Document généré le 29 jan. 2025 19:56

Meta
Journal des traducteurs
Translators' Journal

META

Mise en oeuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie

Anne Condamines et Nathalie Dehaut

Volume 56, numéro 2, juin 2011

Les corpus et la recherche en terminologie et en traductologie
Corpora and Research in Terminology and Translation Studies

URI : https://id.erudit.org/iderudit/1006176ar
DOI : https://doi.org/10.7202/1006176ar

Aller au sommaire du numéro

Éditeur(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (imprimé)
1492-1421 (numérique)

Découvrir la revue

Citer cet article

Condamines, A. & Dehaut, N. (2011). Mise en oeuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie. *Meta*, *56*(2), 266–283. https://doi.org/10.7202/1006176ar

Résumé de l'article

L'utilisation des méthodes de linguistique de corpus est particulièrement pertinente pour analyser la terminologie en prenant en compte la réalité des fonctionnements tels qu'on les appréhende dans les textes. La première partie de l'article présente les arguments en faveur de ces méthodes. Elles se focalisent sur le lexique, mais permettent aussi d'étudier d'autres phénomènes linguistiques, syntaxiques, sémantiques ou discursifs. En fait, l'intérêt de la terminologie textuelle est l'étude des usages réels en prenant en compte les contextes linguistiques, mais aussi extralinguistiques. Ainsi, la terminologie textuelle propose une méthode adaptée non seulement pour définir les termes, mais aussi pour répondre à d'autres objectifs. De cette manière, l'article explique comment les méthodes de linguistique de corpus peuvent être adoptées dans une situation d'interdisciplinarité, concernant l'émergence d'une nouvelle discipline : l'exobiologie. Nous montrons comment trois types d'indices (formel, quantitatif et distributionnel) sont mis en oeuvre pour repérer des phénomènes de polysémie ou de synonymie et des cas d'emprunts. Ces mêmes types d'indices sont utilisés pour mettre en lumière dans quels cas ces phénomènes sont conscients ou non, et surtout pour repérer les cas où ils peuvent être une source de conflit. Dans une telle étude, le rôle des experts est crucial, afin de valider les résultats, mais surtout parce qu'ils en sont les utilisateurs. Enfin, nous pensons, grâce à cette présentation du panorama des phénomènes linguistiques dans un corpus interdisciplinaire concernant l'exobiologie, contribuer à la définition de cette nouvelle discipline.

Tous droits réservés © Les Presses de l'Université de Montréal, 2011

Ce document est protégé par la loi sur le droit d'auteur. L'utilisation des services d'Érudit (y compris la reproduction) est assujettie à sa politique d'utilisation que vous pouvez consulter en ligne.

https://apropos.erudit.org/fr/usagers/politique-dutilisation/

érudit

Cet article est diffusé et préservé par Érudit.

Érudit est un consortium interuniversitaire sans but lucratif composé de l'Université de Montréal, l'Université Laval et l'Université du Québec à Montréal. Il a pour mission la promotion et la valorisation de la recherche.

https://www.erudit.org/fr/

Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie*

ANNE CONDAMINES
Université de Toulouse 2 – Le Mirail, Toulouse, France
anne.condamines@univ-tlse2.fr

NATHALIE DEHAUT
Université de Toulouse 2 – Le Mirail, Toulouse, France
Centre national d'études spatiales (CNES), Toulouse, France
nathalie.dehaut@univ-tlse2.fr

RÉSUMÉ

L'utilisation des méthodes de linguistique de corpus est particulièrement pertinente pour analyser la terminologie en prenant en compte la réalité des fonctionnements tels qu'on les appréhende dans les textes. La première partie de l'article présente les arguments en faveur de ces méthodes. Elles se focalisent sur le lexique, mais permettent aussi d'étudier d'autres phénomènes linguistiques, syntaxiques, sémantiques ou discursifs. En fait, l'intérêt de la terminologie textuelle est l'étude des usages réels en prenant en compte les contextes linguistiques, mais aussi extralinguistiques. Ainsi, la terminologie textuelle propose une méthode adaptée non seulement pour définir les termes, mais aussi pour répondre à d'autres objectifs. De cette manière, l'article explique comment les méthodes de linguistique de corpus peuvent être adoptées dans une situation d'interdisciplinarité, concernant l'émergence d'une nouvelle discipline : l'exobiologie. Nous montrons comment trois types d'indices (formel, quantitatif et distributionnel) sont mis en œuvre pour repérer des phénomènes de polysémie ou de synonymie et des cas d'emprunts. Ces mêmes types d'indices sont utilisés pour mettre en lumière dans quels cas ces phénomènes sont conscients ou non, et surtout pour repérer les cas où ils peuvent être une source de conflit. Dans une telle étude, le rôle des experts est crucial, afin de valider les résultats, mais surtout parce qu'ils en sont les utilisateurs. Enfin, nous pensons, grâce à cette présentation du panorama des phénomènes linguistiques dans un corpus interdisciplinaire concernant l'exobiologie, contribuer à la définition de cette nouvelle discipline.

ABSTRACT

The methods used in corpus linguistics are very relevant in order to analyze how terminology works within texts. The first part of the paper puts forward arguments explaining how these methods are relevant. These methods focus on the lexicon, but also allow to study other phenomena related to linguistics, syntax, semantics or discourse. In fact, the main point in textual terminology is that what is studied is actual uses, since the analysis takes into account both linguistic and extra-linguistic contexts. Textual terminology is therefore relevant not only for defining terms, but also for achieving other kinds of aims. The paper presents how those methods may be used in a multidisciplinary situation dealing with the emergence of a new field, i.e. exobiology. We show how three types of clues (formal, quantitative and distributional) are used in order to identify polysemy, synonymy or loanwords. The approach is also relevant to identify whether these phenomena are deliberate, and when they can be a source of conflict. In such a study, the role of

experts is crucial, not only in order to validate the results but also because they are end users of the results. Finally, we hope that by presenting them with a wide range of linguistic phenomena in a multidisciplinary corpus dealing with exobiology, we can contribute to the definition of this new field.

MOTS-CLÉS/KEYWORDS

exobiologie, interdisciplinarité, sémantique terminologique, rôle des experts, terminologie textuelle
exobiology, multidisciplinarity, terminological semantics, role of experts, textual terminology

1. Introduction

L'étude de la terminologie en situation d'innovation scientifique est riche d'enseignements. En effet, elle oblige à s'intéresser non seulement à la néologie lexicale ou sémantique mais aussi aux liens entre cette néologie et l'innovation conceptuelle. C'est le cas avec l'exobiologie (l'étude de la vie sur les planètes hors du système solaire). Cette « néodiscipline » a pour particularité de s'inscrire dans une interdisciplinarité revendiquée par les différentes disciplines impliquées : chimie, biologie, physique, astronomie, bien qu'il n'y ait encore ni spécialiste ni étudiant en exobiologie.

