

Aurélie Picton*

Marqueurs et contextes riches en connaissances pour observer l'évolution en diachronie courte : Éléments méthodologiques en corpus (Markers and Knowledge-Rich Contexts to Study Short-Term Evolution in Corpora: Elements for a Methodological Framework in French)

Abstract

Knowledge evolution is a central and rapid phenomenon in specialized domains, and particularly in technologically advanced sectors such as space research. Learning to detect it over short-time spans (approx. 10 years) is highly necessary for several applications such as terminological bases updating, competitive intelligence, etc. To detect such an evolution, one possibility is to identify and analyse (semi)automatically clues of evolution in texts. This paper focuses on one possible linguistic clue, based on information extraction and data-mining approaches to learn how to track evolution: the identification of knowledge-rich contexts of evolution.

Résumé

L'évolution des connaissances est un phénomène central et inévitable dans les domaines de spécialité, et en particulier dans les domaines de pointe tels que le spatial. Il est donc nécessaire d'apprendre à repérer cette évolution, en particulier sur de courts intervalles temporels (environ 10 ans), afin de répondre à plusieurs besoins applicatifs tels que la mise à jour de ressources terminologiques, la veille scientifique et technique, etc. Pour ce faire, il est possible d'envisager de mettre au jour des indices linguistiques associables à une évolution des connaissances et repérables (semi) automatiquement dans les textes. Dans cet article, nous proposons un exemple d'indice inspiré de techniques d'extraction d'information et de fouille de textes pour apprendre à repérer l'évolution: l'analyse de contextes riches en connaissances évolutives.

1. Introduction et cadre de l'étude

L'évolution des connaissances est une réalité très présente dans les domaines scientifiques, notamment dans les domaines de pointe tels que le spatial. L'agence spatiale française (le Centre National d'Études Spatiales (Cnes)¹) est pleinement consciente de ce phénomène et a clairement mis en avant le fait que, bien que ces développements et changements soient souhaitables, ils peuvent néanmoins entraîner des difficultés non négligeables dans la pratique des experts du domaine. Les effets de cette évolution sont particulièrement sensibles dans les contextes de mise en place de projets spatiaux dits « de longue durée » (c'est-à-dire une dizaine à une trentaine d'années) pendant lesquels les connaissances convoquées évoluent nécessairement, parfois insidieusement. Cette situation peut entraîner un certain nombre de difficultés auprès des experts telles qu'une mauvaise communication entre ingénieurs « séniors » et ingénieurs « juniors » qui arrivent en cours de projet, oubli du contexte de connaissances dans laquelle projet a été initié ou plus généralement, modification non consciente du sens/de la forme des termes. Cette situation a été repérée par le Cnes comme problématique et il apparaît nécessaire de savoir repérer et identifier ces changements. Pour ce faire, à travers notre point de vue de linguiste, nous essayons de voir dans

1 <http://www.cnes.fr>

* Aurélie Picton
Univ. Toulouse 2, CLLE-ERSS, CNRS & Centre National d'Études Spatiales
5, allées A. Machado
31 058 Toulouse Cedex 9
France
Aurelie.Picton@univ-tlse2.fr

quelles mesures les textes du domaine vont refléter cette évolution, quels indices linguistiques sont mis en jeu dans les textes pour la marquer et quels sont les moyens de les repérer (Condamines et al. 2004; Picton 2009²ii). Parallèlement, dans la mesure où les techniques et outils de traitement automatique ont été considérablement développés, ils permettent d'envisager d'assister le processus de repérage du changement dans les textes, voire de l'automatiser en partie. Plus précisément, nous nous inspirons ici de travaux d'extraction d'information pour mettre en place un repérage automatisé de portions de textes qui contiennent des informations pertinentes sur l'évolution du domaine et que nous appelons *contextes riches en connaissances évolutives*.

Dans les sections suivantes, nous présentons dans un premier temps les travaux desquels s'inspire notre démarche et définissons la notion de *contextes riches en connaissances évolutives*. Nous détaillons ensuite le corpus d'étude construit pour cette étude et la manière dont nous identifions et extrayons les contextes. La quatrième section est dédiée à l'analyse des résultats : nous mesurons la pertinence et la précision des contextes et détaillons plusieurs analyses et études de cas pour montrer leur intérêt. Nous terminons sur quelques conclusions et perspectives.

2. Travaux liés et définitions

2.1. Définitions : *marqueurs et contextes riches en connaissances évolutives*

Les *contextes riches en connaissances évolutives* sont entendus dans cet article comme des portions de textes qui contiennent des informations pertinentes sur l'évolution des connaissances d'un domaine. Cette notion s'inspire de la notion de *contexte riche en connaissances* (ou *Knowledge-Rich Context*, KRC) proposée notamment par Meyer (2001) et que l'auteur définit comme

« indicating at least one item of domain knowledge that could be useful for conceptual analysis. In other words, the context should indicate at least one conceptual characteristic, whether it be an attribute or a relation. » (Meyer 2001 : 281)

Cette notion a vu le jour en terminologie et en ingénierie des connaissances et fait appel à la notion de *marqueurs* pour la caractérisation des contextes. Les marqueurs sont au cœur des tâches de construction et de structuration de ressources termino-ontologiques à partir de textes où l'« on suppose que l'observation de certaines formes linguistiques entre deux ou plusieurs éléments du lexique peut révéler un rapport de sens entre ces éléments » (Aussenac-Gilles/Séguéla, 2000 : 175). Ces formes linguistiques, lorsqu'elles permettent de rendre compte de fonctionnements stables et associables systématiquement à une interprétation sémantique, sont appelées *marqueurs de relations*³iii. Les contextes dans lesquels apparaissent des marqueurs et des termes qui entrent dans la relation décrite sont alors appelés *contextes riches en connaissances*.

Dans la lignée de ces travaux, nous faisons l'hypothèse qu'il est possible de définir des marqueurs d'évolution pour identifier des contextes riches en connaissances évolutives, contextes qui contiennent des informations pertinentes sur l'évolution du domaine, quelle qu'elle soit. C'est ce que nous montrons dans la suite de cet article. Mais avant de poursuivre, il convient de mieux préciser les travaux desquels s'inspire cette étude.

2.2. Travaux liés : entre extraction d'information et repérage de relations sémantiques

Nous l'avons dit, de nombreux travaux en terminologie se sont penchés sur la question des marqueurs et des contextes riches en connaissances. Or, une différence notable existe entre les marqueurs de relations conceptuelles et les marqueurs d'évolution tels que nous les envisageons, ces

2 Cet article est une version modifiée du chapitre 5 de (Picton 2009).

3 Voir L'Homme/Marshman (2006) pour une revue détaillée sur cette question. Sur des relations spécifiques voir notamment Borillo (1996) ou Hearst (1992) sur l'hyponymie/hyperonymie, Condamines (2000) ou Otman (1996) sur la méronymie, Barrière (2001) ou Marshman (2006) sur la relation cause/effet, Hamon (2000) ou Suarez et Cabré (2002) sur la synonymie.

