HAL
open science

Nouvelles perspectives pour la terminologie textuelle

Anne Condamines

To cite this version:

Anne Condamines. Nouvelles perspectives pour la terminologie textuelle. J. Altmanova; M. Centrella; K.E. Russo. Terminology and Discourse, Peter Lang, 2018, 978-3-0343-2415-1. 10.3726/978-3-0343-2414-4 . halshs-01899150

HAL Id: halshs-01899150

https://shs.hal.science/halshs-01899150

Submitted on 19 Oct 2018

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

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

Nouvelles perspectives pour la terminologie textuelle

Anne Condamines

Cognition, Langues, Langage, Ergonomie, (UMR 5263 du CNRS)

Université de Toulouse, CNRS, Université Jean Jaurès

Résumé : Récemment, de nouvelles possibilités pour la terminologie textuelle sont apparues, pour lesquelles les termes et les relations, initialement repérées pour construire des bases de connaissances terminologiques, constituent une entrée possible dans les textes. Ils servent alors de pivots pour l'analyse et c'est le choix et l'interprétation de leurs contextes d'apparition qui devient crucial en fonction de l'objectif de l'étude. D'un point de vue meta, le type d'analyse linguistique à mettre en place est assez similaire d'une étude à l'autre. Elle consiste à utiliser deux types de marqueurs : les uns (les marqueurs descendants), directement liés à l'objectif, sont constitués d'éléments lexico-syntaxiques qui, en langue, jouent un rôle dans l'interprétation recherchée. Les autres (les marqueurs ascendants) sont constitués par les contextes d'apparition des termes, qui sont sélectionnés et interprétés en fonction de l'objectif de l'étude. Deux exemples de travaux utilisant les méthodes de la terminologie textuelle sont présentés, l'un prend en compte l'évolution dans le temps, l'autre la pluridisciplinarité.

Abstract : Recently, new possible uses for textual terminology appeared in which terms and the relationships (spotted initially in order to build terminological knowledge bases) become a way to enter the texts. Terms play the role of pivots and the interpretation of their contexts becomes crucial according to the aim of the study. From a meta point of view, the type of linguistic analysis to be set up is rather similar from a study to the other one. It consists in using two kinds of patterns. The first ones (top-down patterns) are directly linked to the study et are composed of lexico-syntactic elements linked to the sought interpretation. The other ones (bottom-up) are constituted by the contexts in which terms appear, selected and interpreted according to the aim of the study. Two examples of works using the methods of the textual terminology are presented, the first one takes into account the evolution in time, the other one the multidisciplinarity.

Mots-clés : variation des termes, distribution des termes, marqueurs lexico-syntaxique, terminologie textuelle.

Key words : lexico-syntactic patterns, terms distribution, terms variation, textual terminology.

1 Introduction

Ce chapitre rappelle les origines de la terminologie textuelle en montrant comment des méthodes d'analyse outillée ont été mises en place pour construire des bases de connaissance terminologiques (BCT) à partir de corpus, susceptibles de répondre aux besoins émergents de gestion de la documentation (partie 2). La construction de BCT à partir de corpus se fait par l'utilisation de deux types de connaissances linguistiques : d'une part l'utilisation de marqueurs spécifiques (en l'occurrence de marqueurs de relations conceptuelles) et, d'autre part, la mise en œuvre de l'analyse distributionnelle. Or, des connaissances similaires peuvent être mises en œuvre pour d'autres types d'analyse. Les méthodes doivent alors être adaptées mais on retrouve les étapes mises en place pour la construction de BCT. Il apparaît alors que l'objectif de l'analyse, c'est-à-dire l'utilisation finale des résultats élaborés, est crucial tout au long de l'étude, depuis la constitution du corpus jusqu'à l'interprétation des résultats. D'un point de vue linguistique, ces objectifs appliqués s'interprètent le plus souvent comme la nécessité de prendre en compte la variation terminologique, quelle qu'en soit son origine : évolution dans le temps (3.1), évolution en situation pluridisciplinaire (3.2). On voit alors que la prise en compte d'un besoin externe, que l'on intègre dans l'étude comme un paramètre de la

situation extra-linguistique, permet de s'interroger (voire d'expliquer) des fonctionnements linguistiques et d'affiner la connaissance que l'on a sur le fonctionnement des termes.

2 La terminologie textuelle : origines, objectifs, méthodes

La « terminologie textuelle » a émergé au début des années 1990 essentiellement dans la perspective de construire des terminologies qui soient plus proches de la réalité des usages, et avec des méthodes plus systématiques que les terminologies précédentes. Quatre éléments ont permis la mise en place de la réflexion.

a) Dans les entreprises la gestion systématique de la documentation s'est développée, faisant appel à des outils pour améliorer la recherche d'information, l'extraction d'information, l'archivage, la traduction… Le besoin en ressources terminologiques pour alimenter ces outils est apparu comme majeur.
b) En lien avec ces besoins, le constat a été fait d'une inadéquation entre les terminologies de référence, officielles, et la réalité des usages langagiers dans les documentations techniques.
c) Les outils de traitement automatique de la langue dédiés à des tâches particulières se sont développés. Ils ne visaient plus seulement à formaliser le fonctionnement de la langue (comme c'était le cas jusqu'alors) mais à répondre à des besoins précis : étiqueteurs morpho-syntaxiques, extracteurs de termes-candidats, concordanciers…
d) La réflexion sur le rôle des corpus en linguistique et, particulièrement, en sémantique a mobilisé de nombreuses équipes de recherche.

