nos corpus majoritairement que par les termes *bouquet de télévision satellite*, *bouquet satellitaire*, ou *bouquet satellite*.

#### 4.3.2. Terme en cours de diffusion : *navigation par satellite*

Le deuxième exemple est celui de *navigation* qui est, sur la période 2003-2011, la technologie satellitaire la plus représentée dans les corpus. Cet exemple est particulièrement riche et intéressant pour notre étude puisqu'il s'agit d'un terme/concept en pleine diffusion et évolution. Cette évolution est marquée dans nos corpus d'une part par une augmentation du nombre d'occurrences de ce terme et d'autre part, par une diversification de ses contextes d'apparition. Nous présentons l'analyse complète de ce cas dans la section 4.5. où nous détaillons la manière dont l'observatoire permet de saisir l'évolution sémantique en cours de ce terme.

#### 4.3.3. Termes aux prémices d'une diffusion : le thème de la médecine par satellite

Enfin, inversement, plusieurs termes utilisés dans les CP par le Cnes ne sont quasiment jamais repris dans les quotidiens. En effet, si les concepts de la navigation et de la télévision par satellite sont repris de manière significative, certaines autres fonctionnalités sont aussi présentes dans les CP mais inégalement représentées dans VT. C'est le cas par exemple des termes liés aux applications médicales par satellite. Dans les CP, 4 GN peuvent relever de cette catégorie : *dépistage des complications* (1), *télé-médecine* (2), *télé-santé* (2), *sauvetage* (3), soit 8 occurrences sur les 69 qui correspondent à la structure [GN *par satellite*] (11,6%). Or, cette catégorie reste encore très peu présente dans VT : seulement 6 occurrences sur 760 GN (0,79% ; *dépistage des complications* (1), *télé-médecine* (2), *échographie* (1) et *téléconsultation* (2)). Notons que certains quotidiens, tel que *Le Monde*, n'utilisent d'ailleurs aucun de ces syntagmes. Il semble donc qu'il y ait un décalage important entre les CP d'une part, et les quotidiens d'autre part, sur la thématique de la médecine par satellite, d'autant plus qu'aucune autre structure syntaxique n'a pu être rencontrée dans nos corpus pour exprimer ces concepts (telles que [le satellite X permet la téléconsultation] ou encore, les [échographies peuvent être réalisées grâce au satellite X] ou, et c'est une structure moins prévisible, qui semble spécifique de ce corpus : [satellite de GN] (avec GN = *télécommunication*, *localisation*, *observation de la terre*...)).

### 4.4. Comparer le mode de construction des termes complexes

Une autre piste de recherche, plus générale, est apparue, toujours à partir de l'étude des structures [GN *par satellite*]. Il s'agit du mode de construction des termes complexes, et en

particulier de la différence potentielle entre la construction des termes complexes dans les langues de spécialité et dans la langue générale. En effet, en travaillant sur la structure [GN par satellite], nous nous sommes aperçues qu'elle pouvait alterner avec la forme [GN satellite] par exemple:

*observation par satellite/observation satellite*  
*localisation par satellite/localisation satellite*  
*télécommunication par satellite/télécommunication satellite*

Cette alternance a attiré notre attention et nous avons voulu voir comment, plus généralement, se faisait la répartition des structures [GN prep GN] vs [GN GN] (par exemple : *station sol/station au sol, données satellite/données du satellite*). Le Tableau 6 rend compte de cette répartition. Notons que nous sommes parties des structures [GN GN] pour lesquelles nous avons fait l'hypothèse que la structure concurrente [GN prep GN] était possible.

	CP	VT
<i>NN</i>	120	171
<i>N prep N</i>	122	582
<b>Proportion de NN</b>	49.6%	22,7%

Tableau 6 : Répartition des variantes selon leurs modes de construction dans VT et CP

Les résultats obtenus permettent bien de mettre en évidence une tendance : alors que 4 fois sur 5, la langue générale utilise les termes GN prep GN, on ne les sélectionne qu'1 fois sur 2 seulement, dans les textes spécialisés.

L'ellipse du déterminant et de la préposition est souvent considérée comme un indice de terminologisation (Collet, 1997; Portelance, 1991). Dans le même ordre d'idée, pour la théorie des sous-langages, ces formes de « déviations » (Kittredge, Lehrberger, 1983), appelées ainsi car elles ne correspondent pas aux constructions attendues, font partie intégrante de la problématique. Mais aucune étude n'a, à notre connaissance, étudié le processus inverse, de rétablissement du déterminant et de la préposition, dans les cas de dé-terminologisation. Or, il semble bien que, de manière globale, lors du passage des termes dans le discours général, les prépositions soient rétablies.