derniers ne permettant pas d'établir de lien conceptuel entre plusieurs termes/concepts. Plus précisément, le principe des marqueurs de relations est de mettre au jour des patrons dans lesquels entrent en jeu une relation conceptuelle et les termes/concepts pris dans cette relation. Par exemple, le patron classique [X est un type de Y] indique une relation d'hyponymie/hyperonymie entre les termes/concepts X et Y. Dans un texte traitant de botanique, un contexte tel que *la violette blanche et la violette des collines sont deux types de fleurs répandus en Haute-Garonne* permet d'établir un lien conceptuel entre le terme/concept hyperonyme *fleur* et les termes/concepts hyponymes *violette blanche* et *violette des collines*^{4v}. Dans le cas des marqueurs d'évolution par contre, l'information repérable est beaucoup plus disparate et ne vise pas à établir de lien entre termes/concepts. Il ne s'agit donc pas de mettre au jour des patrons qui contiendraient un marqueur d'évolution et des termes concernés par cette évolution, mais plutôt de localiser des informations pertinentes concernant l'évolution des connaissances d'un domaine, que celles-ci impliquent des termes/concepts spécifiques ou non. C'est ce que nous illustrons par la suite. De ce point de vue, notre démarche doit être mise en lien également avec des travaux d'extraction d'informations dites « évolutives » tels que ceux de Ibekwe-Sanjuan (2005) pour la veille scientifique et technologique, ceux de Sándor (2006) pour la veille et la gestion des risques ou encore ceux de Laignelet (2009) sur le repérage de segments obsolescents pour la mise à jour de ressources encyclopédiques. Dans ces travaux, l'extraction d'information repose sur l'identification de portions de textes, tels que des phrases ou des paragraphes, susceptibles de contenir des informations pertinentes sur l'évolution, à savoir la nouveauté pour Ibekwe, la nouveauté et les risques pour Sándor, et l'obsolescence pour Laignelet. Il ne s'agit donc plus de se concentrer sur des patrons locaux, mais sur des zones restreintes de textes.

Ces différents objectifs d'extraction permettent de plus de souligner le fait qu'une approche par marqueurs pour l'identification de contextes est une approche « souple » qui peut être adaptée à de nombreux types d'informations tels que l'hyponymie, la synonymie, la nouveauté, le risque, l'obsolescence, etc. Ceci est un atout pour notre étude puisque cela permet d'envisager de mettre en place le repérage de nombreux aspects d'évolution différents et de mettre au jour des phénomènes riches et inédits de changement. C'est ce que nous argumentons dans les sections suivantes, en commençant par décrire la manière dont les contextes sont définis à partir de marqueurs.

3. Extraction des contextes d'évolution à l'aide de marqueurs

3.1. Corpus d'étude

Afin de correspondre au cadre de la diachronie courte imposée par le contexte des projets spatiaux, le corpus construit pour cette recherche est basé sur les chapitres d'optique et d'optoélectronique d'un cours édité par le Cnes : le Cours de Techniques et Technologies des Véhicules Spatiaux (désormais TTVS). Ce cours a été édité tous les quatre ans depuis 1994 par les éditions Cépaduès (Cnes 1994; 1998; 2002) et a été rédigé par plus de 80 experts à l'attention de semi-experts (Bowker/Pearson 2002). Le tableau suivant présente le nombre d'occurrences pour chacun des sous-corpus.

	TTVS1994	TTVS1998	TTVS2002	Total
Nombre d'occurrences	46 448	78 656	109 505	234 609

Tableau 1. Nombre d'occurrences dans le corpus TTVS

4 Bien qu'il puisse arriver que l'un des termes/concepts concerné par la relation sémantique ne soit pas directement présent dans le patron recherché en corpus (par exemple dans les cas d'ellipses, d'anaphores, etc. (Marshman 2006 : 153sq.)).

3.2. Liste de marqueurs proposée pour le français

L'extraction de contextes riches en connaissances évolutives passe dans un premier temps par la définition de marqueurs d'évolution. Les marqueurs choisis pour cette étude sont définis *a priori* et identifiés sur la base d'une démarche introspective et itérative, que l'on retrouve dans les travaux de Méla (2004) ou Leroy (2004) par exemple. Plus précisément, nous partons d'une première liste de quelques marqueurs définis sur la base de notre intuition de locuteur. Cette liste est projetée dans les corpus à l'aide de l'outil TerminoWeb (Barrière/Agbago 2006) qui permet ensuite de visualiser les extraits de corpus dans lesquels apparaissent les marqueurs. L'observation des résultats permet d'ajuster la première liste intuitive en y ajoutant de nouveaux marqueurs possibles qui apparaissent dans les résultats obtenus mais qui n'avaient pas été identifiés. Ils permettent également d'éliminer les marqueurs qui apportent trop de bruit. Ces allers-retours entre la liste de marqueurs et les extraits en corpus sont reproduits de manière itérative jusqu'à ce que la liste obtenue soit satisfaisante.

La sélection *a priori* des marqueurs n'est pas une étape aisée et repose sur une série de choix justifiés par les besoins de l'analyse. Les deux principaux choix concernent les catégories de marqueurs et les types d'informations recherchés. Dans notre cas, les marqueurs choisis sont des marqueurs lexico-syntaxiques (noms, adjectifs, adverbes, verbes, locutions verbales et syntagmes prépositionnels essentiellement).

Du point de vue des types d'informations, la liste définie ici contient une quarantaine de marqueurs répartis en neuf « groupes » d'interprétations qui couvrent de manière satisfaisante le champ des évolutions possibles en diachronie courte. Nous avons défini ces groupes sur la base de trois observations :

1. l'étude de l'évolution peut se baser sur la datation d'événements;
2. les évolutions typiquement prises en considération sont la nouveauté et la disparition de termes/concepts;
3. les évolutions scientifiques et techniques sont souvent motivées par l'apparition de nouveaux besoins dans le domaine, qui entraînent alors progrès et améliorations.

À partir de ces trois observations les marqueurs proposés sont répartis comme suit :

A. Datation d'événements

A.a. Marqueurs de datation : pour la datation des événements liés à des concepts du domaine;

- **an(s)|année(s) [date]**

- (i) *Les polynômes de Zernike ont pris de l'importance quand les ordinateurs ont permis de les calculer rapidement (ils ont été inventés dans le début des années 40 ! mais jamais exploités)*

- **en [date]**

- (ii) *Le premier terminal est monté sur Spot 4 (lancé en 2000), le second sur ARTEMIS (lancé en 2001)*

- **depuis quelques années**

- (iii) *Depuis quelques années, la technique d'usinage ionique sans contact avec le miroir est opérationnelle*

A.b. Marqueurs d'ancrage dans le présent/l'actualité : marqueurs qui ancrent les informations dans le présent et indiquent que les informations données sont d'actualité au moment de la rédaction des textes

- **à ce jour**

(iv) *on a pu lire que les plus grandes barrettes à ce jour comprennent autour de 12 000 détecteurs*

- actuellement

(v) *Ces lidars sont actuellement extrêmement demandés par la communauté scientifique*

- aujourd'hui

(vi) *le Zérodur, le plus mature, et le Carbure de Silicium (SiC) fritté qui sont utilisés aujourd'hui dans la majorité des cas*

- classique

(vii) *Baffle semi-spéculaire Baffle semi- spéculaire : La structure est celle d'un baffle classique*

- disponible

(viii) *Les outils et les données technologiques disponibles permettent de limiter les risques lors du développement d'une nouvelle fonction*

- exister/existant

(ix) *De nombreuses variantes de ce télescope existent*

- maintenant

(x) *Le signal est stocké dans une mémoire (bande magnétique et maintenant, mémoire à état solide)*

A.c. Marqueurs d'ancrage dans le passé : marqueurs qui ancrent les connaissances dont il est question dans le passé et qui permettent de relever certaines informations valides à un temps antérieur à celui de la rédaction du corpus;

- autrefois

(xi) *Autrefois, la technologie des détecteurs ne permettait pas d'avoir des barrettes ou des mosaïques (CCD)*

- histoire

(xii) *soit au tout début de l'histoire des lasers - passé*

(xiii) *Jusqu'à un récent passé, ces instruments n'étaient pas disponibles à cause de leur difficulté de polissage*

