Variation de la Terminologie dans le Temps : une Méthode Linguistique pour Mesurer l'Évolution de la Connaissance en Corpus[1]

Anne Condamines, Josette Rebeyrolle, Anny Soubeille
Equipe de Recherches en Syntaxe et Sémantique (ERSS)
CNRS et Université Toulouse Le Mirail
5 allées Antonio Machado
F-31058 Toulouse cedex
anne.condamines, josette.rebeyrolle@univ-tlse2.fr, anny.soubeille@wanadoo.fr

Résumé
L'article présente une étude qui porte sur la variation temporelle de la terminologie. Partant du problème spécifique de l'évolution des connaissances dans les projets spatiaux du CNES, cette étude vise à montrer que cette évolution peut être repérée en s'appuyant sur les variations qui affectent la terminologie d'un domaine. Le repérage de l'évolution des connaissances repose d'abord sur la constitution d'un corpus pertinent pour étudier le phénomène de la variation dans le temps. Nous montrons en particulier pourquoi il est nécessaire de contrôler les situations extra-linguistiques pour organiser les textes sous la forme d'au moins deux sous-corpus qui ne varient que du seul point de vue de l'époque de leur rédaction. Partant de l'idée que dans les langues spécialisées, comme dans la langue, c'est le lexique qui évolue le plus rapidement, nous proposons trois angles d'observation de la variation terminologique : la forme des termes, leur distribution et leur fonctionnement sémantique. Ces trois types d'indices serviront de base à la comparaison des deux sous-corpus. Un changement de fonctionnement d'un corpus à l'autre est interprétée comme susceptible d'être un indice d'évolution des connaissances.

1. Introduction

Comme le rappelle Rousseau (2000), l'idée que les changements qui affectent le monde (évolution des connaissances, découvertes scientifiques, développement des techniques, ...) se répercutent dans les discours sous la forme de changements de sens est déjà présente chez le psychologue allemand Wund dans un livre paru en 1900. Selon le point de vue théorique que l'on adopte, on expliquera ces changements sémantiques de diverses manières : d'un point de vue onomasiologique, on s'appuie sur le principe unificateur de l'analogie pour expliquer les changements qui affectent l'organisation sémantique du lexique (Ullmann, 1969, Blank, 1999), d'un point de vue sémasiologique, on explique le changement de sens par un changement de désignation réalisable par différents processus lexicaux, tels que les formations de mots, les phraséologismes, etc. (Koch, 2000).
L'objet de cette contribution est d'aborder la question du changement sémantique sous un angle à la fois proche et différent de celui qui caractérise les travaux qui se situent dans le champ de la linguistique diachronique. Le point de vue que l'on adopte ici est en effet circonscrit par un problème appliqué posé par le Centre National d'Etudes Spatiales (CNES, Toulouse, France) (Condamines et *al.* 2003). Ce problème concerne les projets spatiaux de longue durée, c'est-à-dire des projets dont la durée dépasse dix ans. Il s'agit le plus souvent d'envoyer des sondes vers un objectif lointain, ne pouvant être atteint qu'au terme d'un

voyage de plusieurs années. Or, dans ce type de mission, les spécialistes qui ont participé à l'élaboration des instruments (satellites, sondes, etc.) ne seront pas ceux qui, dix ou vingt ans plus tard, effectueront les opérations prévues. Dans ces projets spatiaux, qui restent en sommeil pendant des années (seul un contrôle mensuel étant effectué), il existe donc un risque de perte des connaissances qui ont été mobilisées lors des phases de conception. De plus, tout en sachant que les connaissances évoluent, on sait que cette évolution se fait souvent à l'insu des acteurs d'un projet, soit parce que ces acteurs eux-mêmes changent (départ à la retraite, déplacement des ingénieurs sur d'autres projets, ...), soit parce que, pris dans la dynamique du projet, ils n'ont pas conscience de changements et/ou, en perdent la trace[2]. Devant ce risque majeur de perte de connaissances, le CNES est à la recherche de méthodes qui lui permettent de repérer les évolutions ou les ruptures de connaissances afin de proposer aux ingénieurs des moyens d'accéder à des fonctionnements pouvant être interprétés comme des indices d'un changement potentiel.
Etant donné ce contexte opératoire, notre problématique se laisse définir ainsi : nous travaillons dans une diachronie restreinte, autrement dit sur quelques années seulement contrairement aux études diachroniques « en langue ». Mais nous espérons de cette situation très particulière qu'elle nous permette de comprendre des phénomènes qui, se déroulant sur plusieurs dizaines voire plusieurs centaines d'années dans le cadre d'une langue, sont souvent difficiles à identifier et à décrire.