Les besoins étaient avérés, les corpus spécialisés étaient disponibles ainsi que les outils pour les analyser et la réflexion théorique pour permettre de définir des méthodes de terminologie textuelle qui s'appuient sur des connaissances linguistiques. Le résultat le plus tangible de la terminologie textuelle a été la définition de bases de connaissances terminologiques. A partir de ce moment, les principales questions se sont concentrées sur la définition de méthodes outillées permettant d'extraire, à partir de corpus, des termes et des relations entre eux.

2.1 Le contexte théorique

Si les besoins des utilisateurs des terminologies ont joué un rôle de dynamiseur dans le développement de la terminologie textuelle, les réflexions des chercheurs en linguistique ou en traduction ont joué aussi un rôle déterminant. Il est ainsi devenu manifeste que les recommandations formulées par Wüster étaient souvent trop éloignées de la réalité du fonctionnement de la langue pour pouvoir être réellement mises en œuvre pour gérer la documentation technique. Il faut dire que la Théorie Générale de la Terminologie s'appuie sur une vision prescriptive de la terminologie, visant à rendre la communication transparente et non-ambigüe, c'est-à-dire assez éloignée du fonctionnement de la langue générale (Wüster 1981). Or, comme le signale Lerat, une langue spécialisée (même dans un domaine technique), ce n'est que l'usage d'une langue dans un domaine particulier « Une langue spécialisée est une langue naturelle considérée en tant que vecteur de connaissances particulières » (Lerat 1995 : 20). C'est donc la situation qui contraint les choix langagiers et, même lorsque cette situation est très circonscrite, il est rare que les manifestations langagières échappent complètement aux difficultés inhérentes à la langue. Il paraît ainsi difficile d'imaginer une langue qui ne serait qu'un véhicule d'information et dans laquelle la terminologie serait biunivoque : un terme/un concept, un concept/un terme.

Le point de vue wüstérien, basé sur le rôle supposé dominant des concepts dans les domaines spécialisés (L'homme 2004) a été battu en brèche par des chercheurs qui se sont, de fait, inscrits dans une vision textuelle de la terminologie, au sens d'une vision qui tenait compte de la réalité des fonctionnements langagiers. C'est le cas de J. Pearson (Pearson 1998), qui a d'emblée voulu étudier les termes dans leurs contextes d'apparition mais aussi de la socio-terminologie (Gaudin 2003), de la terminologie socio-cognitive (Temmerman 2000) ou encore de la théorie communicative de la

terminologie (Cabré 1999). Ce mouvement a aussi eu pour conséquence d'inscrire la terminologie textuelle dans une perspective sémantique. En effet, il ne pouvait plus s'agir d'étudier les termes comme étiquettes de concepts mais bien de comprendre ce qui faisait la spécificité des termes, c'est-à-dire d'éléments sémiotiques utilisés dans des situations de connaissances particulières (Slodzian 2000).

2.2 Les bases de connaissances terminologiques

C'est à I. Meyer et D. Skuce de l'Université d'Ottawa, que l'on doit le terme de « Terminological Knowledge Base » utilisé dans le cadre du projet Cogniterm (Meyer et al. 1992). Presque simultanément, des projets se sont développés dans d'autres équipes, par exemple à l'Université de Surrey (M. Roggers et K. Ahmad) (Ahmad, 1993) ou à l'Université de Toulouse (Condamines et Amsili 1993). Dans tous les cas, on retrouve trois caractéristiques dans les méthodes et objectifs proposés.

a) Le rapprochement avec l'ingénierie des Connaissances

Dans tous les cas mentionnés ci-dessus, les équipes étaient constituées de linguistes/terminologues et d'informaticiens. Le terme même de *base de connaissances terminologiques* rappelle la parenté avec les modes de représentation utilisés en ingénierie des connaissances. De fait, la proximité des représentations en terminologie (cf. b)) et celles de l'ingénierie des connaissances (appelées ontologies) est apparue assez rapidement (Aussenac-Gilles et Condamines 2001). Skuce et Meyer ont même parlé de symbiose pour évoquer les parentés entre la terminologie et l'ingénierie des connaissances (Skuce et Meyer 1991).

b) Le constat de la nécessité de choisir un format de représentation terminologique qui soit compatible avec une utilisation par des outils

Il est apparu que souvent, à partir de définitions de termes, il était possible d'envisager un mode de représentation qui, tout à la fois, soit utilisable par un outil et maintienne les éléments définitoires. Cette représentation, consiste à relier les termes par des relations dont la sémantique est fixe. Ainsi, si A est relié à B par la relation R et que C est relié à D par la même relation, alors on peut dire qu'il y a exactement la même différence (pour employer un terme saussurien) entre A et B qu'entre C et D. Le principal intérêt de ce type de représentation est qu'il permet de rendre compte, dans le cas de l'hyperonymie, de l'héritage des propriétés : ainsi, tous les fils de A héritent des propriétés de A, ce qui permet de ne pas avoir à répéter les relations pour tous les fils de A.

c) La définition de méthodes d'analyse des textes pour construire les réseaux terminologiques.