B. Nouveauté et disparition de termes ou de concepts

B.a. Marqueurs de nouveauté ou d'apparition d'une entité : pour annoncer la présence de termes/concepts présentés comme nouveaux ou récents. Comme l'illustre l'exemple du marqueur *inventer* infra (xvi), l'état de nouveauté n'est pas toujours relatif au moment de la rédaction du corpus, mais peut être relatif à une date précise spécifiée dans le contexte;

- apparaître

(xiv) *Partant de ces produits sont apparues des applications pour les caméras d'observation de la Terre*

- arrivée

(xv) *observation ou imagerie spectrale à haute résolution, surveillance de l'environnement, l'arrivée du multimédia, la télévision numérique à haute définition*

- inventer

(xvi) *ils ont été inventés dans le début des années 40*

- nouveau

(xvii) *Un produit nouveau est apparu depuis quelques années sur le marché, il s'agit de multi barrettes*

- récent

(xviii) *CCD aminci : Le substrat dans lequel est réalisé le détecteur est aminci. [...] Cette technologie, récente, est très appréciée*

B.b. Marqueurs de disparition, d'obsolescence, de passé : marqueurs qui indiquent la présence de termes/concepts qui ont été abandonnés ou sont devenus obsolètes;

- abandonner

(xix) *Ce type d'architecture a été étudié en phase 0 du projet Pléiades pour être finalement abandonnée pour des problèmes d'obscurité*

C. Nouveaux besoins potentiels

C.a. Marqueurs d'insuffisance : marqueurs qui ciblent les lacunes et les insuffisances de certains concepts du domaine. De fait, l'hypothèse est que ces concepts seront potentiellement liés à des progrès à venir pour pallier ces lacunes et répondre aux besoins du domaine;

- malheureusement

(xx) *Malheureusement, ces hybrides nécessitent des photocathodes d'efficacité quantique limitée (10 à 20 % maxi) et ne travaillent que dans le visible*

- pas assez

(xxi) *les mesures ne sont pas assez précises et on ne peut pas obtenir toutes les altitudes et observer partout*

- pas encore

(xxii) *Comme la technologie des diodes laser ne permet pas encore de monter trop haut en puissance [...] on juxtapose sur le même télescope plusieurs diodes laser [...]*

- pas satisfaisant

(xxiii) *la qualité des filtres [...] employés pour la reprographie ne sont pas satisfaisants pour le spatial, et il faut envisager des filtres plus spécifiques*

C.b. Marqueurs de comparaison : marqueurs qui expriment une relation de comparaison entre au moins deux concepts. Il peut s'agir d'une comparaison entre deux concepts différents, mais également entre les différents états ou « générations » d'un seul et même concept qui a évolué;

- meilleur (que)

(xxiv) *Le CNET a réalisé des lasers à puits quantiques dont le courant de seuil est inférieur à 2 et le gain différentiel de l'ordre de 0,3 W/A (trois fois meilleur qu'un laser classique)*

- moins [...] (que)

(xxv) *un modèle macroscopique simplifié, et de ce fait moins précis ou avec un domaine de validité plus restreint, mais plus rapide en terme de simulation*

- plus [...] (que)

(xxvi) *On cite, par exemple les nouveaux matériaux pour la fabrication de miroirs, les allègements plus poussés, les plans focaux plus "intégrés"*

C.c. Marqueurs de remplacement ou de succession : marqueurs qui permettent de repérer les cas où un concept fait ou a fait place à un autre. Le « remplaçant » est susceptible de devenir la nouvelle référence dans le domaine alors que le « remplacé » est susceptible de devenir obsolète

- remplacer

(xxvii) *les détecteurs CMOS vont remplacer les CCD*

- successeur

(xxviii) *Polder, son successeur Parasol, ou encore Modis (USA) [...], sont d'autres exemples*

C.d. Marqueurs d'amélioration ou de développement : marqueurs qui permettent d'extraire trois types d'informations proches sur l'évolution, à savoir des informations sur des concepts améliorés dans le domaine, des informations sur des concepts en cours de développement, en progrès, ou susceptibles de faire l'objet de travaux et des informations sur des concepts en essor ou en développement au moment de la rédaction;

- être à l'étude

(xxix) *Comme il existe d'autres contraintes, il faut utiliser d'autres "astuces" pour parfaire cette réjection [...] et même d'autres principes instrumentaux sont à l'étude*

- amélioration/améliorer

(xxx) *Les améliorations technologiques portent généralement sur les techniques de confinement de la lumière et des porteurs d'électrons (électrons et trous)*

- développement/développer

(xxxii) *Les instruments développés seront donc des caméras d'imagerie à haute résolution spatiale*

- encourageant

(xxxiii) *A la vue des résultats très encourageants obtenus*

- essor

(xxxiiii) *L'optique en plein essor : les techniques "SOL" actuelles*

- évoluer/évolution

(xxxv) *Les lasers solides ont, eux, évolué vers la production de lasers compacts*

- progrès

(xxxvi) *On notera les progrès importants accomplis dans la fabrication de barrettes et mosaïques CCD infrarouges*

- prototype

(xxxvii) *on voit apparaître des prototypes de deux miroirs*

- réduire

(xxxviii) *Ceci conduit à réduire les dimensions des antennes et certains équipements et composants*

C.e. Marqueurs de prédiction (futur) : marqueurs qui permettent de repérer des informations sur des concepts dont les experts/rédacteurs pensent qu'ils seront centraux dans l'avenir.

- avenir

(xxxviii) *On peut facilement imaginer que l'avenir sera dans des instruments plus "simples", miniaturisés au maximum*

- envisager

(xxxix) *Dans le plus proche avenir, il faudra envisager sérieusement des systèmes de télécommunication à très hauts débits*

- futur

(xl) *Mais il est certain que leur évolution est à suivre de près pour les futures applications spatiales*

- imaginer

(xli) *On pourra aussi imaginer d'autres utilisations des lasers semiconducteurs pour les communications intra-satellite ou pour la métrologie à bord*

prometteur

(xlii) *ces instruments (détection directe) semblent très prometteurs*

C'est à partir de cette liste de 40 marqueurs que sont extraits les contextes riches en connaissances évolutives. Pour cela, les contextes recherchés sont définis comme des portions de textes qui contiennent au moins un marqueur d'évolution et un ou plusieurs terme(s)/concept(s) du domaine.

3.3. Extraction de contextes riches en connaissances évolutives

Mais une fois la liste de marqueurs définie, l'une des principales difficultés est de savoir comment délimiter les contextes à observer. En effet, les marqueurs d'évolution ont ceci de particulier qu'ils ne peuvent généralement pas être construits sous la forme de patrons qui les associeraient à des termes/concepts du domaine (voir supra). Dans l'approche choisie ici, l'objectif est donc plutôt de mettre au jour des « zones » (ou des segments à explorer, pour reprendre la terminologie de Laignelet (2009)) que de véritables contextes complets et clos.