2. Observation de variations sémantiques en corpus spécialisé

Afin d'observer les changements de sens qui s'opèrent dans le cadre que nous venons de situer, une analyse de corpus s'impose pour une raison essentielle : étant donné que les concepteurs des instruments embarqués sur des projets spatiaux de longue durée auront disparu au moment où un certain nombre d'actions devront être accomplies, si aucune solution traditionnelle (comme la formation interne) n'est mise en place, la documentation produite sera la seule trace accessible de la connaissance. En conséquence, une méthode linguistique d'observation des évolutions de connaissances s'appuiera sur la documentation écrite qui les accompagne. Une question qui se pose alors immédiatement est celle de savoir quelles sont les propriétés que doit satisfaire un ensemble de textes (*i.e*, un corpus) pour mettre au jour ces évolutions. Pour y répondre, nous avons pris en compte trois types d'exigences :
1) une exigence d'homogénéité : les textes qui composent le corpus doivent être significatifs du point de vue du CNES, plus précisément ils doivent concerner un seul et même projet et contenir des textes relevant d'un même genre textuel[3] ;
2) une exigence de diachronicité : les textes qui composent le corpus devront nécessairement s'échelonner dans le temps afin de rendre possible l'observation de continuités, de ruptures et/ou d'évolutions des connaissances ;
3) une exigence de contrastivité : l'une des façons d'observer l'évolution étant de se fonder sur des comparaisons, il est crucial de partitionner les textes du corpus en plusieurs groupes (au moins deux) en s'appuyant sur des critères externes (Habert et *al.*, 1999) pertinents pour le projet visé. Afin de mettre en relation des fonctionnements linguistiques avec une évolution diachronique, les textes sélectionnés doivent être échelonnés dans le temps et sur une période la plus étendue possible.

Le corpus constitué pour mettre au point la méthode d'analyse a vu ces trois critères satisfaits dans le cadre du projet Doris (Doppler Orbitography and Radiolocation Integrated Satellite)[4]. Il s'agit d'un projet qui a commencé à être développé depuis une dizaine d'années et qui comporte trois types d'éléments : 50 balises au sol, un système de récepteurs à bord de satellites, un centre de traitement des données. Nous avons travaillé sur les balises, qui se sont développées sur trois générations en fonction d'évolutions techniques (parmi lesquelles, la miniaturisation de l'électronique qui a permis de réduire le poids et le volume des balises) : la première a été développée en 1984, la deuxième, de 1996 à 1999 et la troisième, à partir de 2000. A cela s'ajoute un élément capital. On dispose pour ce projet d'un ensemble de locuteurs compétents, autrement dit de spécialistes de l'instrument, aisément accessibles et disponibles. Le corpus constitué est organisé en deux sous-corpus (soit deux groupes de textes d'environ 16 000 mots chacun) : l'un correspond aux première et deuxième générations de balise DORIS (appelé G1_2) et l'autre à la troisième génération de balises (appelé G3).

3. Des changements de forme aux évolutions de connaissances

Les procédés de formation de nouveaux items lexicaux sont bien connus et décrits dans divers cadres théoriques de la sémantique diachronique (Rousseau, 2000 ; Koch, 2000; Blank, 1999) ou sur la grammaticalisation (Traugott & Dasher, 2002). Dans un autre champ, certains comme Teubert (2001) ou Belica (1996), considèrent que l'un des objectifs de la sémantique de corpus est d'étudier les notions de stabilité / changement (diachronique et/ou synchronique). En terminologie, en revanche, cette question de l'évolution de la forme et du sens n'est que rarement posée (voir toutefois Bonnet, 2003). En effet, la terminologie se concentre souvent sur un objectif de normalisation qui fige les fonctionnements à un instant *t* sans que les évolutions possibles soient prévues ou même envisagées.
Notre objectif est donc le suivant : identifier et décrire les formes privilégiées du changement sémantique à partir de l'analyse linguistique d'un corpus spécialisé construit de façon à rendre possible l'observation d'évolutions de connaissances.