Le chantier est ouvert car beaucoup de questions restent en suspens : certaines prépositions sont-elles particulièrement rétablies ? Inversement, d'autres ne sont-elles jamais rétablies ? Ne peut-il y avoir d'erreur dans la restitution de la préposition (*communication satellite* est-il l'ellipse de *communication par satellite* ou de *communication du satellite*)? Y'a-t-il des régularités en fonction des quotidiens ? Néanmoins, cet observatoire non seulement permet de mesurer cette tendance (aspect quantitatif) mais il permettra aussi d'approfondir cette problématique d'un point de vue qualitatif.

#### 4.5. Observer des changements sémantiques

Pour terminer, nous revenons sur le cas de *navigation satellite*. De manière encore plus fine en effet, il est possible avec cet observatoire de repérer des changements sémantiques dont l'origine repose précisément sur la diffusion des termes dans la langue générale. Ainsi, nous l'avons vue, le terme *navigation satellite* se diffuse largement dans la langue générale, notamment à travers le développement et la banalisation des GPS. Les modificateurs de *navigation* répertoriés à ce jour dans les ressources lexicographiques sont *maritime, aérienne* et *informatique*. Les combinaisons entre *navigation* et ces différents adjectifs méritent



toutefois un commentaire. En effet, si on perçoit la parenté sémantique entre *navigation* et *maritime* puis la métaphore, régulièrement attestée<sup>10</sup> entre l'eau et l'espace dans *navigation aérienne*, la création *navigation informatique* demande un effort supplémentaire pour considérer l'internet comme un lieu fluide, dans lequel on peut se déplacer (comme l'eau ou l'espace). Aujourd'hui, 4 contextes apparaissent, qui montrent une évolution sémantique émergente de ce paradigme vers l'ajout de la navigation *terrestre* dont par exemple :

« système de navigation des véhicules terrestres » (VT)

« nous conjugons les données GPS avec les données de la réflectométrie pour obtenir, notamment en matière de navigation terrestre, une précision jusque-là inégalée et aussi une plus grande fiabilité » (VT)

On peut penser que la création de *navigation terrestre* est le fruit d'un double processus : d'une part, une troncation dans le terme « navigation des véhicules terrestres » et, d'autre part, l'influence du paradigme *maritime, aérien, ... terrestre*.

Quoi qu'il en soit, la combinaison créée a pour effet d'augmenter la couverture sémantique de *navigation* dont le sens pourrait être, dans le contexte spatial : guidage par satellite de tout objet se déplaçant dans l'air, l'espace ou sur terre.

## 5. Éléments de discussion et remarques conclusives

Dans cet article, nous avons présenté un regard original de linguistique de corpus et de terminologie textuelle sur une question encore rarement systématiquement étudiée : la diffusion des termes dans la langue générale.

À partir d'une demande pratique du Centre national d'études spatiales, nous avons proposé une réflexion méthodologique sur les outils et les données envisageables pour analyser de manière complète et fine les phénomènes en jeu. Si les pistes que nous abordons à ce jour sont encore parfois embryonnaires, elles restent toutefois nombreuses et riches pour répondre à des besoins pratiques mais aussi pour mettre en œuvre des analyses encore inédites à ce jour, souvent faute de moyens. Nous l'avons illustré dans la section 4, de nombreux aspects, tant quantitatifs que qualitatifs peuvent être questionnés et offrent un portrait inédit et ouvert du passage des termes d'une langue de spécialité à la langue générale. Ces différentes raisons nous mènent, malgré l'imperfection de l'environnement que nous présentons, à proposer le terme d'*observatoire*.

D'un point de vue méthodologique, le défi que nous nous posons reste de taille et beaucoup de paramètres sont à affiner. Parmi les principaux soulevés dans cette recherche, nous en retiendrons trois à ce stade. Le premier concerne la pertinence du genre « presse » pour refléter les usages en langue générale. Comme nous l'avons dit, si sa représentativité a parfois été remise en cause, la presse n'en reste pas moins un genre souvent utilisé, disponible et en quantité, trois qualités encore rares dans d'autres genres pour analyser la langue générale. Mais au-delà de cet aspect, dans notre cas, ce sont surtout les masses de données qui nous posent question : les données disponibles sont-elles trop abondantes ? Limitent-elles la représentativité des données pour observer la diffusion des termes ? Ou alors, comment échantillonner, en gardant l'équilibre entre quantité et représentativité par rapport à notre problématique de diffusion ? Ces questions restent encore pour nous difficiles à trancher, mais doivent être prises en compte de manière centrale dans la suite de nos travaux.