Une fois admis le fait que les réseaux terminologiques pouvaient être construits à partir de textes, l'idée a émergé que les textes à étudier pouvaient être non plus des définitions mais des corpus, c'est-à-dire des textes beaucoup plus volumineux et surtout avec des informations beaucoup moins structurées (Rebeyrolle, 2000). Compte tenu du volume de données à traiter (souvent quelques centaines de milliers de mots), il n'est pas possible d'envisager que ce processus soit fait complètement « à la main ». Du début des années 1990 date aussi le développement d'outils dédiés à la terminologie : extracteurs de termes candidats et extracteurs de relations candidates (Bourigault et al. 2001).

Le schéma 1 synthétise le processus de constitution de terminologies à partir de textes. On peut y noter un élément majeur : l'objectif de la constitution du réseau terminologique joue un rôle crucial. En effet, il permet de construire un corpus pertinent pour cet objectif (qui doit être représentatif de l'état de la connaissance) et riche en marqueurs (cf. ci-dessous) et il intervient aussi dans l'interprétation des résultats. En effet, les choix de modélisation ne sont pas basés sur les seules connaissances linguistiques mais aussi sur ce que le linguiste/terminologue sait de l'utilisation qui va être faite de la ressource. Cela signifie qu'il n'y pas une terminologie par domaine mais, potentiellement, une terminologie par domaine et par type d'application, même si une partie du réseau terminologique est certainement commun à toutes les applications.

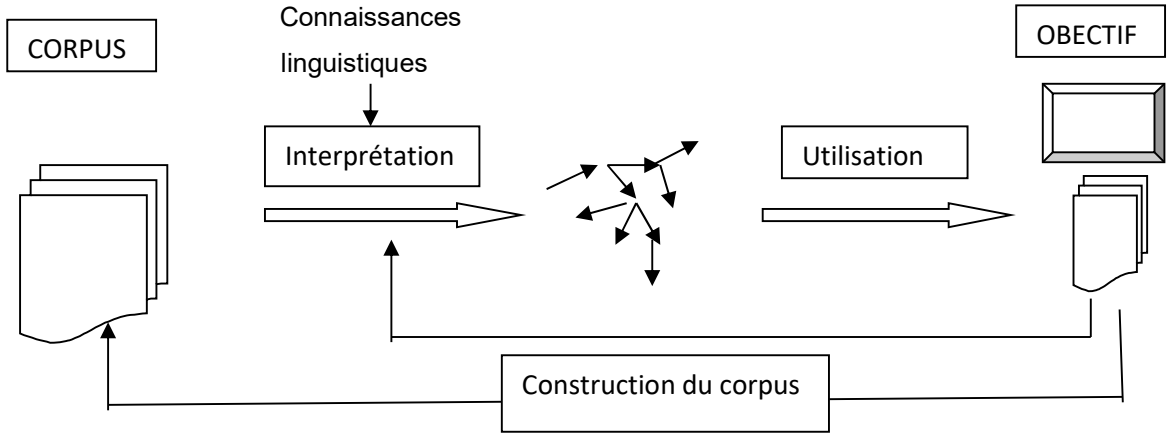

Schéma 1 : Terminologie textuelle pour la construction de réseaux terminologiques

2.3 Limites de la représentation sous forme de réseaux

L'objectif de construire un réseau terminologique à partir d'un corpus spécialisé, de manière systématique voire de manière automatique, pose un ensemble de défis qui ont conduit à la définition d'un grand nombre d'outils ou de méthodes en ingénierie des connaissances, mais assez peu à des réflexions théoriques sur les enjeux linguistiques d'un tel processus. Or, passer d'une forme discursive (linéaire) à une représentation spatiale ne va pas nécessairement de soi et peut conduire à une perte de connaissances sémantiques. Deux phénomènes doivent ainsi être pris en compte.

Décontextualisation
La construction de réseaux terminologiques passe par un phénomène de décontextualisation. Il s'agit de tenir compte de l'ensemble des contextes d'apparition et de les subsumer (généraliser) sous la forme de quelques relations, c'est-à-dire de les extraire de leur forme discursive et de leur donner une forme spatiale. Les flèches dans le réseau remplacent ainsi des formes lexico-syntaxiques. Ce processus se fait nécessairement par une interprétation qui peut générer des erreurs ou des simplifications.

Suppression d'éléments sémantiques (remplacement par une forme spatialisée)
Le choix de la représentation sous forme de relation laisse de côté un certain nombre de données sémantiques, soit parce qu'elles ne sont pas représentables avec ce format, soit parce que les relations seulement binaires ne sont pas adaptées à la représentation discursive (Aussenac-Gilles et Condamines 2009).

2.4 Le rôle des contextes linguistiques dans l'interprétation

Même assistée par des outils, la constitution de terminologies relève d'une linguistique de l'interprétation, voire d'une sémantique interprétative (Rastier 1987) qui ne dispense pas le linguiste/terminologue d'une analyse fine. A partir d'une liste de candidats-termes, il est amené à examiner les contextes d'apparition en mettant en œuvre deux types d'approches. L'une consiste à utiliser des marqueurs de relation, c'est-à-dire des éléments qui sont supposés, en langue, être associés à une relation (méthode descendante), l'autre à faire émerger l'interprétation à partir d'une généralisation des contextes d'apparition des candidats-termes (méthode ascendante).