3.1 Comment mesurer ces évolutions ?

L'approche que nous proposons allie méthode automatique et analyse linguistique. Plus précisément, il s'agit, d'une part, de formuler des requêtes automatiques indépendantes d'un corpus particulier en visant la réutilisabilité sur de nouveaux corpus constitués avec les mêmes critères et, d'autre part, de définir des modes d'interprétation linguistique des résultats obtenus.
Notre méthode est donc une méthode *outillée* faisant intervenir trois types d'outils : un extracteur de termes, Nomino (David & Plante, 1996), un étiqueteur grammatical (Cordial Université[5]), un concordancier, Yakwa (Rebeyrolle & Tanguy, 2000).
Les résultats de ces outils sont utilisés par des programmes PERL, conçus pour aider à identifier les variations que nous décrivons dans la section suivante.

3.2 Quels sont les indices qui manifestent ces évolutions ?

Nous reprenons à notre compte l'hypothèse de travail de Nyckees (2000) : "Détecter un changement de sens, c'est détecter un changement de règles d'usage au travers des énoncés produits".

3.2.1 Variations de la forme des termes

Partant de l'hypothèse que des changements affectant la forme des dénominations sont le reflet de changements touchant à leur contenu, nous accordons une large place aux variations morphologiques, qu'il faut entendre à la fois comme variations repérables d'un point de vue informatique et signifiantes d'un point de vue linguistique, par interprétation des résultats fournis par les outils.

Quatre modes d'évolutions morphologiques sont potentiellement intéressants :

- L'apparition ou disparition de formes : le terme *afficheur* disparaît dans les documents concernant la troisième génération consécutivement à la disparition de ces 'objets' qui sont remplacés par de simples écrans d'ordinateur.
- La composition qui est un procédé très productif en terminologie qui se manifeste pour de nombreuses formes par une perte de leur autonomie lorsqu'elle s'associent avec d'autres formes elles aussi jusque là autonomes, par exemple dans le corpus Doris : *synchronisation* et *mode -> mode de synchronisation.*
- La formation d'ellipses (que Koch (2000), reprenant Ullmann, explique par les besoins communicatifs) consécutive ou non au figement d'un composé est également très répandue : *horloge de la balise > horloge balise, mode de survie > mode survie.*
- L'expansion qui se manifeste le plus souvent par l'ajout d'un modifieur à un nom : *mode de fonctionnement > mode de fonctionnement secouru.*

Tous ces modes de changements peuvent être significatifs d'une évolution. Les apparitions/disparitions peuvent correspondre à des apparitions/disparitions de concepts mais elles peuvent aussi être l'indice d'une évolution dans la dénomination des concepts. La disparition d'expansion peut correspondre à une stabilisation du concept que, pour des raisons d'économie, on ne dénomme plus que par une forme écourtée, avec le risque d'ambiguïté que cela peut entraîner. L'expansion quant à elle pourrait correspondre soit au développement d'un concept existant qui s'affine, se spécialise ce qui oblige à le détailler et à le préciser, soit à la nomination d'une nouvelle technique. Avant de s'imposer, la nouvelle forme se trouve généralement en concurrence avec celle qui a été supplantée : dans le corpus Doris, des deux formes *fonctionner en mode* et *être en mode* une seule demeure *être en mode*, de même *fonctionner* et *être en fonctionnement* coexistent pour ensuite ne laisser place qu'à *être en fonctionnement.*

3.2.2 Variations de la distribution des termes

Les changements du comportement distributionnel des termes sont généralement le reflet de changement de sens (Harris et *al.*, 1989, Rousseau, 2000). Plus précisément, on peut interpréter les variations de constructions syntaxiques comme des indices de conceptualisations différentes : le terme *opérateur* passe ainsi de la position sujet de verbe d'action du type *appuyer*, *taper*, *valider* à la position complément dans des structures du

type : '*écran/interface*$_{[sujet]}$ + *permettre à* + *opérateur*$_{[complément]}$'. Deux types de distributions peuvent apparaître, les unes sont identifiées par l'étude, elles sont propres au corpus et ne sont donc pas décrites *a priori* ; les autres font intervenir une connaissance *a priori* qui permet d'attribuer à certains contextes un rôle particulier de marqueurs de relations conceptuelles.

