

Comment accéder aux éléments définitoires dans les textes spécialisés?

Dans cet article, je montre l'intérêt d'utiliser des corpus de textes spécialisés dans un travail terminographique (identification de termes et constructions de définitions terminologiques). J'essaie d'identifier les types de textes spécialisés qui conviennent à ce travail et je propose un certain nombre de critères pour sélectionner des textes appropriés. Je pars du principe que dans beaucoup de cas de communications spécialisées, les auteurs vont, de façon explicite ou implicite, expliquer certains des termes qu'ils utilisent. En utilisant trois corpus différents, je proposerai une méthodologie pour accéder à ces éléments définitoires.

Termes-clés:
corpus ; extraction ; définition ;
communication spécialisée.

1 Introduction

La terminographie exige des investissements importants en temps, en argent et en énergie. La disponibilité de textes électroniques devrait nous permettre de construire des corpus de textes importants qui se prêteraient à une exploitation terminographique semi-automatique. J'ai l'intention de parler de la conception et de la construction de corpus qui sont adaptés à ces activités et de proposer une méthodologie pour extraire des informations terminologiques.

Le travail sur corpus n'est pas une activité récente. Depuis déjà une trentaine d'années, des chercheurs construisent des corpus de plus en plus grands et les exploitent dans des buts différents, pour construire des dictionnaires, pour étudier la grammaire d'une langue, pour observer l'évolution d'une langue. Ce qui caractérise les grands corpus les plus connus (en langue anglaise tout au moins), c'est qu'ils contiennent des textes très variés. Ces corpus ont été construits pour être représentatifs de la langue telle qu'elle est utilisée tous les jours, l'intention étant de créer une ressource qui reflète les éléments communs d'une langue. Si ces corpus contiennent des textes spécialisés, ceux-ci ne représentent qu'une toute petite partie de l'ensemble du corpus. Par conséquent, ces corpus sont moins intéressants pour ceux qui cherchent à étudier la langue de domaines spécialisés. Le travail sur des corpus

spécialisés est beaucoup plus récent et pendant très longtemps, les quelques chercheurs qui voulaient exploiter des corpus de textes spécialisés ont dû créer leur propre corpus. Pour différentes raisons, ces corpus ont été difficiles à construire : il y avait un manque de textes électroniques, les permissions étaient difficiles à obtenir et, de la part des chercheurs, il y avait un manque d'enthousiasme pour partager ou prêter des ressources existantes. Par conséquent, pendant très longtemps, le nombre et la taille des corpus de textes spécialisés sont restés relativement faibles. La situation a évolué de façon notable ces dernières années. Les éditeurs sont plus prêts à donner leur permission, le volume de textes disponibles sous forme électronique ne cesse de croître. De nos jours, celui qui veut construire un corpus de textes spécialisés a l'embarras du choix, ce qui amène d'autres problèmes sur lesquels je reviendrai plus tard.

J'ai l'intention de me poser plusieurs questions qui me semblent importantes pour ceux qui s'intéressent à un travail sur corpus de textes spécialisés. Je me demanderai d'abord ce qu'est un corpus pour établir ce qui distingue un corpus d'autres collections de textes, telles que des archives par exemple. Je vais aussi essayer de caractériser ce qui distingue une approche à base de corpus d'une approche plus traditionnelle. Ensuite, j'essaierai d'établir ce que nous voulons dire quand nous utilisons le terme *texte spécialisé*. Ce terme est utilisé par des communautés différentes pour décrire un grand nombre de types de textes qui ne se prêtent pas tous à un travail

terminographique. Je tâcherai donc, dans la deuxième partie de mon article, et de définir le terme *texte spécialisé*, et de caractériser le genre de texte spécialisé qui conviendrait à une exploitation terminographique précise, c'est-à-dire l'extraction d'éléments définitoires qui serviront de base pour la formulation de définitions terminologiques. Quand nous aurons caractérisé le genre de texte qui nous intéresse, il nous restera à décider quels critères il faut appliquer pour sélectionner les textes les plus appropriés à notre activité. Dans la troisième partie, je proposerai donc un ensemble de critères que j'utilise dans mon propre travail terminographique.

La plus grande partie de mon article consistera à décrire la méthodologie que j'ai adoptée pour réaliser mon travail. J'ai conçu deux approches différentes, l'une qui consiste à exploiter la présence de certaines structures grammaticales et l'autre qui consiste à utiliser les termes eux-mêmes comme point de départ. Dans cet article, l'accent sera sur la première de ces deux approches. Il est important de noter que ce que je propose n'est pas une solution parfaite; elle ne mène pas à une solution automatique. L'expert aura seulement les éléments avec lesquels il pourra travailler ensuite.

2 Qu'est-ce qu'un corpus?

D'après Sinclair (1994: 2), un corpus est une « *collection of pieces of language that are selected and ordered according to explicit linguistic criteria to be used as a sample of the language* ». D'après Atkins, Clear et Ostler (1992:1) un corpus est « *a subset of an ETL (electronic text library) built according to explicit design criteria for a specific purpose, e.g. the Corpus Révolutionnaire (Bibliothèque Beaubourg, Paris), the Cobuild Corpus, the Longman/Lancaster corpus, the*