2.4.1 Méthode descendante : Les marqueurs de relation

Les marqueurs de relation se rapprochent de ce que Meyer a appelé les « Knowledge Rich Contexts » (Meyer 2001 : 281)) : «a knowledge rich context is a context indicating at least one item of domain knowledge that could be useful for conceptual analysis ».

Dans une perspective de traitement assisté par ordinateur, un marqueur de relation est un élément ou un ensemble d'éléments lexico-syntaxique(s), voire typodispositionel(s), qui peut être utilisé pour repérer plus ou moins systématiquement en corpus, une relation (Condamines 2003). Dans l'exemple [1] ci-dessous (extrait d'un corpus sur le cancer du sein), le marqueur est l'élément *notamment* dans la structure [SN1 notamment SN2]. La relation est une hyperonymie entre N1 (*hormones*) et N2 (*œstrogènes*).

[1] Un traitement hormonal a pour but de neutraliser l'activité de *certaines* hormones *notamment* les œstrogènes chez la femme.

Dans [2] issu d'un corpus de vulcanologie, le marqueur est l'élément i*ncorpore* dans la structure [SN1 incorpore SN2] qui permet de repérer une méronymie entre croûte *plongeante* et *sédiments* et entre *croûte plongeante* et *basaltes.*

[2] Cette croûte plongeante qui *incorpore des* sédiments et *des* basaltes contenant une certaine proportion d'eau, se déshydrate d'abord.

Il faut souvent mettre en œuvre plusieurs marqueurs pour être certain d'obtenir un triplet Terme-Relation-Terme qui soit valide. Souvent en effet, soit il manque un terme soit l'interprétation n'est pas certaine (Aussenac-Gilles et Condamines 2009).

2.4.2 Méthode ascendante : Analyse de la distribution des termes.

L'approche distributionnelle, qu'elle soit associée à un point de vue behaviouriste et mathématique (Harris 1954) ou sociolinguistique (Firth 1957), s'appuie sur l'idée que l'on peut comprendre le sens d'un mot simplement en tenant compte des contextes linguistiques dans lequel il apparaît :

« Difference in meaning correlates with differences in distribution" (Harris 1954 : 43).
« You shall know a word by the company it keeps » (Firth 1957).

La terminologie à partir de corpus relève exactement de cette situation : des éléments lexicaux inconnus (ou plus ou moins connus) du linguiste/terminologue apparaissent dans des contextes connus. En outre, le linguiste/terminologue n'est pas seulement locuteur de la langue mise en œuvre mais il a aussi une connaissance de son fonctionnement qui le rend à même de proposer des interprétations.

Prenons les deux extraits suivants extraits d'un corpus sur le cancer du sein :

[3] La *production du lait* se fait au niveau des lobules glandulaires, l'*excrétion* se fait par les canaux lactifères.
[4] Un cancer est dit "in situ" lorsque les cellules cancéreuses se développent strictement à l'intérieur des canaux qui *transportent le lait* (canaux galactophores).

Plusieurs indices nous amènent à penser que *lactifères* et *galactophores* sont synonymes. Outre leur « synonymie étymologique » (qui pourrait ne pas être suffisante), plusieurs éléments du contexte (en italiques) apparaissent comme similaires dans les deux énoncés. Autant d'indices qui nous amènent à suggérer à l'expert que ces deux termes sont synonymes.

Pour conclure cette partie, on peut constater que la terminologie textuelle s'est développée initialement dans la perspective de tenir compte des usages réels et dans un contexte de développement d'une linguistique de corpus outillée. Les méthodes élaborées et les réflexions théoriques menées ont eu une autre conséquence, plus récemment apparue : la terminologie textuelle peut conduire non seulement à la constitution de bases terminologiques mais aussi à entrer, avec des méthodes systématisées, dans l'analyse de corpus spécialisés pour des objectifs très divers. Cette ouverture vers de nouvelles perspectives est détaillée dans la partie 3.

3 L'analyse des termes comme clé d'entrée dans les textes spécialisés

La terminologie textuelle s'est développée avec la perspective principale de construire des ressources utilisables par des outils et qui tiennent compte du fonctionnement réel des termes dans les discours experts. Des bases de connaissances terminologiques continuent d'ailleurs à être construites (voir par exemple le projet Eco-lexicon, (Faber et al. 2009).

Mais la définition de méthodes d'élaboration mises en œuvre pour construire les BCT a aussi permis de développer une réflexion sur le rôle des contextes, en particulier sur la notion de contexte riche en connaissance. Dans le but de constituer des réseaux terminologiques, « riche en connaissance » est équivalent à « utilisable pour construire une relation ». Or, la notion de « connaissance » peut varier selon les objectifs d'étude des corpus spécialisés. Ainsi, dans certains cas, les réseaux terminologiques ne sont pas un résultat mais plutôt un point de départ qui permet d'entrer dans l'épaisseur sémantique du corpus (Condamines et al 2011), (Condamines et al. 2012).