Notre intervention s'est faite via le Centre national d'études spatiales (CNES). Il s'agit pour nous de comparer les productions de ces disciplines, particulièrement du point de vue lexical, pour faire apparaître les similitudes et les différences dans une double perspective. D'une part, repérer d'éventuelles difficultés de communications entre ces disciplines et, d'autre part, aider les futurs exobiologistes à repérer des « zones de sens » particulièrement denses dont on peut faire l'hypothèse que s'y joue aussi de la créativité conceptuelle. En effet, dans cette situation, les phénomènes qui peuvent apparaître comme ambigus ou vagues pour des linguistes sont peut-être à identifier comme « en évolution » et donc tout à fait intéressants pour les chercheurs.

Dans cette situation, l'utilisation des méthodes de la linguistique de corpus est particulièrement pertinente. En effet, ces méthodes mettent en œuvre une linguistique outillée sur des corpus qui, très souvent, sont organisés en sous-corpus dont le fonctionnement est comparé. Mais il est clair que les résultats doivent être analysés et interprétés au regard du contexte inédit de cette étude.

La première partie de l'article rappelle les enjeux de la linguistique de corpus et les questions qui la traversent tout en montrant sa pertinence pour l'étude de la terminologie, particulièrement dans ce cas d'intense innovation conceptuelle. Dans la deuxième partie, nous présentons comment le corpus a été constitué et comment les différents indices fournis par les outils (quantitatifs, formels et distributionnels) sont utilisés pour l'analyse de ce corpus.

2. Linguistique de corpus et terminologie : une rencontre naturelle ?

Les liens entre la linguistique de corpus et la terminologie pourraient paraître difficiles si l'on s'en tenait à une vision prescriptive de la terminologie, surtout focalisée sur la relation entre concept et terme. Dans une telle vision, impulsée par Wüster, le terme est et doit être maintenu en dehors des discours (écrits ou oraux) dans lesquels il est susceptible d'être utilisé et la terminologie en dehors de la linguistique de la langue générale, trop éloignée d'une vision prescriptive.

> En voulant effectuer le transfert inadmissible à la terminologie des expériences accumulées dans le domaine de la langue commune, on a pendant des années opposé à la normalisation de cette dernière la devise suivante : « la langue ne se laisse pas normaliser » (Wüster 1981 : 65).

Pourtant, une des raisons majeures de l'évolution de la terminologie vers une vision plus descriptive, ces dernières décennies, est venue en grande partie de plus ou moins grands échecs dans les tentatives d'utiliser les terminologies (normatives) existantes pour traiter (archiver, interroger, traduire…) des textes, et tout particulièrement de la documentation d'entreprise. Une réflexion, plus théorique, de la part de lexicologues, a conduit par ailleurs au constat d'une grande artificialité dans le fait de vouloir maintenir la terminologie en dehors du giron de la linguistique, même si on peut reconnaître aux termes quelques spécificités par rapport aux mots « généraux ». Enfin, dans le même temps, la linguistique s'est trouvée fortement secouée par la mise à disposition de volumes très importants de « données textuelles » via internet ou les intranets. Comment prendre en compte ces usages, qui apparaissent parfois comme « déviants » par rapport à l'intuition linguistique ? Ces données doivent-elles constituer l'objet de la linguistique ou bien le linguiste doit-il se contenter de son introspection comme le pensait Chomsky ? Bref, la rencontre entre terminologie et linguistique de corpus ne semblait pas se faire sur un terrain parfaitement balisé où la terminologie n'aurait eu qu'à emprunter des outils théoriques et méthodologiques déjà bien établis. Pourtant, s'il est un domaine où il y a un sens à mettre en œuvre une linguistique de corpus, c'est bien celui de l'étude de la terminologie (Condamines 2005). Nous allons nous employer à expliquer pourquoi dans cette première partie.

2.1. Pourquoi utiliser les méthodes de la linguistique de corpus pour étudier la terminologie ?

Trois éléments semblent plaider en faveur d'une grande pertinence à la mise en œuvre de la linguistique de corpus pour étudier la terminologie.

2.1.1. Corpus

Pour le linguiste, définir un corpus, qui va être son objet d'étude, n'est pas chose facile. En effet, à la suite de Saussure, la linguistique s'est donné comme objet la langue. Il ne s'agit pas d'étudier des usages particuliers, *a priori* infinis, mais bien de dégager ce qui fait système, les règles qui sous-tendent le fonctionnement langagier, d'une langue, voire de toutes les langues. Or, avec un corpus, on n'est plus dans un fonctionnement potentiel mais bien dans la réalité des usages, nécessairement corrélés à des situations de communication particulières. Cette confrontation à la réalité des usages produit un « effet de réel », qui peut être déstabilisant pour le linguiste « introspectif ».

Définir un corpus oblige à circonscrire un objet clos et donc à se donner des règles pour garantir un minimum de représentativité à cet objet, cette représentativité étant elle-même liée à un objectif d'étude. On a ainsi défini un corpus comme : « a subset of [Electronic Text Library] built according to explicit design criteria for a specific purpose […] » (Atkins, Clear, *et al.* 1992 : 1).

Or, une des premières caractéristiques attribuées à la terminologie, même dans sa vision la plus prescriptive, est que toute terminologie est associée à un domaine.

Cette caractéristique constitue une première manifestation (mais aussi une première limite) de ce qui n'est, pour la linguistique introspective, qu'une potentialité de la langue : on n'étudie que les usages relevant d'un certain domaine, prédéfini. Cette caractéristique, qui peut paraître banale, entraîne en fait un changement de paradigme. En effet, même si, les terminologues le savent, cette caractéristique s'avère peu opérationnelle (comment décider de quel domaine relève tel ou tel texte et comment être sûr que le domaine à étudier est bien représenté par les textes sélectionnés [Gaudin 1995]), c'est quand même une première façon de prendre en compte la situation de communication, ce qui, dans les années 1980, n'était considéré en linguistique que comme relevant d'une linguistique de seconde zone, au mieux de la sociolinguistique. Non seulement la langue est ainsi considérée dans son usage mais, qui plus est, la situation de communication est première puisque c'est en fonction d'elle que se constitue le corpus. Ainsi, d'une certaine façon, la terminologie était bien plus à même d'intégrer la notion de corpus puisqu'elle acceptait déjà de donner une contrainte d'usage à son champ d'investigation. D'ailleurs, on perçoit bien ce lien avec la sociolinguistique, y compris dans les appellations données aux approches alternatives aux visées prescriptives : socioterminologie (Gaudin 1995), approche socio-cognitive (Temmerman 2000), terminologie textuelle (Slodzian 2000), etc.

2.1.2. Variation

Prendre en compte la notion de domaine constitue une première délimitation de l'objet d'étude, qui est d'emblée reliée à la prise en considération de la variation (la variation terminologique serait liée à la répartition en domaines). Or, cet élément est une des caractéristiques majeures de la linguistique de corpus. On ne peut pas en effet traiter des volumes de données textuelles, parfois considérables, comme un « sac de mots ». Il faut essayer de les ordonner et cette organisation passe par la prise en compte de données extralinguistiques. D'où un développement significatif, en linguistique de corpus, des recherches en lien avec la notion de genre textuel, qui permet de corréler données extralinguistiques et fonctionnements langagiers (Biber 1988 ; Bahtia 1993 ; Swales 1990). La « terminologie textuelle », alternative à une vision trop prescriptive de la terminologie, a été elle aussi confrontée à la nécessité de prendre en compte la variation due aux éléments extralinguistiques et à ne pas la limiter au seul domaine. D'une part, nous l'avons dit, la notion de domaine est difficile à manier et, d'autre part, certaines variations peuvent être liées à bien autre chose que le domaine, par exemple le niveau d'expertise des locuteurs et des interlocuteurs, l'objectif de la communication, etc. (Pearson 1998).