Distributions identifiées par l'analyse

L'analyse distributionnelle relève dans ce cas-là principalement d'une interprétation linguistique. En effet, il s'agit de constituer des classes dont on fait l'hypothèse qu'elles correspondent au même type de fonctionnement sémantique. Cette catégorisation est souvent difficile à faire. Nous nous sommes donné comme guide la nécessité, pour pouvoir considérer que l'on a à faire à une classe, de pouvoir élaborer un schéma sémantico-syntaxique qui rende compte de cette classe, comme dans le cas des deux modes de fonctionnement d'*opérateur* décrits ci-dessus.
Dans une étude antérieure (Condamines & Rebeyrolle, 1997), nous avions interprété ce type de variation comme étant le reflet de points de vue (différents mais synchroniques), de groupes de locuteurs, sur les concepts auxquels renvoient ces termes. Dans le cas qui nous préoccupe ici, ces points de vue seraient dus à des évolutions dans le temps[6] ; il s'agirait donc de points de vue diachroniques.

Distributions caractérisées avant l'analyse

Les contextes dont il est question permettent de construire des réseaux terminologiques. Il s'agit de marqueurs de relation du type :

```
[dét déf + N1 + Vêtre + dét indéf N2 + relative]
```

(ou « dét déf » désigne un déterminant de la classe des définis et « dét indéf », un déterminant de la classe des indéfinis) qui permettent d'identifier une relation hyperonymique entre N2 et N1 (Condamines & Rebeyrolle, 2002). On peut faire l'hypothèse qu'un changement dans un réseau de relations est un indice d'une évolution sémantique. Pour tester cette hypothèse, il s'agit de construire des réseaux de termes pour chacun des sous-corpus et de comparer ces réseaux afin de repérer des changements. Mais, pour permettre la comparaison totale de réseaux, une grande quantité de données serait nécessaire. Les corpus dont nous disposons étant peu volumineux, ce mode d'exploration ne donne pas des résultats très nombreux. Dans cette étude, nous nous sommes plutôt focalisées sur les contextes non interprétés *a priori*.

4. Résultats

Les résultats présentés ici portent sur les éléments suivants :

- Résultats quantitatifs permettant de repérer les noms et groupes nominaux qui apparaissent *vs* disparaissent d'un sous-corpus à l'autre (cas qualifié précédemment d'apparition/disparition de formes).
- Résultats quantitatifs et « qualitatifs » pour les noms et groupes nominaux communs aux deux corpus. Considérant que seuls les termes qui apparaissent de manière significativement plus importante (ou moins importante) dans l'un ou l'autre corpus devaient faire l'objet d'une étude détaillée, nous avons construit un test statistique qui

permet de mesurer cet écart. Précisément, ce test permet de mesurer la pertinence des écarts d'emplois des termes communs aux deux sous-corpus. Seuls, ces termes communs significatifs ont fait l'objet d'une analyse linguistique fine de leur fonctionnement en contexte portant sur l'expansion et la distribution.

4.1 Résultats quantitatifs

Le dénombrement des noms et des groupes nominaux a été fait avec le logiciel Nomino. Les résultats proposés par ce logiciel ont ensuite fait l'objet d'un nettoyage avec un programme PERL (il s'agissait de supprimer des formes numériques considérées comme des noms, de « récupérer » des noms en majuscule considérés comme des noms propres, de normaliser des formes accentuées vs non-accentuées considérées comme deux formes distinctes...).
Les tableaux ci-dessous rendent comptent des résultats comparés dans les deux sous-corpus.

		G1-2		G3	
Noms	formes	684	48%	793	37%
	occurrences	5358		7002	
Groupes nominaux	formes	755	**52%**	1325	**63%**
	occurrences	1630		2521	
Nbre total de formes		1439	100%	2118	100%

Tableau 1: nombre de noms et des groupes nominaux dans les deux sous-corpus

	Seulement en G1-2 Nbre de formes	Seulement en G3 Nbre de formes	%	Communes à G1_2 et G3 Nbre de formes	%	Total
Noms	357	466	41%	327	**28%**	1150 100 %
Groupes nominaux	659	1229	63%	96	**4 %**	1984 100 %

Tableau 2: répartition des noms et des groupes nominaux dans les deux sous-corpus