La délimitation des contextes se fait le plus souvent en considérant le niveau de la phrase (voir notamment Ibekwe-Sanjuan (2005)). On pourrait considérer alors qu'un contexte riche en connaissances évolutives est une phrase qui contient au moins un marqueur et un terme/concept du domaine. Mais en observant manuellement les résultats obtenus dans nos corpus, nous avons finalement porté notre choix de délimitation sur le paragraphe. En effet, le contexte de la phrase s'est avéré souvent trop court pour extraire des informations pertinentes complètes ; le paragraphe semble de fait un bon compromis entre la phrase et une section titrée, souvent trop longue et trop hétérogène du point de vue de l'information contenue. Néanmoins, certains paragraphes peuvent ne pas contenir toute l'information pertinente. C'est le cas notamment des paragraphes qui reprennent un ensemble d'informations décrites dans les paragraphes précédents. Par exemple, dans le contexte d'évolution suivant (où les marqueurs sont en gras et les termes/concepts sont en gras et entre crochets) :

(xliii) *Ces **nouvelles technologies** n'ont **pas encore** réellement fait leur entrée dans le spatial, car elles sont plus **récentes** que celle des [**semi-conducteurs**]. Elles ont donc encore à prouver leurs potentialités dans ce domaine. Mais il est certain que leur évolution est à suivre de près pour les **futures applications spatiales**, où faibles [**encombrement**] et [**masse**], faible [**consommation électrique**] et bas coût seront des caractéristiques déterminantes.*

l'anaphore en début de paragraphe (*ces nouvelles technologies*) indique que la liste des technologies marquées comme nouvelles ne peut être retrouvée que plus haut dans le texte, dans les paragraphes précédents. Dans ce cas, et avec la méthodologie mise en place dans cette étude, le linguiste/terminologue devra explorer (manuellement) un contexte un peu plus large pour récolter l'ensemble des informations pertinentes. Ces cas sont cependant assez peu fréquents. Notons que d'autres démarches pour la délimitation de contextes sont possibles, par exemple à l'aide de calculs statistiques, sur le modèle du TextTiling de Hearst (1997)⁵.

Les contextes riches en connaissances évolutives sur lesquels nous basons nos observations sont donc des paragraphes dans lesquels au moins un marqueur et un terme/concept sont présents. Nous décrivons les analyses que ces contextes permettent de construire et montrons notamment en quoi la combinaison de plusieurs marqueurs influe sur la pertinence des contextes extraits.

3.4. Observations en corpus : Projection des termes en contextes

Une fois les contextes extraits, l'objectif est de chercher à observer la liste des termes du domaine y gravitant effectivement. La liste de termes retenus ici a été extraite à l'aide de l'outil Syntex (Bourigault 2007), validée par des experts du Cnes, et contient à ce jour près de 2000 termes. À partir de cette liste et des contextes, deux points de vue se complètent pour l'analyse :

- analyser chacun des contextes isolément,
- ou observer la répartition des termes dans ces contextes.

3.4.1. Évaluation de la pertinence des contextes et des marqueurs

De manière générale, les contextes extraits contiennent un bruit assez important : un peu plus de la moitié d'entre eux ne sont pas pertinents, comme l'illustre le Tableau .

	Nombre de contextes d'évolution	Nombre de contextes pertinents	Précision (%)
TTVS1994	255	108	42,35
TTVS1998	420	160	38,10
TTVS2002	527	207	39,28
Total	1202	475	39,52

Tableau 2. Pertinence des contextes d'évolution

Ce bruit est principalement dû à deux phénomènes. Le premier est lié à la présence de marques d'incertitude ou de négation difficiles à anticiper et qui viennent modifier l'interprétation associable *a priori* au marqueur. En effet, la plupart des marqueurs proposés sont susceptibles d'être modifiés, niés ou « nuancés » par le rédacteur dans les textes. Ainsi, dans l'extrait

(xliv) *Dans le domaine strict des communications, on peut se poser la question de savoir si la détection hétérodyne supplantera la détection directe, grâce à ses grandes potentialités, ou si, au contraire, l'évolution très rapide des performances des lasers semi-conducteurs consacrerà cette dernière par sa simplicité d'emploi*

l'évolution est présentée sous forme d'hypothèse. L'interprétation doit donc être nuancée voire ne pas être prise en compte.

Le second phénomène, plus fréquent, renvoie au fait que dans certains contextes, l'information temporelle extraite n'est pas liée au domaine en général mais à une situation donnée, un « contexte de discours », phénomène qui se rapporte à la notion de « relation discursive contextuelle » décrite par Borillo (1996 : 122). L'auteure propose cette notion dans le cadre de l'exploration

5 C'est ce qu'ont tenté notamment Laignelet et Pimm (2007).

de la relation conceptuelle d'hyponymie/hyperonymie. Elle remarque que, dans certains cas, la relation extraite n'est pas stable et lexicale et n'intervient que dans une situation discursive donnée. Dans le cas des contextes riches en connaissances d'évolution, ceci se traduit par des cas où l'évolution en jeu ne renvoie pas à une évolution stable et avérée dans les connaissances générales du domaine, mais à une évolution dans le cadre d'une situation spécifique telle qu'une expérience scientifique. Par exemple :

(xlv) *L'approche est d'**augmenter** progressivement les rapports D/d. Au **début**, on est à la limite de l'occultation centrale : la pupille est "presque pleine".*

Dans cet extrait, on voit que les marqueurs en gras et soulignés renvoient à des actions ou des étapes propres à une expérience décrite et ne peuvent être associés à des connaissances générales du domaine.

Ce cas de figure a été particulièrement présent pour certains marqueurs, tels que *changer/changement* qui a dû finalement être éliminé de la liste finale dans la mesure où il apportait trop de bruit. En effet, bien que certains contextes aient été pertinents :

(xlvi) *Tout semble indiquer que les réseaux de communications vont être **changés** par l'arrivée des technologies de la photonique.*

ce marqueur prend le plus souvent dans nos corpus une valeur propre à un contexte discursif spécifique :

(xlvii) *il ne nous reste plus qu'à implanter par diffusion des impuretés **n** dans le substrat pour **changer** la nature du semi-conducteur dans la zone délimitée par le masque.*

Malgré ces erreurs, la pertinence des contextes reste intéressante, d'autant plus qu'elle peut dépendre du nombre de marqueurs contenus par contexte. Cette observation, que l'on retrouve notamment chez Malaisé et al. (2005), signifie que plus un contexte contient de marqueurs, plus il est pertinent du point de vue de l'évolution. Il est donc possible de filtrer les contextes les plus pertinents en fonction du nombre de marqueurs qu'ils contiennent.

Pour illustrer cet aspect, le Tableau présente la répartition des différents contextes évolutifs en fonction du nombre de marqueurs qu'ils contiennent. Le Tableau présente le nombre et la proportion de contextes **pertinents** en fonction du nombre de marqueurs.

		TTVS1994	TTVS1998	TTVS2002	Total
Nombre de marqueurs par contexte	8	0	1	1	2
	7	1	1	1	3
	6	0	3	4	7
	5	1	5	5	11
	4	7	6	9	22
	3	7	24	38	69
	2	54	83	120	257
	1	185	297	349	831
Total	255	420	527	1202	

Tableau 3. Répartition des contextes en fonction du nombre de marqueurs

		TTVS1994		TTVS1998		TTVS2002		Total	
		Nb	%	Nb	%	Nb	%	Nb	%
Nombre de marqueurs par contexte	8	-	-	1	100	1	100	2	100
	7	1	100	1	100	1	100	3	100
	6	-	-	3	100	4	100	7	100
	5	1	100	5	100	5	100	11	100
	4	7	100	5	83,33 ⁶	9	100	21	94,44
	3	5	71,43	17	70,83	27	71,05	49	71,10
	2	23	42,59	35	42,17	48	40	106	41,59
	1	71	38,38	93	31,31	112	32,09	276	33,93
	Total	108	42,35	160	38,10	207	39,28	475	39,52

Tableau 4. Nombre et proportion de contextes pertinents en fonction du nombre de marqueurs

Ces deux tableaux permettent de confirmer que plus les marqueurs sont nombreux dans un contexte donné plus il est probable que celui-ci soit pertinent. Ainsi, dans les cas où les contextes contiennent de 3 à 8 marqueurs (en gris dans le Tableau 4), de 71 à 100% des contextes sont pertinents, soit 93 contextes sur un total de 114. Par contre, les performances se dégradent assez significativement dès que le contexte ne contient qu'un ou deux marqueur(s). Cependant, il est clair également que plus les contextes contiennent de marqueurs moins ils sont nombreux.