La variation est étudiée en linguistique de corpus à travers la comparaison de sous-corpus constitués sur la base d'éléments extralinguistiques et surtout d'une hypothèse de travail. Cette comparaison, dont les résultats sont ensuite interprétés en fonction de l'objectif de l'étude, est ainsi un des éléments majeurs de cette approche.

Par ailleurs, au contact de la linguistique de corpus et du traitement automatique des langues (TAL), un autre type de variation est apparu comme déterminant pour la terminologie textuelle, il s'agit de la variation en lien avec des besoins d'utilisation des résultats, c'est-à-dire avec l'objectif de l'étude qui peut se décliner en études théoriques mais aussi en besoins applicatifs réels (archivage, représentation des connaissances, traduction…).

2.1.3. Lexique

Paradoxalement, la linguistique de corpus est d'abord une linguistique des mots ou, en tout cas, dans laquelle les études sont basées sur les mots, plus exactement sur les chaînes de caractères. Cette importance accordée au mot, souvent interrogée par les tenants d'une linguistique du texte (Rastier 2001), a été renforcée par l'utilisation quasi incontournable d'outils. En effet, et c'est une autre de ses caractéristiques majeures, la linguistique de corpus est une linguistique outillée (Habert 2004). Cet élément n'est pas intrinsèquement lié au fait qu'un corpus ne peut être travaillé qu'à l'aide d'outils (nous considérons qu'il peut d'ailleurs y avoir des corpus qui ne sont pas électroniques), mais au fait que la plupart des corpus étudiés sont volumineux et que le simple dénombrement de formes peut être largement assisté par un outil. Toutefois, plusieurs remarques doivent être faites. D'une part, les termes ne sont pas seulement des mots, mais bien plus souvent des groupes de mots (avec les variantes qu'ils peuvent comporter), voire des morphèmes, c'est-à-dire des parties de mots. D'autre part, dans la linguistique de corpus, si « l'entrée » dans les corpus se fait par les mots, les phénomènes étudiés vont très souvent au-delà des mots et concernent des aspects syntaxiques, discursifs, sémantiques… La même situation apparaît avec les corpus spécialisés : les mots peuvent être considérés comme un accès à la « profondeur » des textes, cette dimension textuelle pouvant avoir une pertinence plus grande que le seul fonctionnement des mots, cela en fonction des besoins. C'est le cas avec l'étude sur la constitution d'une nouvelle discipline (l'exobiologie) qui est présentée ci-dessous.

2.2. Linguistique de corpus pour la terminologie : quels enjeux, quelles méthodes ?

Acceptons donc l'hypothèse et la nécessité de la mise en place d'une linguistique de corpus pour l'étude de la terminologie textuelle et voyons comment elle peut se mettre en place.

2.2.1. Outils

Il ne s'agit pas ici de faire un descriptif approfondi de tous les types d'outils utiles pour la linguistique de corpus spécialisés, mais d'en brosser plutôt un rapide panorama pour préciser les méthodes sous-tendant ces outils et les biais éventuels.

Les outils utilisés pour travailler sur le contenu des corpus sont essentiellement les concordanciers. Les corpus peuvent être étiquetés ou non, avec analyse syntaxique ou non ; dans le premier cas, il n'y a pas désambiguïsation et toutes les étiquettes grammaticales possibles sont conservées pour chaque chaîne de caractères. Ainsi, *ferme* est étiqueté comme un nom, comme un verbe ou comme un adjectif. Ces outils, permettant de faire des interrogations sur la base d'expressions régulières, constituent une possibilité très intéressante d'accéder aux fonctionnements langagiers, ne serait-ce que pour étudier la distribution d'un mot. *In fine*, c'est toujours vers eux que le linguiste-terminologue se tourne, préférant souvent même les outils qui proposent le moins possible d'interprétation pour avoir un accès direct aux données.

Pour les corpus spécialisés, des outils beaucoup plus « dédiés » ont été conçus : extracteurs de termes et extracteurs de relations. Le développement de ces outils a

bénéficié de l'essor de la constitution d'ontologies à partir de textes qui, en ingénierie des connaissances, est devenu un enjeu majeur (Bourigault et Jacquemin 2000). Avec prudence, la plupart des concepteurs parlent plutôt de « termes candidats » et de « relations candidates », ce qui souligne à la fois le fait que les outils peuvent « faire des erreurs », mais aussi qu'il n'y a pas une définition consensuelle de « terme ». Ces outils fonctionnent sur deux approches majeures : la reconnaissance de forme (et la répartition de ces formes) d'une part et les statistiques d'autre part, les deux étant assez souvent combinées. Il faut reconnaître que la définition de ces outils ne s'est pas toujours accompagnée d'une réflexion très poussée sur la manière dont ils peuvent être réellement utilisés ni sur la façon dont les résultats proposés interagissent avec les connaissances des linguistes/terminologues ou des experts du domaine et encore moins sur ce que les résultats obtenus peuvent apporter comme connaissances sur le fonctionnement de la langue. Souvent peu au fait des études linguistiques, les informaticiens ne se penchent que rarement sur des questions de sémantique qui sont pourtant au cœur du fonctionnement langagier, y compris dans les textes spécialisés. Beaucoup pensent ainsi que la sémantique se réduit à l'établissement de liens entre des termes. Enfin, très peu d'outils prennent en compte la variation d'un corpus à l'autre en fonction d'éléments extralinguistiques alors que des travaux se développent dans ce sens en terminologie, que ce soit en ce qui concerne la variation des termes ou des marqueurs de relations conceptuelles, en fonction du genre textuel (Condamines 2008 ; Marshman, L'Homme, *et al.* 2008 ; Rogers 2000). Le problème principal des outils dédiés est certainement dû au fait qu'ils sont développés alors même que l'étude du fonctionnement des termes est elle-même en plein essor. Il faut donc, lorsqu'ils sont utilisés pour la recherche en terminologie, savoir quelles en sont les limites, ce qui suppose d'avoir un minimum de compréhension des principes sur lesquels ils fonctionnent.

2.2.2. Définir un objectif d'étude

Ainsi que nous l'avons vu, le côtoiement de la terminologie textuelle avec le TAL, et surtout l'ingénierie des connaissances, a fait apparaître la nécessité de définir un objectif pour l'étude d'un corpus spécialisé à travers les termes. Cette nécessité existe déjà pour la linguistique de corpus « en général » mais pour la linguistique de corpus spécialisé, elle devient incontournable. Elle permet en effet non seulement de guider l'interprétation des éléments langagiers mais aussi de définir le corpus et de donner un cadre de validation aux experts.