NB : le nombre total de formes (N et GN) n'est pas égal à la somme des formes présentées dans le tableau précédent étant donné qu'apparaissent ici les formes communes aux deux sous-corpus que l'on ne recompte pas deux fois.
Ces résultats appellent les commentaires suivants :

- La proportion de groupes nominaux est plus élevée que celle des noms (respectivement, 52% pour G1_2 et 63% pour G3). Cette situation est habituelle dans les corpus spécialisés. On peut noter toutefois que la proportion de groupes nominaux augmente dans le corpus le plus récent. On peut interpréter cette augmentation comme étant le signe d'un accroissement de la précision des termes (plus la couverture sémantique des noms se précise, plus le nombre de modifieurs augmente).
- Le nombre de noms communs aux deux sous-corpus (327, près de 30% de l'ensemble des noms des deux corpus) et surtout de groupes nominaux (96, soit 4% seulement de la somme totale des groupes nominaux (1984)) est peu élevé. Autrement dit, 70% des noms ont disparu ou sont apparus d'un sous-corpus à l'autre et 96% des groupes nominaux. On

peut en conclure que beaucoup de groupes nominaux comportent des noms communs aux deux corpus mais avec des expansions différentes.

4.2 Résultats qualitatifs sur les noms et groupes nominaux communs aux deux sous-corpus

Le test statistique nous permet d'isoler, parmi l'ensemble des noms communs aux deux sous-corpus, les 116 noms et 16 groupes nominaux dont le fonctionnement est particulièrement significatif. Dans cet ensemble, nous avons retenu uniquement ceux qui présentent au moins trois occurrences dans chacun des corpus, en considérant que pour pouvoir caractériser le fonctionnement en contexte de ces unités, on ne peut se contenter d'un hapax ni de deux occurrences. Avant de présenter les résultats, il convient de préciser un certain nombre d'éléments.

Pour les groupes nominaux, nous avons travaillé sur la tête et sa distribution (par exemple, pour *spécification technique*, nous avons travaillé sur le terme *spécification)*. En effet, il nous a semblé préférable de déconstruire les groupes (identifiés automatiquement, rappelons-le) pour mieux travailler le processus de variation en contexte.

Nous nous sommes particulièrement intéressées aux fonctionnements suivants : expansion, distribution.

Expansion : nous avons considéré comme expansion un groupe nominal qui peut être considéré comme un nom suivi ou précédé d'un modifieur et ce, même si ce nom n'apparaît pas seul dans un des deux corpus (par exemple, *autotest séquenceur* est une expansion de *autotest*, *carte de commande* est une expansion de *carte*, *autorisation d'émission sur un satellite* est une expansion d'*autorisation*). Une comparaison entre les deux sous-corpus permet de repérer si l'expansion est stable ou différente d'un corpus à l'autre.

Distribution : la distribution concerne tous les contextes, droits ou gauches qui n'ont pas déjà été considérés dans le cadre de l'expansion. D'une certaine façon, l'expansion concerne les relations à l'intérieur d'un syntagme et la distribution les relations entre syntagmes.

Par ailleurs, ainsi que nous l'avons déjà souligné, nous recherchons une modélisation de ces contextes. Il s'agit en effet de comparer les distributions et cela n'est réellement possible qu'entre des abstractions de ces distributions (en particulier, mais pas seulement, en raison de la petite quantité de données à évaluer). Par exemple, dans le premier sous-corpus, *autotest* apparaît dans deux structures : [autotest déceler dét N], avec N = *anomalie, panne, défaillance*...et [prep temporelle autotest] :

Si l'autotest n'a décelé aucune anomalie lors du test de télémesure...
Le signal SI est généré uniquement pendant l'autotest manuel de la balise

Cette modélisation nous permet de dire que dans le premier cas, *autotest* est considéré comme un acteur et dans le second comme un processus. Dans le second corpus, *autotest* apparaît dans la structure [prep temporelle N] mais pas dans la structure [autotest déceler dét N]. En revanche, il apparaît dans la structure suivante : [sanction de l'autotest] :

Paramètres : - [...], - sanction de l'autotest

Considérant que seul un acteur (assimilé à un humain) pouvait sanctionner, nous avons fait le choix de considérer que la présence de ce nouveau contexte n'était pas le signe d'une évolution puisque la catégorisation abstraite (en tant qu'acteur) avait déjà été repérée. Nous

avons donc considéré que la distribution d'*autotest* était à peu près équivalente dans les deux sous-corpus.