Ainsi, les deux méthodes d'analyse, descendante et ascendante sont également adaptées pour l'analyse de corpus spécialisés, pour d'autres types d'objectifs que la seule construction de réseaux terminologiques. Comme dans la constitution de terminologies, l'objectif de l'étude joue un rôle fondamental car il permet de guider la constitution du corpus et l'interprétation des contextes, nourrie aussi par les connaissances sur le fonctionnement de la langue. Deux exemples d'études sont donnés dans la suite de cette partie.

3.1 Exemple de l'évolution de la connaissance dans le temps

Parmi les types de variation qui peuvent affecter les termes, l'une concerne l'évolution liée au temps. On pourrait penser que ce changement se fait de manière très lente, qui ne met pas en péril la stabilité des dictionnaires terminologiques. C'est sans doute exact pour des disciplines où les connaissances évoluent lentement. Cependant, dans certains domaines, les connaissances évoluent rapidement, soit en raison d'une demande sociétale, soit en lien avec un intérêt spontané des chercheurs. C'est le cas dans un certain nombre de domaines du spatial. A la demande du CNES (Centre National d'Etudes Spatiales), nous avons travaillé à la définition d'une méthode basée sur le repérage de l'évolution des connaissances via l'analyse du fonctionnement des termes. L'objectif n'était plus ici de construire un réseau terminologique mais de repérer des variations du fonctionnement des termes dans les corpus rédigés à différentes périodes. Cette étude a fait l'objet de plusieurs conventions d'étude avec le CNES (Condamines et al. 2011) et du financement de la thèse d'A. Picton (Picton 2009).

Le corpus
Le corpus est constitué de deux sous-corpus, l'un concernant le projet DORIS (lui-même organisé en deux sous corpus, correspondant à deux périodes différentes, 36 400 mots)) et l'autre un manuel de formation interne appelé Techniques et Technologie des Véhicules Spatiaux (TTVS) (organisé en trois sous corpus, 234 000 mots).
Comme pour la construction de terminologies, deux types de contextes ont été travaillés : ceux (descendants) qui sont en lien avec l'objectif de l'étude et ceux, ascendants, qui proviennent de l'interprétation de la distribution des termes en lien avec l'objectif.

Méthode descendante : marqueurs d'obsolescence
Comme pour la recherche de relations, certains éléments langagiers peuvent être directement associés à la connaissance qui est recherchée. En l'occurrence, il s'agit de marqueurs d'obsolescence (Laignelet 2009) qui, dans le corpus le plus récent, peuvent signaler la conscience chez le locuteur d'un usage ancien, voire dépassé d'un terme. C'est le cas de l'adverbe *autrefois* et de l'adjectif *nouveau* dans les exemples [5] et [6].

[5] *Autrefois,* <u>la technologie des détecteurs</u> ne permettait pas d'avoir des barrettes ou des mosaïques.

[6] Un produit *nouveau* est apparu depuis quelques années sur le marché, il s'agit de multi barrettes.

Méthode ascendante
L'analyse comparée de la distribution des termes à différentes époques permet de repérer des évolutions. Le tableau 1 extrait de (Picton, 2009) synthétise le cas du terme « lumière parasite », qui voit sa couverture sémantique augmenter entre 1994 et 2002.

Lumière parasite	TTVS 1994	TTVS 2002
Nombre de mots dans le corpus	46 448	109 505
Nombre d'occurrences du terme	40 (0,086 %)	142 (0,129 %)
Co-occurrents les plus fréquents	Filtrage de … Piégeage de … Suppression de …	Filtrage de … Piégeage de … Suppression de … Lutter contre … Réduire… Contrôle de … Spécification de … Étude de …

Tableau 1 : Fonctionnement distributionnel de *Lumière parasite* à deux époques dans le TTVS

On note qu'en 2002, un nouvel environnement distributionnel est apparu, plus orienté vers la nécessité de limiter la *lumière parasite(lutter contre, réduire…*).

L'évolution du fonctionnement des termes dans le temps peut être due à un ensemble de phénomènes, internes à la langue et/ou liés à des évolutions de connaissances ou à la demande sociétale. Le retour vers les experts permet de comprendre et de catégoriser ces types de variations. Dans sa thèse, A. Picton a ainsi défini 17 types d'évolution terminologique dans le temps (Picton 2009).

3.2 Exemple de la variation terminologique en situation de pluridisciplinarité

Dans le cadre d'un projet financé par le Groupement De Recherche du CNRS (Centre National de la Recherche Scientifique) : « Origine des Planètes et de la Vie » et via une thèse en partie financée par le CNES, nous nous sommes intéressées à la variation terminologique dans un contexte d'innovation disciplinaire. La discipline concernée est l'exobiologie, qui s'intéresse à la vie dans les planètes hors du système solaire. Cette néo-discipline se construit au carrefour de quatre disciplines existantes : l'astronomie, la biologie, la chimie et la géologie. Ces disciplines sont caractérisées par une terminologie établie, des manuels de référence et des experts reconnus, on peut ainsi dire qu'elle constitue chacune un paradigme disciplinaire au sens kuhnien (Kuhn 1962). La mise en commun et/ou la confrontation de ces savoirs constitue un lieu d'observation particulièrement riche pour étudier les mouvements terminologiques. En effet, les usages qui sont faits des termes communs peuvent être divers et leur analyse peut permettre la mise au jour de points de vue, plus ou moins compatibles.