Sur la base de cette observation, ne seront donc pris en considération dans la suite de ce chapitre que les contextes qui contiennent au moins 3 marqueurs dans la mesure où il s'agit des cas où plus de la moitié des contextes extraits sont pertinents. Le taux de précision moyen atteint alors 81,57% dans l'ensemble du corpus TTVS (Tableau 5).

Souscorpus	Nombre de contextes d'évolution	Nombre de contextes pertinents	Précision
TTVS1994	16	14	87,50 %
TTVS1998	40	32	80 %
TTVS2002	58	47	81,03 %
Total	114	93	81,58 %

Tableau 5. Pertinence des contextes d'évolution

À partir de ces contextes, la précision des marqueurs a été évaluée : pour chacun, la proportion d'occurrences pertinentes est calculée par rapport à l'ensemble des occurrences du marqueur dans les contextes. Par exemple, dans un contexte tel que :

(xlvi) Une **nouvelle** classe d'instruments devient nécessaire : l'interférométrie qui permet d'augmenter la résolution spatiale. Cette technique est associée à une autre, la synthèse d'ouverture qui permet de reconstruire progressivement l'image de l'objet à partir des informations de son spectre en fréquences spatiales (contraste et phase). Cette technique fait appel à l'interférométrie à grande base. Déjà implantée au sol (VLTi), pour des missions relativement simples, et grâce aux **progrès récents** des technologies de l'optique, l'interférométrie pourra devenir la **nouvelle** composante de l'instrumentation optique spatiale. (TTVS)

on compte 4 occurrences de marqueurs (en gras), dont 2 pour le marqueur *nouvelle*. Ces 4 occurrences sont pertinentes.

⁶ Ce chiffre, relativement bas comparé aux 100% de précision obtenus, doit être lu avec précaution. En effet bien que la pertinence soit dégradée de près de 17%, dans les faits un seul contexte contenant quatre marqueurs n'est pas pertinent sur les 22 contextes concernés. Relativement au nombre de contextes concernés, il ne s'agit pas d'un bruit considérable malgré sa proportion de 17%.

Par contre, dans un contexte tel que :

- (xlix) *Les exemples ne manquent pas pour convaincre le lecteur que, dans le plus proche avenir, il faudra envisager sérieusement des systèmes de télécommunication à très hauts débits et suffisamment évolutifs pour s'adapter à des besoins très particuliers.* (TTVS)

on compte 3 occurrences de marqueurs, mais seulement 2 pertinentes (*évolutifs* indique ici une propriété des systèmes de télécommunication, et non pas une information sur l'évolution du domaine). Le contexte reste cependant pertinent.

Dans un second temps, nous nous intéressons à la performance des marqueurs seuls.

Le Tableau 6 indique la précision des marqueurs en fonction des trois groupes principaux d'interprétation auxquels ils appartiennent (voir *supra*).

	Occurrences totales	Pertinents	Non pertinents	Précision (%)
A. Marqueurs de datation	91	77	14	89,01
B. Nouveauté et disparition de termes/concepts	68	61	7	89,86
C. Nouveaux besoins potentiels	320	229	91	71,47

Tableau 6. Précision des trois groupes principaux de marqueurs (TTVS)

Chacun des trois groupes présente une précision tout à fait satisfaisante, supérieure à 70%. Le groupe « Nouveaux besoins potentiels » est cependant un peu moins précis que les deux autres, ce qui s'explique en partie par le fait qu'il contient beaucoup de marqueurs « multi-valeurs », c'est-à-dire les marqueurs dont l'information temporelle est susceptible de s'ancrer dans un contexte discursif spécifique et qui n'indiquent alors pas une évolution des connaissances du domaine (voir Borillo 1996, *supra*).

Le Tableau 7 ci-dessous détaille pour chacun des sous-corpus les précisions de chaque classe de marqueurs.

	Sous-corpus								
	TTVS1994			TTVS1998			TTVS2002		
	Occurrences	Pertinents	Précision	Occurrences	Pertinents	Précision	Occurrences	Pertinents	Précision
Datation	7	7	100	31	29	93,55	53	45	84,91
Nouveauté et disparition de termes/concepts	13	10	76,92	22	19	86,36	34	33	97,06
Nouveaux besoins potentiels	51	43	84,31	124	76	61,29	144	111	77,08

Tableau 7. Précision des marqueurs en fonction de leur classe dans chaque sous-corpus

Dans les trois sous-corpus, les marqueurs « de datation » sont parmi les plus précis, suivis par les marqueurs de « nouveauté et disparition ». Néanmoins, il apparaît également nettement que la performance des marqueurs choisis dépend du sous-corpus, bien que ces variations restent difficiles à expliquer précisément. Ce type d'indice linguistique pour traiter l'évolution doit donc être considéré avec précaution et devra être mis en œuvre dans d'autres types de corpus (et avec d'autres marqueurs également) afin d'étudier au mieux les possibilités qu'il offre. Néanmoins, les performances globales présentées ici restent tout à fait satisfaisantes pour la démarche d'exploration qui est la nôtre et offrent une base solide pour observer qualitativement les informations ainsi extraites.

3.4.2. Observations de contextes

Le premier point de vue possible pour l'observation est d'analyser chacun des contextes extraits pour retracer leur évolution. Pour ce faire, deux perspectives sont possibles et complémentaires : la synchronie dynamique et la diachronie.

3.4.2.1. Point de vue synchronique dynamique

Une première manière d'aborder l'analyse des contextes extraits est de les observer en synchronie. Plus exactement, le point de vue adopté est un point de vue synchronique dynamique (Gilbert 1975; Houdebine 1985), c'est-à-dire d'observer des traces d'évolution et de changement sans comparer différents sous-corpus, et donc en synchronie.

Comme les neuf classes de marqueurs définies l'indiquent, les informations observables en contextes peuvent être de différentes natures, mais certaines présentent un intérêt particulier pour apprendre à connaître l'histoire, les étapes ou la chronologie du domaine dont on analyse l'évolution. C'est le cas notamment des datations. En effet, les contextes qui contiennent des marqueurs de datation permettent assez rapidement au linguiste de placer les premiers repères de l'évolution qu'il tente de saisir. Par exemple, dans le TTVS2002 des contextes tels que^{7vii}

- (i) *Depuis quelques années, la [technique d'usinage ionique] sans contact avec le [miroir] est opérationnelle.*
- (ii) *Un produit nouveau est apparu depuis quelques années sur le marché, il s'agit de [multi barrettes], c'est-à-dire que sur la même [puce détectrice] sont implantées plusieurs [barrettes].*

permettent de retenir que les multi-barrettes ou la technique d'usinage ionique pour les miroirs sont des concepts récents en 2002, au moment de l'édition du cours. Plus précisément, à l'aide des marqueurs de nouveauté, le premier extrait souligne des progrès récents de la technologie des miroirs, alors que le second indique de nouvelles technologies pour les puces détectrices. Ces extraits illustrent la complémentarité qui existe entre les différentes classes de marqueurs. La datation des informations n'est cependant pas toujours liée à la date de rédaction du corpus, mais plutôt à une chronologie donnée dans les contextes. Par exemple, les extraits du TTVS 2002 ci-dessous offrent des éléments intéressants pour retracer l'histoire des lasers dans le spatial :

- (iii) *L'émission dans un [barreau de verre dopé terre rare] a été démontrée dès les années 60, soit au tout début de l'histoire des [lasers].*
- (liii) *D'énormes progrès ont été faits dans le domaine des [lasers] depuis le démarrage de ce projet (années 80)*
- (liv) *Les progrès de la technologie [fibre optique] ont ensuite permis, dans les années 80, d'utiliser des [fibres monomodes] améliorant le [guidage], donc l'efficacité du [laser].*

Ces contextes datent l'apparition des lasers dans les années 60 et permettent de dater le début de l'utilisation des fibres monomodes dans les années 1980. Cette période est ainsi marquée comme centrale dans l'histoire et les progrès des technologies laser.