En fonction de ces deux critères, expansion et distribution, les formes peuvent être versées dans quatre configurations :

1. soit l'expansion est différente dans les deux corpus et la distribution est identique,
2. soit l'expansion est identique et la distribution est différente,
3. soit l'expansion et la distribution sont différentes,
4. soit l'expansion et la distribution sont similaires.

Nous donnons et commentons ci-dessous les 72 noms que nous avons finalement examinés, du point de vue de leur expansion et de leur distribution dans les deux sous-corpus.

Configuration 1 : expansion différente et distribution identique (14 noms)
autotest, connecteur, délai, Doris, durée, énergie, gradient, instant, liaison, nombre, secteur, spécification, test, transfert
Dans ce cas, où seule l'expansion est différente, on peut faire l'hypothèse que les concepts ont évolué vers un affinement, vers une précision. La distribution restant stable, on peut penser, en revanche, que cela s'explique par une stabilité de la fonction des concepts.
Par exemple, dans le passage d'un sous-corpus à l'autre un terme comme *liaison* voit son extension se modifier puisqu'on passe de *liaison* à *liaison directe, liaison distante, liaison filaire, liaison descendante, liaison de télégestion.* Cependant, la distribution ne varie pas, puisqu'il s'agit toujours de *réaliser la liaison*, ou d'*établir la liaison.*
Configuration 2 : expansion identique et distribution différente (10 noms)
action, amplificateur, bord, courant, face, intermédiaire, recalage, résolution, satellite, tension
Dans ce cas, où seule la distribution a évolué, l'évolution s'est sans doute faite vers une couverture sémantique plus grande du concept initial (avec potentiellement création de polysémie).
Action par exemple a peu évolué du point de vue de l'expansion ; ainsi, il est utilisé très souvent seul dans les deux sous-corpus. En revanche, il apparaît dans des contextes assez différents. Dans le premier sous-corpus, un des contextes, particulièrement fréquent, peut être modélisé comme [déverbal V dét par dét action sut dét N] (*tout contrôle doit se terminer par une action sur la touche VAL*) alors qu'un des contextes caractéristique du second corpus peut être schématisé par la structure [action Vparticipe passé par dét N] (*lecture par l'opérateur des actions réalisées par le terminal*). Cette évolution semble le signe que les actions sont plutôt réalisées par des humains dans le premier sous-corpus et par des outils dans le second.
Configuration 3 : expansion et distribution différentes (31 noms)
anomalie, autorisation, BM, carte, circuit, commande, date, données, émission, environnement, essai, fonction, format, FOUS, gestion, identification, interface, jour, ligne, longueur, marche, Max, mesure, module, mot, opération, panne, période, point, synchronisation, température
Dans ce cas, l'expansion et la distribution ont évolué de concert. On peut s'attendre ici à trouver des changements majeurs pour le domaine. C'est d'ailleurs le cas le plus fréquent puisque près de la moitié des formes que nous avons analysé présentent ce type d'évolution.

Pour illustrer cette configuration, on citera le cas du nom *essai*. D'un sens exclusif de processus dans le premier sous-corpus, on passe à un sens de document dans le second. Le sens de processus est toujours présent (*durée de l'essai*)mais il se trouve clairement supplanté par le sens de « document qui consigne les résultats obtenus lors des essais ». On passe en effet de constructions comme *avant essai* ou *après essai* à *plan d'essai, documentation d'essai, revue d'essai*. On notera enfin que dans le premier sous-corpus, le terme *essai* se présente sans modifieur alors que les expansions sont très nombreuses dans la seconde partie, comme le montrent : *essai de qualification, essai de recette, essai de performances, essai de gradient thermique*, etc.

Configuration 4 : expansion et distribution identiques (17 noms)

coffret, embase, figure, fin, gamme, humidité, indicateur, intérieur, lieu, passage, place, position, pression, protection, puissance, synchro, valeur

Dans les cas où les deux fonctionnements sont similaires, on peut considérer qu'il y a une grande stabilité d'usage dans les deux corpus et, qu'en principe, le concept n'a pas évolué. Il s'agit dans ce cas de concepts stables pour le domaine ou en tous cas soumis à une évolution moins rapide. C'est le cas par exemple de *indicateur* qui conserve son sens comme le montrent à la fois ses expansions, *indicateur de verrouillage*, notamment et aussi sa distribution en complément de verbes comme *fournir* dans *fournir un indicateur*, par exemple.