2.2.3. Utiliser des indices mais les interpréter : décorréler les résultats d'outils de l'interprétation

Les résultats proposés par les outils d'analyse de corpus doivent être considérés comme autant d'indices qu'il faudra évaluer, mettre en balance et interpréter en fonction de l'objectif de l'étude. Compte tenu des possibilités des outils, qui ne manipulent que des formes, et de la nature comparative de la linguistique de corpus, ces indices peuvent concerner trois types de variations entre sous-corpus : quantitatives, formelles ou distributionnelles.

a) Variation quantitative

La variation quantitative concerne le fait qu'un phénomène langagier peut être bien plus présent dans un sous-corpus que dans un autre. De manière plus sophistiquée, les méthodes quantitatives s'appuient souvent sur des recherches statistiques, ce qui permet de prendre en compte le fait qu'un phénomène est significativement plus présent dans un ou l'autre sous-corpus. Il faut alors donner un sens à ce fonctionnement.

b) Variation de forme

Il s'agit de variation dans la forme des mots (des chaînes de caractères), éventuellement des groupes de mots. Ainsi, peuvent être rapprochés des mots au singulier *vs* pluriel, des mots ayant la même base, des groupes avec ou sans une certaine forme (par exemple avec ou sans déterminant)…

c) Variation de distribution

Cette variation concerne la variation dans les contextes dans lesquels apparaissent les mots ou les groupes de mots. Cet indice vise plus nettement à prendre en compte le sens. Il prend appui en effet sur le point de vue bloomfieldien selon lequel un changement de distribution est le signe d'un changement de sens. Beaucoup d'outils de TAL, s'inspirant des travaux de Harris sur le distributionalisme, fonctionnent sur cette base.

2.2.4. Prendre en compte les experts tout au long de l'étude

Une particularité de la linguistique de corpus spécialisés est que, dans la plupart des cas, elle nécessite de collaborer avec des experts du domaine, tout au moins de les faire intervenir aussi souvent que possible dans le processus qui va de la définition de l'objectif jusqu'à la restitution des résultats, en passant par la constitution du corpus et la validation des interprétations.

3. Terminologie dans le cadre de l'émergence d'une nouvelle discipline : l'exobiologie

Cette partie va nous amener à montrer comment les méthodes de la linguistique de corpus peuvent être mises en œuvre dans une situation somme toute assez rare : l'émergence d'une nouvelle discipline, l'exobiologie. Nous montrerons comment cette problématique tire bénéfice des réflexions existant en linguistique de corpus, ce qui suppose d'adapter les résultats d'outils et de les intégrer à une réflexion plus générale, essentiellement d'ordre sémantique, voire épistémologique, qui concerne l'idée d'innovation.

L'exobiologie étudie l'émergence de la vie sur Terre et la possibilité d'une vie en dehors de la Terre. Il s'agit d'une discipline en cours d'élaboration, qui ne concerne pas le simple côtoiement de plusieurs disciplines mais bien la mise en commun délibérée de différentes disciplines pour en constituer une nouvelle, en lien avec l'évolution des connaissances sur la vie. Cette particularité conduit à tenir compte d'éléments importants :

a) Il n'y a pas à proprement parler d'expert en exobiologie

Compte tenu de la situation d'innovation, il n'y a pas réellement d'expert en exobiologie ni de formation unifiée sur ce sujet mais des experts des différentes disciplines concernées, persuadés de la nécessité de constituer cette nouvelle science. L'étude des corpus à partir de la terminologie peut ainsi contribuer à définir cette nouvelle science.

b) Pertinence de la comparaison

Le contexte dans lequel émerge l'exobiologie rend très pertinente l'étude comparative. Les disciplines concernées sont à la fois assez proches pour qu'il soit justifié de les comparer et elles ont un point de vue suffisamment distinct pour qu'on puisse imputer les différences langagières à ces différences de points de vue.

c) Importance de la conscience vs non-conscience des phénomènes langagiers

Dans le processus de création scientifique en jeu dans la constitution d'une nouvelle science, la dimension langagière joue un rôle crucial : « […] la réception du nouveau est une dimension par elle-même, qui rejoint l'aventure historique du sens » (Stengers et Schlanger 1991 : 94). C'est dans la différence de sens, infime ou majeure, accordé à tel ou tel terme et dans l'accord possible ou non sur des définitions que se forge, au moins en partie, la création d'idées neuves. Les différences de sens d'une discipline à l'autre sont parfois clairement perçues ; mais parfois, le travail du linguiste-terminologue permet d'aider à les mettre au jour. Le linguiste-terminologue peut ainsi jouer un rôle de maïeuticien, en tout cas de diagnostiqueur, qui peut contribuer à éclairer l'émergence de nouvelles idées.

3.1. Présentation de l'étude et du corpus

L'étude que nous sommes en train de mener consiste à mesurer l'impact d'une situation interdisciplinaire sur la terminologie d'un domaine. Que ce soit dans des domaines techniques ou scientifiques, il est très courant de voir différentes disciplines collaborer sur un objet commun d'étude.

> La nécessité de l'interdisciplinarité est devenue une quasi-évidence dans notre société où la spécialisation est la règle : il n'est pratiquement aucun problème concret qui puisse recevoir une solution appropriée sans faire appel à diverses spécialités et à divers spécialistes. Les pratiques interdisciplinaires sont indispensables puisque les approches monodisciplinaires ne fournissent que rarement des réponses appropriées à des questions concrètes (Fourez, Mathy, *et al.* 1993 : 119).

Notre préoccupation consiste à établir une sorte de panorama du fonctionnement lexical dans cette situation. Ce panorama peut être utilisé pour repérer des différences qui peuvent être interprétées soit comme des difficultés (problème de communication), soit comme des lieux d'émergence de nouveaux sens, voire de nouvelles conceptualisations scientifiques.

Le domaine de l'exobiologie constitue un champ d'investigation particulièrement pertinent puisqu'il est reconnu comme étant nécessairement interdisciplinaire.

Comme nous l'avons déjà signalé dans la première partie de cet article, il n'existe pas à proprement parler d'expert en exobiologie, donc il n'existe pas non plus de

textes d'exobiologie ou même de revue d'exobiologie. Ce sont des chimistes, des biologistes, des géologues ou des astronomes qui se placent sous la bannière de l'exobiologie parce qu'ils ont un point commun de recherche : l'origine de la vie sur Terre et la recherche de la vie ailleurs.

Grâce à des discussions avec les experts de chaque discipline concernée, les textes pouvant constituer le corpus ont pu être identifiés ainsi que les personnes pouvant jouer le rôle d'expert dans notre étude. Finalement, le choix du corpus s'est arrêté sur des manuels qui ont été écrits après des écoles CNRS. Ces écoles rassemblent des chercheurs qui situent leurs recherches dans le domaine de l'exobiologie. Le corpus est composé de deux ouvrages : *L'environnement de la Terre primitive* (2001)[1], *Les traces du vivant* (2003)[2]. Ces ouvrages sont constitués de 38 articles de 40 auteurs différents.

L'intérêt des manuels, outre qu'ils relèvent tous d'un seul genre est aussi que, on le sait, ils sont particulièrement riches en marqueurs définitoires du fait même qu'ils s'adressent à des lecteurs moins experts que les rédacteurs.