Ceci constitue l'un des intérêts majeurs des contextes qui, associés à une interprétation humaine, facilitent la déduction et la reconstruction d'éléments sur l'évolution du domaine, même si l'analyste n'est lui-même pas expert du domaine qu'il observe. Ceci fait de cette approche une démarche riche pour entamer le repérage de l'évolution d'un domaine pour lequel aucun *a priori* n'est posé.

Un autre exemple est illustré par ces extraits (toujours du TTVS2002) :

7 À partir de cette section, dans les exemples présentés les marqueurs seront mis en évidence en gras et les termes du domaine présents dans le contexte seront mis en évidence en gras et entre crochets.

(lv) *Le premier projet européen d'utilisation de [diodes lasers] sur un satellite est le projet [Silex] (Semi-conductor laser Intersatellite Link EXperiment). Ce projet a pour objet la démonstration d'une [liaison optique inter-satellite] (LEO-GEO) à 0,8 μm . Le premier terminal est monté sur Spot 4 (lancé en 2000), le second sur ARTEMIS (lancé en 2001).*

(lvi) *Les [télécommunications intersatellites] : les résultats récents de la liaison Silex*

L'intérêt de ces contextes est qu'ils renvoient à un projet spécifique, le projet Silex. Ils indiquent que le projet a été mis en application autour de 2000/2001, et qu'il exploite les télécommunications intersatellites. Le champ des télécommunications intersatellites apparaît en évolution, en lien avec les technologies lasers. En effet, il se démarque de projets plus anciens par l'utilisation de diodes lasers sur les satellites pour la télécommunication. Ce projet représente donc un tournant dans l'évolution des lasers au début des années 2000.

Un second type d'information intéressant à observer en synchronie dynamique sont les remplacements et successions, dans la mesure où ils permettent de suivre les différentes générations d'une même technologie ou d'un même concept, mais dans un corpus synchronique. Ainsi par exemple (TTVS1998) :

(lvii) *Le [pompage optique] se faisait dans un premier temps par [lampe flash], transversalement, sur des [fibres multimodes]. Les progrès de la technologie [fibre optique] ont ensuite permis, dans les années 80, d'utiliser des [fibres monomodes] améliorant le [guidage], donc l'efficacité du [laser]. La [fibre] peut alors être utilisée à la fois pour guider [l'onde] à [amplifier] mais aussi l'[onde de pompage] ([pompage longitudinal]). Ceci nécessite toutefois une source puissante et cohérente, avec des qualités de [faisceau] suffisantes. On remplace alors les [lampes] par des [diodes lasers].*

Dans ce contexte on observe que dans les années 1980 le pompage optique se faisait à l'aide de fibres monomodes et lampes flash. Les progrès des fibres ont entraîné le remplacement des lampes flash par les diodes lasers.

Ces quelques exemples montrent la richesse des informations que l'on peut mettre au jour par l'observation des contextes en synchronie. Cependant, bien que le dynamisme des connaissances transparaisse en synchronie, il reste souvent intéressant de compléter l'observation par une comparaison diachronique des contextes pour retracer l'évolution des connaissances à l'intérieur de l'intervalle temporel du corpus.

3.4.2.2. Point de vue diachronique

Le point de vue diachronique suppose que l'on dispose d'un corpus comparable, comme c'est le cas pour le corpus constitué dans notre recherche.

La comparaison de contextes permet d'accéder à des évolutions suivies telles que celles de la détection infrarouge et du refroidissement. En effet, en 1994, le contexte suivant indique :

(lviii) *En outre, en utilisant la propriété de couplage radiatif d'une plaque échangeant de l'énergie avec [l'espace froid], on réalise des [radiateurs passifs], capables de [refroidir] [détecteurs infrarouges], filtres et optiques associées, à des températures bien stabilisées (quelques degrés) autour de 80 à 100° K. La puissance thermique que peuvent dissiper ces instruments est en constante augmentation, grâce aux récents progrès de [machines cryogéniques actives] (typiquement 500 mW à 1 W autour de 80° K) ; ces machines permettront de [refroidir] des grands plans focaux, constitués par un grand nombre de [détecteurs infrarouges] ([barrettes] ou [mosaïques]) et ouvrent ainsi la voie à la haute résolution spectrale dans le domaine des fenêtres atmosphériques thermiques (3 à 5 μm et 8 à 12 μm essentiellement) (TTVS1994)*

Alors qu'en 2002, la même thématique est reprise dans des contextes d'évolution quelque peu différents :

(lix) *[L'infrarouge] est toujours lié à la [capacité de refroidir] : optiques et [détecteurs]. Si on arrive à s'affranchir de cette contrainte, on peut simplifier considérablement l'instrument. Or, récemment, de grands progrès ont été accomplis dans le domaine des [mosaïques de détecteurs]*

thermiques non refroidis] (*matrices de [microbolomètres], par exemple*). Ces **[détecteurs thermiques]** (*le principe est la mesure d'un échauffement produit par le rayonnement optique*) ont une sensibilité spectrale très large (*le pouvoir d'absorption de la lumière n'est pas lié à la longueur d'onde*) et peuvent être utilisés pour des missions très diverses. Simplicité d'emploi et économie d'un système de refroidissement sont les deux avantages principaux de ces **nouveaux [instruments infrarouges]**. Par contre, leur sensibilité est **plus faible que celle des [instruments refroidis]** utilisant des **[détecteurs quantiques]**. (TTVS2002)

Ces contextes indiquent l'importance de la question du refroidissement pour la détection infrarouge et ce, dès 1994. Cependant, si les progrès envisagés en 1994 impliquent l'utilisation de machines cryogéniques, on voit dans l'édition 2002 que la solution finalement envisagée à cette époque est toute autre : plutôt que d'envisager des refroidissements très puissants à l'aide des machines cryogéniques actives, en 2002, la nouvelle solution est d'avoir recours à des systèmes dits « non-refroidis ».

Un second exemple de suivi dans le temps est celui de la technologie BiCMOS :

- (lx) *En ce qui concerne la [technologie bi-CMOS], ce n'est qu'en 1993 que nous venons d'avoir accès aux logiciels [...] de conceptions de circuits.* (TTVS1994)
- (lxi) *Aujourd'hui, [...], ce sont les technologies [CMOS] qui font l'objet d'améliorations incessantes. [...] On constate que les technologies [BiCMOS] ne sont qu'une amélioration des technologies [CMOS] existantes.* (TTVS2002)

Ce qui est intéressant dans ce cas est que ces contextes permettent de dater l'apparition des BiCMOS au Cnes en 1993, date où l'accès aux logiciels de conception a été permis, mais ils permettent également de proposer un premier retour sur ces technologies qui, bien que potentiellement intéressantes pour le spatial, ne sont présentées en 2002 que comme des « améliorations » des CMOS (et non pas comme les successeurs des CMOS, comme on aurait pu l'envisager à la lecture de ce dernier contexte).

L'analyse de chacun des contextes fournit donc des informations intéressantes et centrales sur l'évolution des connaissances, et permet entre autres de poser des « jalons » dans la chronologie d'un domaine pour retracer son évolution. Ces contextes mettent en jeu des informations variées et très riches qu'il est possible cependant de compléter par un second type d'analyse : l'observation des termes qui ont tendance à se regrouper dans ces contextes.