Le corpus
Le corpus est constitué par des articles scientifiques émanant des quatre disciplines concernées, et ayant un intérêt pour l'exobiologie (participation des auteurs à des colloques scientifiques), environs 310 000 mots au total.
Les marqueurs descendants.
Dans cette étude, les marqueurs riches en connaissance sont constitués par les éléments langagiers qui manifestent la conscience, de la part des locuteurs, d'usages terminologiques différents d'une discipline à l'autre. Il peut s'agit de locutions comme : *dans le/au sens de, si l'on considère telle discipline, du point de vue de.*

[7] Il n'y a pas eu de lithosphère archéenne (*dans le sens rhéologique du terme*).

[8] Du *point de vue chimique*, une oxydation est définie comme la perte d'un ou plusieurs électrons par un atome ou une molécule.

Les marqueurs ascendants.
L'analyse de la distribution des termes est particulièrement fructueuse dans une situation pluridisciplinaire. En effet, loin d'être problématiques lors de l'apparition d'une néodiscipline, des différences de distribution peuvent au contraire rendre compte de conceptualisations au moins en partie différentes qui se mettent en place à l'insu des experts. Ces différences sont ainsi plutôt le signe d'un foisonnement sémantique, pas toujours conscient, dans lequel le linguiste/terminologue peut contribuer à mettre de l'ordre.

La première étape de cette étude a consisté à repérer, à l'aide du logiciel wordsmith, les mots communs à au moins deux corpus. Voici les premiers mots communs à tous les corpus, avec leur nombre d'occurrences, respectivement pour l'astronomie, la biologie, la chimie et la géologie : *atmosphère (479, 30, 144, 476), eau (340, 266, 101, 318), température(228, 175, 91, 341), planète (467, 40, 30, 250), acide (40, 120, 515, 14), vie (178, 174, 139, 169), formation (180, 51, 195, 211), molécule(138, 139, 241, 75)…*

Pour chacun de ces mots, des types de fonctionnement ont été établis. Par exemple, pour *fractionnement*, on constate, dans le tableau 2 des différences de fonctionnement selon les corpus. Tout d'abord, on note que la chimie n'utilise jamais ce terme, ni au singulier, ni au pluriel. La biologie et l'astronomie l'utilisent mais seulement au singulier. La biologie l'utilise avec le seul sens de processus et l'astronomie avec le seul sens de résultat du processus. Seule la géologie l'utilise pour renvoyer aux deux sens (classiques pour une nominalisation déverbale). Ces deux types de sens sont nettement repérables par des éléments de contexte : *lors de, au moment* de orientent vers une interprétation de processus alors que *mesuré* après *fractionnement* ou *présente* avant ce même terme orientent plutôt vers un sens de résultat tangible.

Fractionnement	ASTRONOMIE	BIOLOGIE	CHIMIE	GEOLOGIE
Nombre de mots dans le corpus	77010	65 589	80190	88 815
Nombre d'occurrences du terme	6/0 (0,08 %)	15/0 (0,022 %)	0	32/13 (0,05 %)
Sens d'action		*Le fractionnement au moment du changement de phase*		*Fractionnements lors de l'accrétion terrestre*
Sens de résultat de l'action .	*fractionnement isotopique du carbone, mesuré dans ces sédiments*			*Le xénon atmosphérique présente un fractionnement*

Tableau 2 : Fonctionnement de *fractionnement* dans deux disciplines intervenant dans l'exobiologie

Résultats obtenus.
Dans cette étude des termes et de leur environnement en situation d'innovation disciplinaire, douze types de fonctionnements ont pu être établis. Ils s'organisent autour de trois phénomènes sémantiques, qui se mettent en place entre les corpus comparés (c'est-à-dire entre les disciplines) :

- Polysémie : un même terme a plusieurs sens d'un corpus à l'autre (voire, dans le même corpus), ou en tout cas un sens plus étendu : c'est le cas de *fractionnement*

- Emprunt : un terme est emprunté à une autre discipline, avec un sens similaire ou en partie différent. Ces cas sont assez difficiles à identifier. En effet, dans l'idéal, il faudrait disposer d'un corpus diachronique afin de repérer la discipline qui, la première, a fait l'usage du terme. Mais d'autres méthodes sont possibles dans un corpus synchronique. Tout d'abord, les locuteurs peuvent avoir conscience de cet emprunt ou bien on peut voir un usage significativement plus faible du terme emprunté dans la discipline emprunteuse. On peut aussi noter un environnement lexical bien moins riche dans le corpus emprunteur.
- Synonymie (plusieurs termes ont le même sens). Dans ces cas, on constate un environnement lexico-syntaxique similaire pour des termes différents (cf. ci-dessus *canaux galactophores* et *canaux lactifères* en [3] et [4]).

Tous ces fonctionnements peuvent être conscients, ce qui se repère généralement par des marqueurs lexico-syntaxiques, ou inconscients, ils sont alors repérés par l'interprétation des contextes d'apparition. Ils peuvent aussi être soit conflictuels : ils manifestent des points de vue incompatibles entre disciplines, soit non-conflictuels : ils cohabitent dans les différents corpus sans que cela semble poser de problèmes. Cette dimension conflictuelle vs non conflictuelle peut, elle aussi, être consciente (indiquée par des éléments lexico-syntaxiques ou non-consciente (mise au jour par analyse des contextes).