---

<sup>10</sup> Voir par exemple *vaisseau spatial*.

Le deuxième est que s'il est clair que la présence des termes que nous avons observés est influencée par les activités du Cnes diffusées, elle peut l'être également par d'autres médias tels que des communiqués d'agences scientifiques majeures autres comme le CNRS ou l'ESA (European Space Agency), ou encore, obtenues par d'autres sources, traduites, etc. Il faudra donc veiller à retravailler le corpus CP pour le compléter par d'autres sources et améliorer ainsi sa représentativité.

Le troisième problème de taille dans notre environnement est un problème encore une fois relativement classique dans les approches sur (grands) corpus : celui de la désambiguïsation. En effet, nos corpus ne sont pas désambiguïsés et il est clair que nombre des termes que nous observons sont polysémiques (*espace, satellite, atmosphère, étoile*, etc.). De fait, nombre de contextes que nous obtenons ne sont pas pertinents. Cependant, notre perspective diachronique implique une nouvelle contrainte sur cette question : puisque le sens est susceptible d'évoluer dans le temps, dans quelle mesure la désambiguïsation systématique serait-elle pertinente et efficace? Nous l'avons vu, la diffusion des termes entraînent des changements de sens (et notamment des métaphores) que nous cherchons précisément à observer et que nous ne pouvons pas toujours anticiper. De fait, plusieurs tests devront également être mis en place pour en mesurer l'impact.

De manière générale, si nous croyons que notre observatoire pose des pistes originales et prometteuses pour l'analyse de la diffusion des termes, les ressources et les méthodes que nous impliquons doivent encore être questionnées et ajustées pour offrir une analyse mieux contrôlée. Cependant, nous restons convaincues que cet observatoire met à disposition un cadre méthodologique global et reproductible qui permet d'articuler analyses globales (quantitatives) et locales de tous ordres : linguistiques, lexicométriques, statistiques, etc.

Dans cet article, nous avons d'ores et déjà ouvert plusieurs pistes de travaux à explorer. D'un point de vue quantitatif, nous avons montré que la comparaison CP/VT montre tantôt une harmonisation, tantôt un décalage. Il est nécessaire alors d'interroger ces résultats en fonction d'événements qui ont pu avoir lieu dans le domaine de l'espace et de leur impact sur le grand public. Cet aspect du rôle de l'extra-linguistique devra (et pourra) être interrogé plus systématiquement dans son rôle sur les phénomènes de néologie. D'un point de vue qualitatif, plusieurs axes de recherche vont être approfondis. Tout d'abord, d'un point de vue global, le rétablissement de la préposition et du déterminant dans le passage des CP au corpus de presse, semble un indice assez net de dé-terminologisation, qui serait le pendant de la suppression de ces éléments dans le passage du général au spécifique. Mais, nous l'avons vu, cette hypothèse pose beaucoup de questions qui pourront être examinées grâce à l'observatoire que nous avons construit. Un autre phénomène qui a été mis au jour concerne la conjonction d'éléments d'explication pour justifier l'apparition d'un terme comme *navigation terrestre*. Nous pensons que nous pourrions trouver d'autres néologismes de ce type, ce qui contribuera à établir une typologie de la néologie liée au passage de la langue spécialisée à la langue générale. Enfin, sur un autre point (mais le propre de ce type de recherche exploratoire est de faire émerger des questions sur des plans que nous n'avions pas imaginés dans un premier temps), il est apparu une structure que l'on pourrait certainement considérer comme un marqueur de fonctionnalité : [GN1 par GN2] et son pendant [GN2 de GN1]. Dans la perspective de systématiser la recherche d'information de ce type, il sera nécessaire d'évaluer la pertinence de ce marqueur, dans les corpus du spatial mais aussi dans des corpus d'autres domaines.

D'un point de vue théorique et épistémologique, nous pensons que ce cadre est donc garant d'une observation non seulement plus systématique, reproductible et fiable, mais également inductive. En d'autres termes, cet observatoire permet de mettre au jour des phénomènes de diffusion pertinents et centraux à observer mais non connus ou soupçonnés *a priori*. Ceci permettra, nous l'espérons, d'aller vers une meilleure compréhension de ces phénomènes et surtout vers la mise au jour d'une typologie plus exhaustive et fine du passage des termes d'une langue de spécialité à la langue générale.