3.4.3. Observations de termes récurrents en contextes

Comme nous l'avons défini au début de cet article, les contextes riches en connaissances évolutives contiennent des marqueurs d'évolution, ainsi que des termes du domaine. Il est donc possible de recenser les termes qui tendent à apparaître fréquemment dans ces contextes. Ces termes peuvent être triés de deux manières :

- un tri en fonction de la valeur d'évolution exprimée dans le contexte où ils apparaissent,
- un tri en fonction des sous-corpus dans lesquels ils apparaissent dans ces contextes.

Le premier tri, bien que prometteur, s'avère cependant très délicat à mener. En effet, dans un même contexte d'évolution peuvent coexister différentes valeurs d'évolution et il n'est pas toujours facile d'interpréter clairement une seule évolution pour un terme donné. Par exemple, dans le contexte :

- (lxii) *A la vue des résultats très encourageants obtenus, trois programmes aujourd'hui envisagent la conception des chaînes vidéo en [ASIC]. Il s'agit de la nouvelle génération des [satellites d'observation de la terre] (Spot5 et HELIOS2) et de l'instrument [IASI], [interféromètre] chargé de l'analyse de l'[atmosphère] à des fins météorologiques qui a la particularité d'être demandeur d'électronique (conçue en ASIC) à très faible consommation devant fonctionner à 100 K.*

certaines satellites sont marqués comme nouveaux (les satellites Spot5 et HELIOS2) mais – malgré l'intervention nécessaire des experts - il reste difficile d'interpréter avec précision l'évolution liée au terme/concept *ASIC* : s'agit-il d'un concept récent ? Les ASIC sont-ils connus depuis longtemps, mais disponibles seulement récemment dans le spatial ? Ont-ils connu des améliorations qui entraînent ces résultats encourageants ? Si oui, lesquelles ? De la même manière, *atmosphère* est un terme du domaine extrait par Syntex et présent dans ce contexte auquel on ne peut lier aucune évolution.

Pour cette raison, nous adoptons la seconde possibilité de tri et observons les termes récurrents dans les contextes riches en connaissances évolutives dans chacun des trois sous-corpus du TTVS.

Ce type d'observation permet de repérer par exemple qu'en 1998, des termes liés au projet Silex sont récurrents dans les contextes d'évolution : *constellations de satellites, constellations, liaison intersatellite, polynômes*. Ces termes permettent de mettre en avant l'importance de ce projet dans l'évolution de l'optique au milieu des années 1990.

En 2002, on rencontre la thématique des miroirs et la question des parasites. Ainsi, les termes *arêtes, couches minces optiques, couche mince, miroir asphérique, miroir plan, miroir primaire* ou encore *flux parasite, image parasite* sont récurrents dans les contextes évolutifs du TTVS 2002.

Cependant, ces termes n'ont pas tous le même statut. En effet, dans cet exemple, les termes *miroir asphérique, miroir plan, miroir primaire* ne sont jamais les « sujets » de l'évolution, mais plus des termes/concepts anciens auxquels on fait référence dans ces contextes parce qu'ils sont liés à une évolution. Par exemple, dans le TTVS 2002 :

(lxiii) *Depuis quelques années, la technique d'[usinage ionique] sans contact avec le [miroir] est opérationnelle. Un faisceau d'ions vient arracher la matière. Malheureusement, cette technique ne peut être utilisée qu'en phase finale car un faisceau de trop grande énergie provoque un échauffement important du [miroir] ; ainsi le taux d'arrachement de matière est faible. Toutefois, elle offre l'avantage d'être très déterministe et laisse entrevoir la possibilité de corriger les défauts de [quilting] générés par les phases précédentes. S'affranchir des contraintes de [quilting] dans le [dimensionnement] est la solution pour un gain de masse précieux pour les grands [miroirs] envisagés dans le futur.*

Dans ce contexte, on compte 3 occurrences du terme *miroir*. Cependant, le terme/concept *miroir* n'y est pas présenté comme cible de l'évolution. Ce contexte présente plutôt des évolutions liées à l'usinage ionique (donc au polissage) des miroirs. De fait, ce n'est pas directement le concept *miroir* dont on décrit l'évolution, mais plutôt celle des techniques qui lui sont associées. Cependant, la présence de ce terme n'est pas anodine dans ces contextes puisqu'il s'agit d'un terme/concept ancien que l'on rappelle pour aborder la question des progrès dans lesquels il est impliqué. Autrement dit, des termes/concepts connus ou anciens qui gravitent dans certains contextes d'évolution pourraient être considérés comme des marqueurs « indirects » de l'évolution d'autres termes/concepts.

4. Remarques conclusives et perspectives

Dans cet article, nous avons mis en avant le potentiel de techniques d'extraction d'information pour la détection de contextes riches en connaissances évolutives. Malgré certaines difficultés à caractériser et définir les marqueurs (section 2), ce type de démarche présente des intérêts méthodologiques certains pour observer l'évolution des connaissances à partir de textes.

Le premier est l'adaptabilité de cette approche à différents types d'informations recherchés et la diversité des informations que l'on cherche à extraire (nouveau, obsolescence, datation, amélioration, etc.). Ceci en fait une approche très souple, caractéristique très riche pour entreprendre une description de l'évolution des connaissances d'un domaine.

De plus, les contextes permettent de repérer des évolutions sur au moins deux types différents d'éléments : les concepts du domaine et leurs propriétés. Par exemple dans un contexte tel que :

(lxiv) Ces **nouvelles technologies** n'ont **pas encore** réellement fait leur entrée dans le spatial, car elles sont plus **récentes** que celle des [semi-conducteurs]. Elles ont donc encore à prouver leurs potentialités dans ce domaine. Mais il est certain que leur **évolution** est à suivre de près pour les **futures applications spatiales**, où faibles [encombrement] et [masse], faible [consommation électrique] et bas coût seront des caractéristiques déterminantes. (TTVS1998)

on extrait à la fois des informations sur l'évolution de concepts (les nouvelles technologies citées en début de paragraphe et les semi-conducteurs), ainsi que sur les propriétés recherchées dans les futures applications spatiales : faible encombrement, faible masse, faible consommation, bas coût. Cet élément est important dans la mesure où ces deux types d'informations sont susceptibles d'être centraux dans certaines applications pour lesquelles on met en place un repérage de l'évolution. Par exemple, dans le cas de la mise à jour de ressources termino-ontologiques, il peut s'avérer important de repérer à la fois les évolutions des concepts, mais également de leurs propriétés. Néanmoins, il reste difficile de nuancer ces deux aspects à ce stade dans la mesure où ils sont marqués de manière semblable dans les contextes.

Le deuxième intérêt méthodologique de cette approche est que l'analyse des contextes s'avère une étape riche pour commencer l'exploration de l'évolution d'un domaine pour lequel on ne dispose que de peu de connaissances. En effet, l'observation de contextes est relativement facile à mettre en oeuvre (dépendamment de la finesse de l'extraction souhaitée), et permet à l'analyste de s'immerger directement et globalement dans les textes et dans le domaine. Il lui est alors facile de dégager des éléments dans la chronologie du domaine par déduction, même s'il n'est pas expert de ce domaine et s'il ne dispose que de peu d'informations *a priori*. L'analyse manuelle de ce type de contextes, que ce soit en synchronie (dynamique) ou en diachronie, lui permet également de travailler à partir de pans entiers de textes qui traitent explicitement de l'évolution du domaine observé et de s'immerger ainsi directement dans les corpus et les connaissances.