	CONSCIENT	NON-CONSCIENT	CONFLICTUEL	NON-CONFLICTUEL
POLYSEMIE	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

Tableau 3 : Différents types de fonctionnement lexical en situation pluridisciplinaire.

Les exemples [9] à [11] ci-dessous montrent comment certaines caractéristiques (correspondant aux lignes grisées dans le tableau 3) se mettent en œuvre dans le discours.

Exemple de synonymie non-consciente non conflictuelle

[9] *La non-détection de O_2 ou de O_3 ne pourra en aucun cas permettre de conclure à l'absence de vie sur une exoplanète*

[10] *La détection de la vie sur les planètes extrasolaires ne peut se faire qu'en analysant les signaux que nous recevons*

Exemple de polysémie consciente, conflictuelle

[11] Il y a peu de concepts [entropie] qui soient si souvent mal utilisés, tout particulièrement en biologie

Ce qui est notable dans les deux études que nous avons présentées, c'est que la nature des interprétations et les catégories de fonctionnements obtenues (17 pour l'évolution dans le temps et 12 pour la pluridisciplinarité) sont très liées à la nature même de l'étude. Toutefois, ce qui est constant d'une étude à l'autre, c'est la nature des contextes étudiés. Ils sont soit interprétables a priori (en langue), soit interprétables en contexte. Lorsqu'ils sont interprétables a priori, ils relèvent d'un

fonctionnement méta-linguistique que l'on peut associer à une dimension de conscience, plus ou moins claire, du locuteu. Lorsqu'ils demandent une analyse et une interprétation plus fine, ces contextes semblent relever d'une mise en discours plus inconsciente (plus épilinguistique) de la part des locuteurs (Condamines à paraître).

Les besoins d'analyse linguistiques dans les domaines spécialisés sont sans doute assez bien circonscrits. Il devrait donc être possible, pour chaque catégorie de besoins (évolution dans le temps, pluridisciplinarité, dé-terminologisation (Condamines et Picton 2014)) de dresser une liste de marqueurs métalinguistiques pertinents. Nul doute que ce matériau linguistique rendrait service à la communauté des linguistes-terminologues.

4 Conclusion

Si la terminologie textuelle s'est d'emblée focalisée sur la constitution de terminologies à partir de corpus, les méthodes de linguistique outillée, qui ont fait leur preuve pour ce type d'objectif, peuvent désormais être mises en œuvre pour d'autres types d'objectifs, en lien avec des besoins assez bien identifiés. Qu'il s'agisse de l'analyse de l'évolution de la connaissance dans le temps, de la prise en compte de la pluridisciplinarité, du changement géographique, du passage d'un domaine spécialisé vers la langue générale (Condamines et Picton 2014)…, la méthode d'analyse est similaire. La première étape, cruciale, concerne la constitution du corpus, qui sera l'objet d'étude tout au long de l'analyse. Son mode de constitution est directement lié à l'objectif de l'étude. Dans tous les cas que nous avons rencontrés, le corpus est organisé en sous-corpus (toujours en fonction de l'objectif d'étude) qui sont comparés. Il faut ensuite se donner un ensemble de termes-candidats communs à ces sous-corpus (ou à certains de ces sous-corpus) ; cette première liste de candidats-termes peut être fournie par les experts ou obtenue à partir d'extracteurs de termes-candidats. Il peut s'agir aussi de termes issus d'une base de connaissance terminologique (qui sert alors de point de départ à l'étude et plus de point d'arrivée). Le fonctionnement de ces candidats-termes est alors analysé grâce à deux types de méthodes. La première, descendante, permet de rechercher des structures métalinguistiques qui, en langue, sont directement en lien avec l'objectif de l'étude. La seconde, ascendante, vise à prendre en compte l'ensemble des contextes d'apparition afin de les interpréter, toujours en lien avec l'objectif de l'étude. La dernière étape vise à valider avec les experts du domaine les résultats obtenus afin de co-construire une interprétation.

Avec des objectifs variés mais bien identifiés et des méthodes d'analyse outillée systématisées, la terminologie textuelle devient un moyen fiable d'analyser les corpus spécialisés. Elle se rapproche ainsi de la linguistique de corpus mais avec une adaptation assumée à la nature et aux besoins de ce type de corpus.

Ahmad K., 1993 : « Terminology and Knowledge Acquisition : A Text Based Approach », dans K.D Schmitz, éd., *Proceedings of TKE'93 : Terminology and Knowledge Engineering*, Frankfurt, Indeks verlag., pp.56-70.

Aussenac-Gilles N., Condamines A., 2001 : « Entre textes et ontologies formelles : les Bases de Connaissances Terminologiques », dans M.Zacklad et M.Grundstein, éds, *Ingénierie et capitalisation des connaissance,* Paris, Hermes, pp. 153-176.

Aussenac-Gilles, N. Condamines, A. 2009 : « Variation syntaxique et contextuelle dans la mise au point de patrons de relations sémantiques », dans J.-L. Minel, éd., *Filtrage sémantique*, Paris, Hermes/Lavoisier, pp. 115-149.

Bourigault, D., Jacquemin, C. & L'Homme, M.-C., 2001 : *Recent Advances in Computational Terminology*, Amsterdam/Philadephia, John Benjamins Publishing Company.