Une des contraintes imposées par l'utilisation d'outils de traitement automatique est le format du fichier puisque ces logiciels fonctionnent généralement avec des fichiers en format (.txt). Le fait de devoir changer de format de fichier a une conséquence sur le contenu des textes. Par exemple, dans ce cas précis, nous avons dû passer d'un format PDF à un format (.txt). Ceci a eu pour conséquence la suppression des images, des tableaux et le passage de leurs légendes ainsi que des notes de bas de page à l'intérieur du corps du texte. Ces éléments ont dû être retrouvés et placés en fin de chapitre afin de garder une cohérence textuelle et de ne pas fausser la recherche de contextes d'un terme. Il s'agit ici d'un premier biais qui nous a été imposé par les outils.

Le deuxième traitement effectué sur le corpus a été son organisation en sous-corpus. La constitution de sous-corpus par discipline a été guidée par l'objectif de l'étude puisqu'il s'agit de repérer des différences et des similitudes lexico-sémantiques entre les disciplines concernées par l'exobiologie. Là encore, les experts ont joué un rôle crucial puisqu'ils nous ont permis de déterminer les disciplines représentées majoritairement en exobiologie. Quatre disciplines ont donc été retenues : l'astronomie, la biologie, la chimie et la géologie. Les articles constituant le corpus ont donc été répartis parmi ces quatre disciplines, d'après la discipline d'origine des rédacteurs, pour former quatre sous-corpus. La répartition obtenue est présentée dans le tableau 1.

Tableau 1

Répartition des articles en sous-corpus

Discipline	N (articles)	N (mots)
Astronomie	12	88 815
Biologie	8	65 589
Chimie	8	88 190
Géologie	8	77 010

Le corpus a été lemmatisé et étiqueté par TreeTagger[3]. Ce logiciel effectue, en premier lieu, un étiquetage morpho-syntaxique pour ensuite pouvoir associer un lemme à chaque mot du texte. À savoir que certains mots (nombre, mot étranger, mot inconnu) ne peuvent pas être lemmatisés et qu'aucune information ne leur est associée.

3.2. Les indices utilisés

Nous avons expliqué dans la première partie de cet article la pertinence d'utiliser une linguistique de corpus pour les études en terminologie, une linguistique de corpus qui implique, souvent, une linguistique outillée. Comme nous l'avons vu, ces outils, compte tenu de leurs possibilités, influencent le choix des indices que nous allons utiliser. Ces indices peuvent être de trois natures : quantitatifs, formels ou distributionnels.

En linguistique de corpus, il est courant de situer son approche comme relevant d'une méthode guidée par des hypothèses (*hypothesis-driven*), par opposition à une méthode guidée par des données (*data-driven*) (Leech 1992 ; Biber, Johansson, *et al.* 1999 ; Tognini-Bonelli 2001 ; Rayson 2002). Ces approches se différencient par rapport à la place que l'on va donner aux hypothèses linguistiques. Dans la première approche, le corpus va être utilisé comme moyen pour vérifier des hypothèses linguistiques « introspectives ». Dans la seconde approche, le corpus va servir d'objet d'étude qui permettra de définir des phénomènes linguistiques. Lors d'études terminologiques précédentes, nous avions repris cette double possibilité pour caractériser le fonctionnement des outils d'extraction terminologique en utilisant les termes « approches descendantes » (*top-down*) (faisant intervenir des connaissances *a priori* sur le fonctionnement terminologique) par opposition aux « approches ascendantes » (*bottom-up*) dans lesquelles les fonctionnements sont conçus comme « émergeant » du corpus (Condamines et Rebeyrolle 1997).

Pour cette étude, nous n'avons pas voulu opter pour une approche plutôt que pour une autre et nous avons préféré combiner les deux approches puisque chacune apporte son lot d'avantages et d'inconvénients. En effet, notre approche est avant tout « guidée par l'objectif » (pour nous, repérer des différences de fonctionnement lexico-sémantiques). Il est clair que nous essayons de rester au plus près des phénomènes qui apparaissent dans le corpus et d'en faire émerger les spécificités. De ce point de vue, l'approche guidée par des hypothèses est plus adaptée.

> The problem with [hypothesis-driven] approach is that during the investigation, we can search only for evidence, or lack of evidence, for what we expect to find. The alternative to hypothesis-driven research is *data-driven* research, in which we are informed by the corpus data itself and allow it to lead us in all sorts of directions, some of which we have never thought of (Rayson 2002 : 1).

Dans le même temps, nous ne nous interdisons pas d'utiliser des connaissances *a priori* sur la langue, dont nous faisons l'hypothèse qu'elles ont un intérêt pour l'étude en cours. Plus précisément, nous savons qu'à certaines formes langagières peuvent être associées de manière régulière des interprétations ; nous faisons l'hypothèse que ce fonctionnement se retrouve dans le corpus étudié.

C'est pourquoi nous divisons nos indices selon ces deux approches : les indices guidés par des hypothèses et les indices guidés par des données.

3.2.1. Les indices guidés par des données : indices quantitatifs et première catégorie d'indices de variation de distribution

Les indices quantitatifs nous permettent de calculer les spécificités de chacun des sous-corpus et de repérer des termes communs aux sous-corpus. Les spécificités sont les mots les plus caractéristiques d'un sous-corpus et non pas les mots les plus fréquents

de celui-ci. Afin de pouvoir les identifier, une comparaison a été effectuée entre un sous-corpus et l'ensemble des autres sous-corpus. La méthode des mots-clefs, dont on se sert pour rechercher les spécificités, est implémentée dans le logiciel WordSmith Tools[4] que nous avons utilisé dans cette étude. Après suppression des hapax, après avoir filtré les mots grammaticaux, noms propres et noms les plus fréquents de la langue générale, nous avons recensé 151 formes spécifiques pour l'astronomie, 178 pour la biologie, 248 pour la chimie et 167 pour la géologie.

Étant donné qu'un des objectifs de l'étude est de déceler des différences (ou ressemblances) dans l'emploi d'un terme selon les disciplines, la recherche des termes communs à au moins deux disciplines s'est avérée indispensable afin de pouvoir travailler sur leurs distributions et de repérer d'éventuelles polysémies : « le caractère monosémique ou polysémique d'une unité linguistique se caractérise par des contextes sémantiquement homogènes, si elle est monosémique, ou sémantiquement hétérogène si elle est polysémique » (Bertels, Speelman, *et al.* 2006 : 76).

Pour le dire autrement, il s'agit de repérer des similarités ou dissimilarités sémantiques à partir des similarités ou dissimilarités distributionnelles, par le classement et la catégorisation des contextes linguistiques (Condamines et Rebeyrolle 1997 ; Teubert 2001). Il s'agit comme le dit Habert (2005) de dégrouper les sens : « c'est trouver le moyen de repérer les cas où un mot en cache un, voire plusieurs autre(s) » (Habert 2005 : 279).

Par exemple dans notre corpus, nous avons pu remarquer, lors de nos analyses, que le terme *accrétion* apparaît davantage en géologie et en astronomie dans une structure [préposition + *accrétion*] (*pendant/au début/après l'accrétion…*), structure nettement absente des deux autres disciplines (biologie et chimie). Cette observation a mis en évidence que le terme *accrétion* était utilisé en géologie et en astronomie dans le sens de *processus* tandis que dans les autres disciplines, il était utilisé comme un *résultat* (avec des syntagmes comme *matière d'accrétion* ou *disque d'accrétion*, où ici *accrétion* est équivalent à *matière accrétée*).