5. Discussion des résultats et conclusion

Au stade où nous en sommes, deux types d'interprétation peuvent être proposés, l'une concerne le point de vue de la lexicologie, l'autre le point de vue du CNES.

5.1 Point de vue de la lexicologie

A notre connaissance, aucune étude poussée sur corpus n'a été menée pour étudier en profondeur le mode d'évolution des termes dans le temps. Les résultats que nous proposons ne sont évidemment pas définitifs mais ils donnent une première base et permettent de dessiner des tendances. Ainsi, il apparaît que lorsque, statistiquement, des mots apparaissent comme nettement différents d'un corpus à l'autre, dans la plupart des cas (43 %), ces mots évoluent à la fois du point de vue de l'expansion et de la distribution. Peu sont stables du point de vue de la distribution et de l'expansion (24 %). Quant à ceux qui évoluent soit du point de vue de l'expansion, soit du point de vue de la distribution, ils sont aussi peu nombreux, respectivement, 19 % et 14 %.

Ces résultats mériteraient d'être encore travaillés. Il faudrait poursuivre l'exploration du point de vue du fonctionnement des termes (par exemple, il serait intéressant de comparer les fonctionnements de *essai* et *test*, *a priori* synonymes mais qui n'appartiennent pas à la même catégorie d'évolution) . Il faudrait ensuite tenir compte de l'évaluation de l'expert (*cf.* ci-dessous) afin de voir si les évolutions qu'il qualifie d'intéressantes peuvent être corrélées avec des fonctionnements en contexte (rappelons que seront intéressantes pour l'expert des évolutions dont il n'aurait pas été conscient par opposition aux évolutions qui relèvent d'un choix (de nouvelles dénominations, par exemple) ou qui relèvent de la disparition d'un objet technique. Enfin, il serait nécessaire de refaire l'étude sur un autre corpus. Rien ne garantit en effet que la répartition des mots dans les quatre configurations soit stable pour un autre

corpus. Seule une étude réalisée sur un nouveau corpus, voire sur de nouveaux corpus permettrait de le dire. En tous cas, cette étude correspond à une première étape d'un projet de plus grande envergure sur le mode d'évolution des termes dans un corpus spécialisé.

5.2 Point de vue du CNES

Les résultats que nous avons obtenus ont permis de dresser une première liste des termes qui, à la fois du point de vue de leur expansion et du point de vue de leur distribution, manifestent une évolution importante. Cette sélection très fine nous permet de solliciter les experts avec une liste de départ contenant peu de termes . Cette situation est très différente de celle, classique, qui consiste à demander à un expert de valider des listes de termes, souvent très longues : c'est souvent un travail très fastidieux pour l'expert. Par ailleurs, dans ce cas précis, il ne s'agit pas de valider des termes mais de donner un avis sur les éléments qui peuvent expliquer un changement de comportement d'un corpus à l'autre. Avec un tel point de vue, la liste de départ sert plutôt à solliciter la réflexion de l'expert qui va essayer de mettre en place une explication globale (et pas pour chacun des termes les uns après les autres) des évolutions. Il est donc important que cette première liste ait été très travaillée en amont afin qu'elle corresponde à des termes qu'un faisceau d'indices (pertinence statistique, expansion, distribution) désignent comme manifestant une évolution majeure. Cette analyse est assistée, autant que faire se peut, par des outils mais la catégorisation des contextes ne peut être réellement faite que par des humains, en l'occurrence des linguistes, qui ont l'avantage d'être extérieurs au CNES et donc d'avoir un regard « neutre ».
Il reste à travailler réellement avec l'expert pour étudier comment il réagit par rapport aux résultats qu'on lui propose et à voir comment, en fonction de ces résultats, la méthode pourrait être validée.