Cabré, M.-T., 1999 : *Terminology. Theory, Methods and Application*, Amsterdam/Philadelphia, John Benjamins Publishing Company.

Condamines, A., 2003 : *Sémantique et corpus spécialisés : constitution de bases de connaissances terminologiques*, Mémoire d'habilitation à diriger les recherches, Université Toulouse 2 Le Mirail.

http://w3.erss.univ-tlse2.fr/textes/pagespersos/acondami/HDR-2.pdf

Condamines, A., à paraître : « Identification de relations conceptuelles en corpus spécialisés : mise en saillance définitoire métalinguistique vs épilinguistique », dans M. Bilger et H. Tyne, éds., *Langue française mise en relief(s)*. Presses Universitaires de Perpignan.

Condamines, A., et Amsili, P., 1993 : « Terminology between Language and Knowledge: an example of Terminological Knowledge Base », dans K.-D. Schmitz, éd., *TKE 93 Terminology and Knowledge Engineering*, Frankfurt, Indeks Verlag, pp. 316-323.

Condamines, A., et Dehaut, N., 2012 : « Rôle du temps et de la pluridisciplinarité dans la néologie sémantique en contexte scientifique: étude outillée en corpus », dans Gérard, C. et Katabek, J., éds., *Néologie sémantique et analyse de corpus*, dans Cahiers de lexicologie, 100, pp. 161-184.

Condamines, A., Dehaut, N., et Picton, A. : 2011 : « Mise en œuvre des méthodes de la linguistique de corpus pour étudier les termes en situation d'innovation disciplinaire : le cas de l'exobiologie », dans *META*, n° 56(2), pp. 266-283.

Condamines, A., Picton, A., 2014 : « Étude du fonctionnement des nominalisations déverbales dans un contexte de déspécialisation », dans *Actes du Congrès Mondial de linguistique Française* (CMLF). Berlin, 19-23 juillet 2014.
http://www.linguistiquefrancaise.org/articles/shsconf/pdf/2014/05/shsconf_cmlf14_01238.pdf

Faber, P., León Araúz, P. & Prieto Velasco, J.A., 2009 : «Semantic relations, dynamicity, and terminological knowledge bases», dans *Current Issues in Language Studies*, 1(1), pp.1-23.

Firth, J.R., 1957 : *Papers in Linguistics 1934–1951* (1957), London, Oxford University Press.

Gaudin, F., 2003 : *Socioterminologie : Une approche sociolinguistique de la terminologie*, Bruxelles, Duculot.

Harris, Zellig., 1954 : Distributional structure, dans *Word* 10(23), pp.146–162.

T. Kuhn, 1962 : *The Structure of Scientific Revolutions.* University of Chicago Press.

Laignelet, M., 2009 : *Associer analyse syntaxique et analyse discursive pour le repérage automatique d'informations potentiellement obsolescentes dans des documents encyclopédiques*, Thèse de l'Université Toulouse 2 Le Mirail.

Lerat, P., 1995 *: Les langues spécialisées*, Paris, PUF.

L'homme, M.-Cl., 2004 : *La terminologie : principes et techniques*, Presses Universitaires de Montréal.

Meyer I., Douglas S., Bowker L., et Eck K., 1992 : «Towards a new generation of terminological resources: An experiment in building a terminological knowledge base», *Proceedings 16th International Conference on Computational Linguistics*, Nantes, pp.956-957.

Meyer, I., 2001: « Extracting knowledge-rich contexts for terminography: A conceptual and methodological framework», dans Bourigault, D., Jacquemin, C. et L'Homme, M.-C, éds., *Recent Advances in Computational Terminology*, Amsterdam/Philadelphia, John Benjamins, pp.279-302.

Pearson, J., 1998 : *Terms in Context,* Amsterdam/Philadelphia, John Benjamins.

Picton, A., 2009 : *Diachronie en langue de spécialité. Définition d'une méthode linguistique outillée pour repérer l'évolution des connaissances en corpus. Un exemple appliqué au domaine spatial*. Thèse de l'Université Toulouse 2.

Rastier F., 1987 : *Sémantique interprétative*. Paris : PUF.

Rebeyrolle, J., 2000 : *Forme et fonction de la définition en discours*. Thèse de Sciences du langage, Université Toulouse Le Mirail.

Skuce D., Meyer I.,1991 : « Terminology and Knowledge Engineering: Exploring a Symbiotic Relationship », dans *Proceedings 6th International Workshop on Knowledge Acquisition for Knowledge-Based Systems* (Banff).

Slodzian M., 2000 : « L'émergence d'une terminologie textuelle et le retour du sens », dans H.Béjoint et P.Thoiron, éds, *Le sens en terminologie*, Lyon , PUL. pp. 61-85.

Temmerman, R. 2000 : *Towards New Ways of Terminology Description. The Sociocognitive Approach*, Amsterdam/Philadelphia: John Benjamins Publishing Company.

Wüster, E., 1981 : « L'étude scientifique générale de la terminologie, zone frontalière entre la linguistique, la logique, l'ontologie, l'informatique et les sciences des choses », dans G. Rondeau et H. Felber, éds., *Textes choisis de terminologie*, Québec : GIRSTERM, Université de Laval, pp. 55-108.