3.2.2. Les indices guidés par des hypothèses : marqueurs liés à l'objectif de l'étude, contextes définitoires complets et incomplets (marqueurs de relations conceptuelles)

Un de ces premiers indices guidés par des hypothèses est, dans le cadre de ce travail, la recherche de marqueurs liés directement à l'objectif de l'étude. C'est-à-dire, dans notre cas, des marqueurs qui montrent la conscience de l'auteur de la possible ambiguïté du terme qu'il utilise. Ces marqueurs sont parfois recensés comme marqueurs de glose (Julia 2001). Des structures comme *au sens de [telle discipline]* ou encore *en [telle discipline]* ont ainsi été recherchées.

(1) Il y a peu de concepts en chimie physique macroscopique qui soient aussi abstraits que le concept d'entropie ; il y a peu de concepts qui soient si souvent mal utilisés, tout particulièrement en biologie et de manière plus générale dès qu'il s'agit de décrire des processus irréversibles conduisant à des états plus structurés que l'état de départ.

(Gargaud, Despois, *et al.* 2001 : 332)

(2) De nombreux auteurs ont émis l'hypothèse que la tectonique des plaques ne fonctionnait pas pendant l'Archéen et donc il a même été proposé que, compte tenu

des forts flux thermiques, il n'y ait pas eu de lithosphère archéenne (dans le sens rhéologique du terme).

(Gargaud, Despois, *et al.* 2001 : 279)

Dans le même ordre d'idées, nous faisons l'hypothèse que le contexte interdisciplinaire amène une sorte de procédé de « vulgarisation » ou au moins de « didactique » interdisciplinaire, ce phénomène étant renforcé par le genre des textes constituant le corpus (des manuels), nous pouvons ici citer les propos de Delavigne (2003 : 83) : « La collaboration d'experts d'origines diverses engage des partages de compétences et des négociations discursives qui ne sont pas sans rapport avec la problématique de la vulgarisation ».

Aussi, dans la perspective d'un repérage automatique des marqueurs, nous avons décidé de nous concentrer sur un procédé caractéristique des discours de vulgarisation : les énoncés définitoires. Nous avons procédé à un relevé des « termes définis » afin de pouvoir les observer dans les différents sous-corpus. Les formes que peut prendre la définition en discours ont fait l'objet d'une analyse détaillée (Rebeyrolle 2000) que nous résumons ici et que nous illustrons par des exemples de notre corpus « exobiologie » :

a) les énoncés définitoires de désignation

(3) On observe dans ce cas un phénomène assimilable à de la résolution cinétique (Kagan *et al.*, 1988) où le terme « résolution » désigne un processus permettant de séparer des énantiomères.

(Gargaud, Despois, *et al.* 2003 : 88)

b) les énoncés définitoires de dénomination

(4) La région située au-dessus de la tropopause se nomme la stratosphère (par analogie avec la structure thermique de la Terre).

(Gargaud, Despois, *et al.* 2001 : 181)

c) les énoncés définitoires de signification

(5) Si G0' a une valeur négative, la réaction est dite « exergonique », ce qui signifie qu'il se produit une libération d'énergie libre que la cellule pourra conserver sous forme d'ATP.

(Gargaud, Despois, *et al.* 2003 : 244)

d) les énoncés définitoires parenthétiques

(6) Le taux de cette décroissance est donc lié au taux « d'astration » (la vitesse de formation des étoiles).

(Gargaud, Despois, *et al.* 2001 : 80)

e) les énoncés définitoires introduits par *c'est-à-dire*

(7) Cette préservation est cependant partielle dans le cas des espèces volatiles, c'est-à-dire des espèces entrant dans des composés se condensant à relativement basse température, et les chondrites sont appauvries en volatils (H, C, N, S, gaz rares) de plusieurs ordres de grandeur par rapport à la NPS.

(Gargaud, Despois, *et al.* 2001 : 201)

Une des façons de repérer des différences d'un corpus à l'autre (d'une discipline à l'autre dans notre cas) consiste à repérer des structures conceptuelles différentes. La méthode la plus courante de repérage des relations conceptuelles est celle de l'utilisation des marqueurs de relations qui constituent des portions de contextes définitoires (Alarcon Martinez 2009). Par exemple grâce à des structures comme *un terme X [tel que/comme] Y, Z, W*, nous pouvons observer si, dans chaque sous-corpus, un hyperonyme reprend les mêmes hyponymes.

Ainsi, dans

(8) […] éléments volatils comme les chondrites carbonées. (géologie)
(Gargaud, Despois, *et al.* 2003 : 40)

(9) […] espèces volatiles comme l'hélium, le néon, l'azote, les molécules organiques complexes et l'eau de constitution des phases hydratées. (astronomie)
(Gargaud, Despois, *et al.* 2001 : 102)

on voit apparaître ce qui pourrait être considéré comme deux structures différentes avec le même hyperonyme (élément/espèce volatil(e)). Afin de s'assurer de la pertinence de cette différence, il faut évidemment, d'une part, ne pas s'en tenir à ces seuls contextes et vérifier si d'autres contextes ne viennent pas compléter les structures et, d'autre part, faire intervenir des experts pour valider ces structures. En effet, ce n'est pas parce que le corpus ne fait pas mention des toutes les relations hyperonymiques possibles que ces relations ne sont pas reconnues et admises dans la discipline.

3.3. Résultats

L'analyse des résultats fournis par les différents indices mis en œuvre nous a permis de repérer différents types de phénomènes :

- des phénomènes de type sémantique : synonymie, polysémie, emprunt « complet » ;
- des fonctionnements concernant la conscience que les locuteurs ont de l'interdisciplinarité sur leur production langagière, c'est-à-dire de la polysémie et de la synonymie ;
- des fonctionnements sur les conséquences de ces phénomènes sémantiques : conflictuels ou non conflictuels.

3.3.1. Phénomènes sémantiques

3.3.1.1. Synonymie

Certaines disciplines emploient des termes différents pour référer au même concept. Ainsi, nous pouvons donner le cas des termes *exoplanète* et *planète extrasolaire* pour désigner une planète qui ne fait pas partie du système solaire. Dans le sous-corpus « astronomie », les deux termes apparaissent quasiment à la même fréquence (*exoplanète* : 19 fois ; *planète extrasolaire* : 15 fois). Tandis que dans le corpus « chimie », nous ne trouvons que le terme *exoplanète* (3 fois) et, à l'opposé, dans le corpus biologie, nous n'avons que des occurrences du syntagme *planète extrasolaire* (2 fois).