Il reste cependant beaucoup de perspectives pour affiner cet indice, la première étant la définition d'autres marqueurs. Dans cette optique, il serait particulièrement intéressant de prendre en compte la dimension discursive et la structure des textes (Laignelet, 2009 ; Péry-Woodley, 2000 ; Toussaint, 2004). En effet, il apparaît clairement dans nos corpus que certaines sections des textes sont propices à l'extraction de contextes, en particulier les introductions, les conclusions et les sections dont le titre contient lui-même un marqueur d'évolution. De plus, sous réserve de les caractériser plus précisément, on pourrait envisager de mettre au jour des configurations de marqueurs en fonction des évolutions à repérer (Ho-Dac & Péry-Woodley, 2008 ; Laignelet, 2009 ; Péry-Woodley, 2000).

Enfin, nous espérons avoir montré l'intérêt de chercher à observer l'évolution des connaissances sur de courts intervalles temporels. La diachronie, et *a fortiori* la diachronie courte, est en effet une perspective encore très marginale en terminologie. Or, à la lecture des résultats obtenus dans cette recherche, il apparaît clairement que l'évolution en corpus spécialisés est un phénomène très présent et multi-facettes pour lequel le besoin de description est patent (Dury/Picton à paraître). Il est donc nécessaire de continuer à décrire, développer des méthodologies et des indices et apprendre à théoriser au mieux la question de la dimension diachronique en langue de spécialité (Picton 2009).

5. Bibliographie

- Aussenac-Gilles, N./Séguéla, P. 2000 : Les relations sémantiques : du linguistique au formel. In *Cahiers de Grammaire, « Sémantique et Corpus »* 25, 175-198.
- Barrière, C. 2001 : Investigating the Causal Relation in Informative Texts. In *Terminology* 7(2), 135-154.
- Barrière, C./Agbago, A. 2006. TerminoWeb: A Software Environment for Term Study in Rich Contexts. In *Actes de la conférence internationale "Terminology, Standardisation and Technology Transfer" (TSTT 2006)*, Beijing, Chine.

- Borillo, A. 1996 : Exploration automatisée de textes de spécialité : repérage et identification automatique de la relation lexicale d'hyponymie. In *Linx* 34-35, 113-121.
- Bowker, L./Pearson, J. 2002: *Working with Specialized Language: a Practical Guide to Using Corpora*. London/New York : Routledge.
- Charlet, G./Zacklad, M./Kassel, G./Bourigault, D. (eds.) *Ingénierie des connaissances, évolutions récentes et nouveaux défis*. Paris : Eyrolles, 225-241.
- Cnes 1994, 1998, 2002 : *Cours de techniques et technologies des véhicules spatiaux*. Cnes : Cépaduès Éditions.
- Condamines, A. 2003a : *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*. Carnet de Grammaire, Rapports Internes de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), Habilitation à Diriger les Recherches : Toulouse 2, vol. 13.
- Condamines, A. 2003b : Vers la définition de genres interprétatifs. In *Actes de la conférence "Terminologie et Intelligence Artificielle" (TIA 2003)*, Strasbourg, France.
- Condamines, A./Rebeyrolle, J. 2000 : Construction d'une base de connaissances terminologiques à partir de textes : expérimentation et définition d'une méthode. In Condamines, A./Rebeyrolle, J./Soubeille, A. 2004 : Variation de la terminologie dans le temps : une méthode linguistique pour mesurer l'évolution de la connaissance en corpus. In *Actes de la conférence "European Association for Lexicography" (EURALEX 2004)*, Lorient, France.
- Dury, P./Picton, A. (à paraître) : Terminologie et diachronie : vers une réconciliation théorique et méthodologique? In *Revue Française de Linguistique Appliquée*.
- Guilbert, L. 1975 : *La créativité lexicale*. Larousse Université : Paris.
- Hearst, M. 1992 : Automatic Acquisition of Hyponyms from Large Texts Corpora. In *Actes du 15ème colloque international en linguistique informatique (Coling 1992)*, Nantes, France.
- Hearst, M. 1997: TextTiling: Segmenting Text into Multi-Paragraph Subtopic Passages. In *Computational Linguistics* 23(1), 33-64.
- Ho-Dac, L. M./Péry-Woodley, M.-P. 2008. Méthodologie exploratoire outillée pour l'étude de l'organisation du discours. In *Actes du Congrès Mondial de Linguistique Française (CMLF 2008)*, Paris, France.
- Houdebine, A.-M. 1985 : Pour une linguistique synchronique dynamique. In *La Linguistique* 21, 7-36.
- Ibekwe-Sanjuan, F. 2005 : Repérage et annotation d'indices de nouveautés dans les écrits scientifiques. In *Actes du colloque "Indice, Index, Indexation"*, Université Charles de Gaulle, Lille, France.
- Laignelet, M. 2009 : *Associer analyse syntaxique et analyse discursive pour le repérage d'informations obsolètes dans les documents encyclopédiques*. Thèse de doctorat en Sciences du Langage. Université Toulouse 2, France.
- Laignelet, M./Pimm, C. (2007) : Utiliser une segmentation thématique TextTiling pour le repérage de segments d'information évolutive dans un corpus de textes encyclopédiques. In *Actes de la 14ème conférence "Traitement Automatique des Langues Naturelles" (TALN'2007) (communication affichée)*, Toulouse, France.
- Leroy, S. 2004 : Extraire sur patrons : allers et retours entre analyse linguistique et repérage automatique. In *Revue Française de Linguistique Appliquée* IX(1), 25-43.
- L'Homme, M.-C./Marshman, E. 2006. Terminological Relationships and Corpus-based Methods for Discovering them: An Assessment for Terminographers. In Bowker, L. (ed.) *Lexicography, Terminology, and Translation. Text-based studies in honour of Ingrid Meyer*. Ottawa : University of Ottawa Press, 67-80.
- Malaisé, V./Zweigenbaum, P./Bachimont, B. 2005: Mining Defining Contexts to Help Structuring Differential Ontologies. In *Terminology* 11(1), 21-53.
- Marshman, E. 2006 : *Lexical Knowledge Patterns for Semi-automatic Extraction of Cause-effect and Association Relations from Medical Texts: A Comparative Study of English and French*. Thèse de doctorat en Traduction, option Terminologie. Université de Montréal, Canada.
- Méla, A. 2004 : Linguistes et 'talistes' peuvent coopérer : repérage et analyse des gloses. In *Revue Française de Linguistique Appliquée* IX(1), 63-82.
- Meyer, I. 2001: Extracting Knowledge-Rich Contexts for Terminography : A Conceptual and Methodological Framework. In Bourigault, D./L'Homme, M.-C./Jacquemin, C (eds.), *Recent Advances in Computational Terminology*. Amsterdam/New York : John Benjamins Publishing Company, 279-302.
- Otman, G. 1996 : *Les Représentations sémantiques en terminologie*. Paris : Masson.
- Péry-Woodley, M.-P. 2000 : *Une pragmatique à fleur de texte : approche en corpus de l'organisation textuelle*. Carnets de Grammaire, Rapports Internes de l'ERSS (Équipe de Recherche en Syntaxe et Sémantique), Habilitation à Diriger les Recherches : Toulouse 2, vol. 8.

- Picton, A. 2009 : Diachronie et langues 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, France.
- Suarez, M. /Cabré, M.T. 2002 : Terminological Variation in Specialized Texts : Linguistic Traces for Automatic Retrieval. In *Actes du VIIIème Symposium ibéroaméricain de terminologie (RITERM)*, Carthagène des Indes, Colombie.
- Toussaint, Y. 2004 : Extraction de connaissances à partir de textes structurés. In Pinon, J.-M./Rumpler, B. (eds.), *Fouille de textes et organisation de documents*. Paris : Documents numériques, Hermès 8(3), 11-34.