Et en effet, si cette question de la diffusion est parfois abordée aujourd'hui par les linguistes et terminologues, elle reste encore mal connue comme en témoigne l'hétérogénéité des dénominations qui la concernent. À l'aide de meilleurs outils, il sera alors possible de mieux saisir les liens entre *migration*, *diffusion*, *vulgarisation*, *déterminologisation* ou encore *néologie*, tous phénomènes à la frontière si difficile à saisir entre langue de spécialité et langue générale.

## Références

- ALEXANDRU C., GAUDIN, F. (2005), Les contextes à la source du terme, In *Mots, termes et contextes*, 59-67, Bruxelles, EAC-AUF.  
(<http://perso.univ-lyon2.fr/~thoiron/JS%20LTT%202005/pdf/Gaudin.pdf>, consultée le 31.10.2012)
- BIBER D. (1993), Representativeness in Corpus Design, *Literary and Linguist Computing*, Vol. 8(4), 243-257.
- CABRÉ M.T., DE YZAGUIRRE L. (1995), Stratégie pour la détection semi-automatique des néologismes de presse, *Traduction Terminologie Rédaction*, Vol. VIII(2), 89-100.
- CABRÉ M.T., NAZAR R. (2011), Towards a New Approach to the Study of Neology, *Neologica*, Vol. 6, 63-80.  
(<http://www.termisti.org/semin2011/cabrenazar.pdf>, consultée le 31.10.2012)
- 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*, 71-85, Paris, Université Paris VII, UFR EILA.
- COLLET T. (1997), La réduction des unités terminologiques complexes de type syntagmatique, *Meta : journal des traducteurs*, Vol. 42(1), 193-206.
- CONDAMINES A., DEHAUT N. (2011), Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie, *Meta : journal des traducteurs*, Vol. 56(2), 266-283.
- CONDAMINES A., 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*, 190-192, Paris, Gallimard.
- CONDAMINES A., DEHAUT N. PICTON A. (2012), Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique. Études outillées en corpus, In GÉRARD C., KATABEK J. (dirs.) *Néologie sémantique et analyse de corpus, Cahiers de Lexicologie*, Vol. 101, 161-184, Paris, Garnier Classiques.
- DROUIN P. (2003), Term Extraction Using non-Technical Corpora as a Point of Leverage, *Terminology*, Vol. 9(1), 99-117.
- DROUIN P., PAQUIN A., MÉNARD N. (2006), Extraction semi-automatique des néologismes dans la terminologie du terrorisme, In *Actes des 8<sup>e</sup> Journées internationales*