Deux types d'indices nous permettent de proposer cette hypothèse de la synonymie : une similarité de forme et une similarité de distribution. En effet, les deux termes comportent la forme *planète* et des préfixes qui, l'un en grec (exo-) l'autre en latin (extra-), signifient /en dehors de/. Nous trouvons aussi une similarité de contextes d'usage :

(10) Signatures spectroscopiques de vie sur les exoplanètes.
(Gargaud, Despois, *et al.* 2003 : 343)

(11) Le cas de la Terre et de sa biosphère durant la première partie de son histoire prouve que la non-détection de O2 ou de O3 ne pourra en aucun cas permettre de conclure à l'absence de vie sur une exoplanète.
(Gargaud, Despois, *et al.* 2003 : 351)

(12) […] la détection de la vie sur les planètes extrasolaires ne peut se faire qu'en analysant les signaux que nous recevons (télédétection).
(Gargaud, Despois, *et al.* 2001 : 467)

(13) […] chercher la vie sur les planètes extrasolaires par la détection de raies d'oxygène dans leur spectre.
(Gargaud, Despois, *et al.* 2001 : 468)

(14) […] Parmi les projets de télescopes spatiaux visant à la détection d'exoplanètes telluriques […].
(Gargaud, Despois, *et al.* 2003 : 344)

(15) Dans un premier temps, d'ici 2015-2020, on « verra » par détection optique directe les premières planètes extrasolaires telluriques […].
(Gargaud, Despois, *et al.* 2001 : 471)

3.3.1.2. Polysémie

Le second cas est celui de la polysémie : un terme qui a différents sens, plus ou moins proches, selon la discipline qui l'utilise. Le principal indice pour repérer ces cas est celui de la différence de distribution avec le principe déjà cité plus haut : un contexte sémantiquement homogène marque la monosémie, un contexte sémantiquement hétérogène marque la polysémie comme le cas d'accrétion présenté en (*3.2.1.*).

La polysémie peut aussi être repérée grâce à des marqueurs distributionnels interprétés *a priori*, grâce auxquels l'auteur signale sa conscience de ce phénomène.

(16) Du point de vue chimique, une oxydation est définie comme la perte d'un ou plusieurs électrons par un atome ou une molécule. Une réduction est, elle, définie comme l'addition d'un ou plusieurs électrons à un atome ou une molécule.
(Gargaud, Despois, *et al.* 2003 : 244)

3.3.1.3. Emprunt

Cette catégorie concerne le cas où un terme/concept a été emprunté tel quel à une discipline par les autres disciplines. Compte tenu de la nature « synchronique » de notre corpus, il n'est pas possible de repérer des cas où l'emprunt se serait fait de manière non consciente. Il faudrait pour cela constituer un corpus « diachronique » prenant en compte des états de langue antérieurs pour chaque discipline afin de repérer quelle discipline est à l'origine de tel ou tel terme/concept. Même en élargissant le corpus, il serait difficile d'avoir des textes suffisamment anciens pour « tracer » le cheminement d'un terme et identifier le moment de son passage d'une discipline à l'autre. Seuls les emprunts conscients peuvent donc être repérés.

3.3.2 Consciences vs non-conscience, conflit vs cohabitation

Les phénomènes conscients sont repérés grâce à des marqueurs spécifiques (interprétés *a priori*). Ainsi, pour les emprunts, il existe des marqueurs utilisés et utilisables.

Notamment, dans l'exemple suivant d'emprunt, le terme *inerte* est toujours accompagné de *chimiquement* dans les corpus « astronomie » et « géologie », mais il est utilisé seul dans le corpus « chimie ».

(17) Les gaz rares des planètes (Ne, Ar, Kr, et Xe) sont chimiquement inertes. (astronomie)

(Gargaud, Despois, *et al.* 2001 : 142)

(18) Cependant, si cet appauvrissement est particulièrement marqué dans le cas des gaz rares, qui sont chimiquement inertes et donc peu retenus dans les silicates et le métal […]. (géologie)

(Gargaud, Despois, *et al.* 2001 : 201)

(19) Une fois récupérés sous atmosphère inerte, ces produits sont soumis à différentes analyses afin de déterminer leurs propriétés optiques, leur solubilité dans différents solvants, leur structure moléculaire […]. (chimie)

(Gargaud, Despois, *et al.* 2003 : 374)

Pour une des occurrences présentes dans le corpus « biologie », il est difficile de dire si *inerte* est utilisé dans le sens chimique ou non :

(20) De plus, ces travaux sont antérieurs à la publication de Darwin (1859) sur l'origine des espèces, et, même après 1859, seuls les esprits les plus avancés s'autorisaient à penser que la vie avait évolué à partir de la matière inerte. (biologie)

(Gargaud, Despois, *et al.* 2001 : 387)

Les phénomènes non conscients sont repérés par l'analyse des contextes distributionnels.

Les fonctionnements conflictuels ou potentiellement conflictuels peuvent être eux aussi reconnus comme tels (marqueurs), comme dans :

(21) […] en chimie physique macroscopique qui soient aussi abstraits que le concept d'entropie ; il y a peu de concepts qui soient si souvent mal utilisés, tout particulièrement en biologie et de manière plus générale dès qu'il s'agit de décrire des processus irréversibles conduisant à des états plus structurés que l'état de départ.

(Gargaud, Despois, *et al.* 2001 : 332)

Mais ils peuvent ne pas être conscients et conduire à des incompréhensions ou des discussions sans fin. Dans ce cas, à nouveau, c'est l'analyse des distributions qui permet de repérer les phénomènes sémantiques et de les soumettre aux experts.

Si l'on combine les trois types de fonctionnements, on devrait obtenir 12 cas de figure possibles comme le montre le tableau suivant. Mais, ainsi que nous l'avons souligné, les emprunts non conscients ne sont pas détectables dans le corpus actuel.

Tableau 1
Impact de l'interdisciplinarité sur la terminologie d'un domaine

	Conscient	Non conscient	Conflictuel	Non conflictuel
Polysémie	x		x	
	x			x
		x	x	
		x		x
Synonymie	x		x	
	x			x
		x	x	
		x		x
Emprunt	x		x	
	x			x
		x	x	
		x		x

3.3.3. Rôle des experts

Certains exemples, notamment l'exemple 20, montrent le rôle essentiel des experts dans ce type d'étude. En effet, ceux-ci vont nous aider à faire les bonnes interprétations pour des cas qui nous posent problème et, surtout, ils vont nous permettre de définir si les phénomènes repérés sont conflictuels ou non.

Les outils nous permettent d'avoir accès à des indices – indices que nous choisissons parce qu'ils sont repérables grâce aux outils. L'interprétation de ces indices se construit au fur et à mesure de l'analyse des exemples repérés grâce aux outils qui nous amènent vers d'autres exemples ou d'autres indices à rechercher. À cela vient s'ajouter le dialogue avec les experts du domaine qui nous permet de valider nos intuitions par rapport au corpus.

Le rôle des experts est donc primordial, de la constitution du corpus à l'interprétation des résultats. De plus, s'ils nous apportent une aide incontestable, cette relation n'est pas unilatérale, puisque au cours des échanges, nous apportons des éléments de réflexion aux experts, surtout dans le cas précis de cette étude qui concerne l'émergence d'une nouvelle discipline et la volonté de construire une interdisciplinarité. En effet, les résultats présentés aux experts et discutés avec eux leur permettent de mettre en lumière des phénomènes *a priori* inexistants pour eux parce que jusque-là inconscients. Le dialogue entre le linguiste et l'expert constitue donc un enrichissement pour les deux parties grâce à un véritable échange.