Notes

1. L'article présenté a été réalisé dans le cadre d'un projet Recherche et Développement financé par le CNES : « Méthodes et Outils de Data Mining pour le projets spatiaux » qui fait intervenir l'INRIA (Institut Nationale de la Recherche en Informatique et en Automatique), l'ERSS, l'IRIT (Institut de Recherches en Informatique de Toulouse) et le LSP (Laboratoire de Statistiques et probabilités).
2. On prétend ainsi que les américains ne seraient plus en mesure de mobiliser les connaissances pertinentes pour aller sur la lune...
3. La notion de genre textuel est prise au sens de Bakhtine (Bakhtine, 1984). Le genre textuel permet souvent à lui seul de lever certaines ambiguïtés. Par exemple, le terme 'durée de vie' a deux emplois nettement distincts selon qu'il est utilisé dans un document de spécifications techniques, où il s'agit de la durée de vie de l'instrument à spécifier, ou dans un document de spécifications de management, où il désigne la durée de vie du projet.
4. Nous remercions M. Escudié, chef de la Division Altimétrie, qui a mis les corpus à notre disposition et a accepté de valider les résultats.
5. Distribué par la société Synapse Développement : http://www.synapse-fr.com
6. Ce point souligne s'il en était besoin encore la nécessité de constituer un corpus sur la base de critères très contrôlés, qui permettent d'associer les évolutions détectées à une variation diachronique.

Références

Bakhtine M. 1984. *Esthétique de la création verbale.* Paris : Gallimard, Tel.

Belica, C. 1996. Analysis of temporal changes in corpora. *International Journal of Corpus Linguistics,* 1 (1), pp.61-73.

Biber, D. 1990. Methodological issues regarding corpus-based analyses of linguistic variation. *Literary and linguistic computing,* 5 (4), pp. 257-270.

Blank, A. 1999. Why do new meanings occur? A cognitive typology of motivations for semantic change. In P. Koch (Ed.), *Historical semantics and cognition* (pp. 61-89). Berlin/New York: Mouton de Gruyter.

Bonnet, V. 2003. Pour une terminologie diachronique. *Les Travaux du CERLICO* (16), pp. 27-47.

Condamines, A. & Rebeyrolle, J. 1997. Point de vue en langue spécialisée. *Meta : Journal des traducteurs,* 42 (1), pp.174-184.

Condamines, A. & Rebeyrolle, J. 2002. Searching for and identifying conceptual relationships via a Corpus-based approach to a Terminological Knowledge Base (CTKB): Method and Results. In M.-C. L'homme, C. Jacquemin & D. Bourigault (Eds.), *Recent Advances in Computational Terminology.* Amsterdam/Philadelphia: John Benjamins Publishing Company, pp.127-148.

David, S. & Plante, P. 1990. De la nécessité d'une approche morpho-syntaxique dans l'analyse de textes. *Intelligence artificielle et sciences cognitives au Québec,* 3 (3), pp.140-154.

Habert, B., Folch, H. & Illouz, G. 1999. Sortir des sens uniques : repérer les mots "mouvants" dans le domaine social. *Sémiotiques* (17), pp.121-152.

Harris, Z.S., Gottfried, M., Ryckman, T., Mattick, J., Daladier, A. & Harris, T.N. 1989. *The form of information in science. Analysis of an immunology sublanguage.* Dordrecht: Kluwer Academic Publishers.

Koch, P. 2000. Pour une approche cognitive du changement sémantique lexical : aspect onomasiologique. In J. François (Ed.), *Théories contemporaines du changement sémantique* (pp.75-96). Leuven: Peeters (Société de Linguistique de Paris).

Nyckees, V. 2000. Changement de sens et déterminisme socio-culturel. In J. François (Ed.), *Théories contemporaines du changement sémantique* (pp.31-58). Leuven: Peeters (Société de Linguistique de Paris).

Rebeyrolle, J. & Tanguy, L. 2000. Repérage automatique de structures linguistiques en corpus : le cas des énoncés définitoires. *Cahiers de Grammaire,* 25, pp. 153-174.

Rousseau, A. 2000. L'évolution lexico-sémantique : explications traditionnelles et propositions nouvelles. In J. François (Ed.), *Théories contemporaines du changement sémantique* (pp.11-30). Leuven: Peeters (Société de Linguistique de Paris).

Teubert, W. 2001. Corpus linguistics and lexicography. *International Journal of Corpus Linguistics,* Special Issue, pp.125-153.

Traugott, E. & Dasher, R. (2002). *Regularity in semantic change*: Cambridge University Press.

Ullmann, S. 1969. *Précis de sémantique française.* Berne: A. Francke.