- d'Analyse statistique des Données Textuelles*, JADT 2006, en ligne <http://lexicometrica.univ-paris3.fr/jadt/jadt2006/PDF/033.pdf> (consultée le 31.10.2012)
- DURY P. (1997), *Étude comparative et diachronique de l'évolution*, Thèse de doctorat en Linguistique – Lexicologie et Terminologie multilingues – Traduction, Université Louis-Lumière, Lyon 2.
- DURY P. (1999) Les variations sémantiques en terminologie : étude diachronique et comparative appliquée à l'écologie, In DELAVIGNE V., BOUVERET M. (dirs.) *Sémantique des termes spécialisés*, 17-33, Rouen Publications de l'Université de Rouen.
- DURY P. (2004), Building a Bilingual Diachronic Corpus of Ecology: The Long Road to Completion, *Icame Journal*, Vol. 28, 5-16.
- DURY P. (2006), La dimension diachronique en terminologie et en traduction spécialisée : le cas de l'écologie, In GAUDIN F., CANDEL D. (dirs.), *Aspects diachroniques du vocabulaire*, 109-124, Rouen, Publications des Universités de Rouen et du Havre.
- GALISSON R. (1978), *Recherches de lexicologie descriptive, la banalisation lexicale*. Paris, Nathan.
- GENTILHOMME-KOTYRINE Y. (1994), Regards sur la terminologisation en lexicologie, *Meta : journal des traducteurs*, Vol. 39(4), 546-560.
- GOUADEC D. (1990), *Terminologie : Constitution des données*. Paris, AFNOR.
- GUILBERT L. (1975), *La créativité lexicale*, Paris, Larousse.
- HUMBLEY J. (2011), Vers une méthode de terminologie rétrospective, *Langages*, Vol. 183(3), 51-56.
- JACOBI D. (1986), *Diffusion et vulgarisation, itinéraires du texte scientifique*, Besançon, Presses Universitaires de Franche-Comté.
- JOSSELIN-LERAY A. (2005), *Place et rôle des terminologies dans les dictionnaires généraux unilingues et bilingues. Étude d'un domaine de spécialité : volcanologie*. Thèse de doctorat en Linguistique – Lexicologie et Terminologie multilingues – Traduction, Université Louis-Lumière, Lyon 2.
- KITTREDGE R., LEHRBERGER J. (dirs.) (1983), *Sublanguage: Studies of Language in Restricted Semantic Domains*, Berlin/New York, W. de Gruyter
- L'HOMME M.-C., VANDAELE S. (dirs.) (2007), *Lexicographie et terminologie. Compatibilité des modèles et des méthodes*, Ottawa, Presses de l'Université d'Ottawa.
- MANIEZ F., DURY P., ARLIN N., ROUGEMONT, C. (dirs.) (2008), *Corpus et dictionnaires de langues de spécialité*, Travaux du CRTT, Grenoble, Presses Universitaires de Grenoble.
- MATHIEU Y. (1998), Étude pour une extraction automatique de néologismes, In CLAS A., MEJRI S., BACCOUCHE T. (dirs.), *La Mémoire des mots*, Actes V<sup>e</sup> Journées Scientifiques du réseau Lexicologie, Terminologie, Traduction, Tunis, 25-27 septembre, 455-461, Montréal, Aupelf.
- MEILLET A. (1921), *Comment les mots changent de sens*, Année sociologique 1904-1905, édition de 1921, en ligne, ([http://fr.wikisource.org/wiki/Comment\\_les\\_mots\\_changent\\_de\\_sens](http://fr.wikisource.org/wiki/Comment_les_mots_changent_de_sens) (consultée le 31.10.2012).
- MEYER I., MACINTOSH K. (1999), 'L'étirement' du sens terminologique : aperçu du phénomène de la déterminologisation, In BÉJOINT P., THOIRON P. (dirs.), *Le sens en terminologie*, 198-217, Lyon, Presses Universitaires de Lyon.
- MOIRAND S. (2007), *Les discours de la presse quotidienne. Observer, analyser, comprendre*, Paris, Presses universitaires de France, Linguistique nouvelle.
- NICOLAE C., 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, 143-155, en ligne [http://www.licorn-ubs.com/jlc6/ACTES/Nicolae\\_JLC09.pdf](http://www.licorn-ubs.com/jlc6/ACTES/Nicolae_JLC09.pdf) (consultée le 31.10.2012).
- NYCKEES V. (2000), Changement de sens et déterminisme socio-culturel, In FRANÇOIS J. (dir.), *Théories contemporaines du changement sémantique*, 31-58, Leuven, Peeters.
- PEARSON J. (1998), *Terms in Context*, Amsterdam/Philadelphie, John Benjamins
- 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.
- PORTELANCE C. (1991), Fondements linguistiques de la terminologie, *Meta : journal des traducteurs*, Vol. 36(1), 64-70.
- PUTNAM H. (1990), *Représentation et réalité*, Paris, Éditions Gallimard.
- RASTIER F., VALETTE M. (2009), De la polysémie à la néosémie, *Le français moderne*, MEJRI S. (dir.), *La problématique du mot*, Vol. 77, 97-116.
- RENOUF A. (1994), Corpora and Historical Dictionaries, In LANCASHIRE I., RUSSON WOOLDRIDGE T. (dirs.), *Early Dictionary Databases*, Center for Computing in the Humanities, 219-235, Toronto, University of Toronto.
- 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.
- TOLÉDANO V., CANDEL D. (2006), Mouvance terminologique, l'espace d'une courte diachronie, dans les dictionnaires et dans la presse : le cas de "mondialisation-globalisation", In CANDEL D., GAUDIN F. (dirs.), *Aspects diachroniques du vocabulaire*, 157-183, Rouen, Presses Universitaires de Rouen et du Havre.
- UNGURANEU L. (2003), *L'interpénétration langue générale-langue spécialisée dans le discours d'internet*, Thèse de doctorat, Université Paris 13/Université Technique de Moldova.
- 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> (consultée le 31.10.2012)
- WIERZBICKA A. (1985), *Lexicography and conceptual analysis*, Ann Arbor: Karoma.

Cette étude a été menée dans le cadre d'une convention avec le Centre National d'Etudes Spatiales que nous remercions pour la confiance dont il nous a témoigné.