4. Conclusion

Cette étude est encore en cours. D'une part, tous les résultats n'ont pas été dépouillés et il serait nécessaire, comme nous l'avons dit, d'étendre le corpus. D'autre part, même si une première réunion a permis de travailler sur une première série de résultats, il reste à présenter l'ensemble de ces résultats aux experts afin d'affiner la méthodologie et, surtout, de prendre en compte leur réaction pour affiner la méthode de mise au jour des phénomènes. Avec cette étude, nous sommes au cœur d'une problématique qui dépasse des éléments strictement langagiers puisqu'elle concerne la création scientifique.

Nous sommes loin d'un point de vue normatif sur la terminologie, qui n'a aucune pertinence ici, en tout cas à ce stade de l'étude. Il se peut que dans un second temps

nous participions à la définition des termes, ce qui permettra de stabiliser les concepts (au moins pour un temps). Ici, la terminologie est un point d'entrée dans l'étude des fonctionnements sémantiques. La focalisation sur les aspects lexicaux permet d'étudier ces fonctionnements de manière systématique. Il est clair que cette possibilité nouvelle de percevoir la terminologie et ses rapports avec les textes a été ouverte par le développement de la linguistique de corpus. Elle dessine des perspectives tout à fait passionnantes pour l'intervention des linguistes-terminologues dans le champ de l'innovation scientifique.

NOTES

* Cette étude est financée, d'une part, sous la forme d'une bourse doctorale conjointement par le CNES et le CNRS et, d'autre part, par le Programme interdisciplinaire du CNRS « Origine des planètes et de la vie ».

1. Gargaud, Muriel, Despois, Didier et Parisot, Jean-Paul (2001) : *L'environnement de la terre primitive*. Bordeaux : Presses universitaires de Bordeaux.
2. Gargaud, Muriel, Despois, Didier, Parisot, Jean-Paul, *et al.* (2003) : *Les traces du vivant*. Bordeaux : Presses universitaires de Bordeaux.
3. Développé par Helmut Schmid, ICLUS (Institute for Computational Linguistics of the University of Stuttgart).
4. Développé par Mike Scott, Oxford University.

RÉFÉRENCES

Alarcon Martinez, Rodrigo (2009) : *Descripción y evaluación de un sistema basado en reglas para la extracción automática de contextos definitorios*. Thèse de doctorat non publiée. Barcelone : Universitat Pompeu Fabra.

Atkins, Sue, Clear, Jeremy et Ostler, Nicholas (1992) : Corpus Design Criteria. *Literary and Linguistic Computing*. 7(1):1-16.

Bahtia, Vijay Kumar (1993) : *Analysing Genre – Language Use in Professional Settings*. London : Longman.

Bertels, Ann, Speelman, Dirk et Geeraerts, Dirk (2006) : Analyse quantitative et statistique de la sémantique dans un corpus technique. *In* : *Actes de la 13e Conférence sur le traitement automatique des langues naturelles*, 73-82.

Biber, Douglas (1988) : *Variation Across Speech and Writing*. Cambridge : Cambridge University Press.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, *et al.* (1999) : *Grammar of Spoken and Written English*. London : Longman.

Bourigault, Didier et Jacquemin, Christian (2000) : Construction de ressources terminologiques. *In* : Jean-Marie Pierrel, dir. *Ingénierie des langues*. Paris : Hermes, 215-233.

Condamines, Anne (2005) : Linguistique de corpus et terminologie. *Langages*. 157(1):36-47.

Condamines, Anne (2008) : Taking Genre into Account for Analyzing Conceptual Relation Patterns. *Corpora*. 3(2):115-140.

Condamines, Anne et Rebeyrolle, Josette (1997) : Point de vue en langue spécialisée, *Meta*. 42(1):174-184.

Delavigne, Valérie (2003) : Quand le terme entre en vulgarisation. *In* : C*onférence TIA* (Strasbourg, 31 mars et 1er avril 2003). Consultée le 10 mai 2011, <http://tia.loria.fr/TIA/IMG/pdf/TIA2003.pdf>.

Fourez, Gérard, Mathy, Philippe et Engiebert-Lecomte, Véronique (1993) : Un modèle pour un travail interdisciplinaire. *ASTER*. 17:119-142.

Gaudin, François (1995) : Champs, clôtures et domaines : Des langues de spécialités à la culture scientifique. *Meta*. 40(2):229-238.

HABERT, Benoît (2004) : Outiller la linguistique : de l'emprunt de techniques aux rencontres de savoirs. *Revue française de linguistique appliquée*. 9(1):5-24.

HABERT, Benoît (2005) : Des décalages de distribution aux divergences d'acception. *In* : Anne CONDAMINES, dir. *Sémantique et corpus*. Paris : Lavoisier/Hermes-Sciences, 277-318.

JULIA, Catherine (2011) : *Fixer le sens ?* Paris : Presses Sorbonne Nouvelle.

LEECH, Geoffrey (1992) : Corpora and Theories of Linguistic Performance. *In* : Jan SVARTVIK, dir. *Directions in Corpus Linguistics*. Berlin/New York : Mouton de Gruyter, 105-122.

MARSHMAN, Élisabeth, L'HOMME, Marie-Claude et SURTEES, Victoria (2008) : Portability of Cause-Effect Relation Markers across Specialized Domains and Text Genres : A Comparative Evaluation. *Corpora*. 3(2):141-172.

PEARSON, Jennifer (1998) : *Terms in Context*. Amsterdam/Philadelphie : John Benjamins.

RASTIER, François (2001) : *Arts et Sciences du texte*. Paris : PUF.

RAYSON, Paul E. (2002) : *Matrix : A Statistical Method and Software Tool for Linguistic Analysis through Corpus Comparison*. Thèse de doctorat non publiée. Lancaster : Lancaster University.

REBEYROLLE, Josette (2000) : *Forme et fonction de la définition en discours*. Thèse de doctorat non publiée. Toulouse : Université Toulouse-le Mirail.

ROGERS, Margaret (2000) : Genre and Terminology. *In* : Anna TROSBORG, dir. *Analysing Professional Genres*. Amsterdam/Philadelphie : John Benjamins, 3-21.

SLODZIAN, Monique (2000) : L'émergence d'une terminologie textuelle. *In* : Henri BÉJOINT et Philippe THOIRON, dir. *Le sens en terminologie*. Lyon : PUL, 61-85.

STENGERS, Isabelle et SCHLANGER, Judith (1991) : *Les concepts scientifiques*. Paris : Gallimard.

SWALES, John (1990) : *Genre Analysis, English in Academic and Research Settings*. Cambridge : Cambridge University Press.

TEMMERMAN, Rita (2000) : *Towards New Ways of Terminology Description. The Sociocognitive-Approach*. Amsterdam/Philadelphie : John Benjamins.

TEUBERT, Wolfgang (2001) : Corpus Linguistics and Lexicography. *International Journal of Corpus Linguistics*. 6(special issue):125-154.

TOGNINI-BONELLI, Elena (2001) : *Corpus Linguistics at Work*. Amsterdam/Philadelphie : John Benjamins.

WÜSTER, Eugen (1981) : L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses. *In* : Guy RONDEAU et Helmut FELBER, dir. *Textes choisis de terminologie*. Québec : GIRSTERM, Université de Laval, 55-108.