Oxford Pilot Corpus ». Puisque ces définitions sont très proches des définitions proposées par d'autres linguistes, nous pouvons les considérer comme étant représentatives. Ces définitions sont révélatrices et nous permettent de voir en quoi un corpus est différent d'autres collections de textes. Tout d'abord, les textes que l'on trouve dans un corpus ne s'y trouvent pas par hasard. Il ne s'agit pas de prendre n'importe quel texte tout simplement parce qu'il est disponible. Chaque texte est sélectionné selon des critères linguistiques précis. Deuxièmement, si l'on crée un corpus, on le fait dans un but précis; on sait pourquoi on le crée, on sait d'avance à quoi il va servir. Troisièmement, les textes que l'on trouve dans un corpus sont des textes authentiques et naturels, c'est-à-dire qu'ils n'ont pas été édités. Bien qu'un corpus contienne des textes qui ont été rédigés pour être lus, il ne sera jamais utilisé de cette façon-là. S'il nous arrive d'en lire des bouts, ces bouts viendront sans doute de textes différents et même quand ils viennent du même texte, ils ne se seront pas suivis dans le texte d'origine. On utilise des outils informatiques (tels que *WordSmith Tools*) pour accéder au corpus; ces outils nous permettent de voir des mots dans leur contexte même si, d'une certaine façon, on les voit aussi hors de leur contexte. Le fait de lire plusieurs occurrences à la fois nous permet d'identifier des tendances que l'on n'aurait peut-être pas observées avec une approche plus conventionnelle. En résumé, un corpus est un produit artificiel qui contient des textes authentiques qui ont été sélectionnés selon des critères précis pour être utilisés comme un échantillon de la langue. En ce qui concerne notre travail, nos corpus contiendront des textes spécialisés, ce qui nous mène à la question suivante, à savoir, qu'est-ce qui caractérise les textes spécialisés?

3 Caractérisation du texte spécialisé

Quand nous utilisons le terme *texte spécialisé*, nous le faisons habituellement pour différencier les textes dits *de langue générale* et les textes dits *de langue de spécialité*. On peut utiliser des critères différents pour caractériser les textes spécialisés et les critères que l'on utilisera dépendront en grande partie de la raison pour laquelle on étudie ces textes. Certains chercheurs s'intéressent aux aspects lexicaux, d'autres aux aspects grammaticaux et d'autres encore aux aspects de typologie du texte. Ceux qui s'intéressent aux aspects lexicaux, tels que les terminologues, diront qu'un texte spécialisé se caractérise par son vocabulaire spécialisé, c'est-à-dire par la fréquence de termes techniques utilisés dans le texte. Ceux qui s'intéressent aux aspects grammaticaux, tels les chercheurs qui se consacrent aux sous-langages, diront que les textes spécialisés se caractérisent non seulement par le vocabulaire utilisé mais aussi et peut-être plus par leurs structures grammaticales. C'est-à-dire que souvent, la grammaire de textes spécialisés est différente de la grammaire pour la langue générale, différente parce que limitée en ce qui concerne les structures utilisées, différente parce que les structures grammaticales qu'on y trouve n'obéissent pas aux règles de grammaire normales (utilisation d'ellipses, construction de termes composés longs et compliqués). Ceux qui s'intéressent à la typologie des textes s'occupent principalement de la fonction et de l'organisation des textes. Si, par exemple, ils étudient la langue de la médecine, ils vont plutôt essayer de cerner les différences entre les différents *types* de documents, que ce soit des articles de recherche, des notices d'emploi pour des

médicaments, des rapports de laboratoire.

Quand on veut construire un corpus de textes spécialisés qui va servir à un travail terminographique, il se trouve qu'aucune des façons de caractériser les textes spécialisés citées ci-dessus ne convient à nos besoins. Contrairement à ce que l'on pourrait penser, tout texte spécialisé ne convient pas à une exploitation terminographique. Rappelons que notre but principal est de sélectionner des textes qui contiennent des éléments définitoires, c'est-à-dire des textes où certains des termes utilisés sont définis ou de façon implicite, ou de façon explicite. Si nous utilisons uniquement le critère de la fréquence des termes pour sélectionner nos textes, nous nous trouverons avec beaucoup de textes qui sont peu riches en éléments définitoires. Il en est le même du critère des structures grammaticales inhabituelles. Il semblerait que le critère de typologie puisse convenir le mieux, mais ce critère pose aussi des problèmes. Il me semble que toute discussion de ce que c'est qu'un texte spécialisé doit tenir compte d'un élément très important qui est extra-textuel, c'est-à-dire le rapport entre l'auteur et le lecteur.

Cet aspect est primordial quand nous voulons décider si un texte doit faire partie de notre corpus. Nous cherchons à construire un corpus qui nous permette non seulement de repérer des termes mais aussi de repérer des éléments définitoires. Si nous pouvons nous attendre à trouver un nombre considérable de termes techniques dans tout texte spécialisé, il n'en est pas le même des éléments définitoires. Alors qu'il y a beaucoup de situations où des auteurs expliqueront de façon explicite certains des termes qu'ils utilisent, il y en a autant, sinon plus, où ils ne le font pas. Par exemple, on ne s'attend pas à en trouver dans des bulletins de météo, ni dans des recettes de cuisine, ni dans des notices d'emploi de

médicaments par exemple. Par contre, on s'attend à en trouver dans des livres de formation, dans des normes industrielles, et même quelquefois dans des articles de recherche s'il s'agit de créer de nouvelles notions ou de redéfinir des notions existantes.

Les rapports entre l'auteur et le lecteur détermineront combien d'explications seront fournies dans un texte donné et ceci est valable pour toute sorte de texte. Je crois qu'il y a trois types de rapports auteur-lecteur qui sont susceptibles de nous intéresser. Le premier concerne la communication entre experts. Les textes rédigés par des experts pour des experts vont avoir une très haute fréquence de termes techniques. Considérons par exemple des articles de recherche publiés dans des revues spécialisées. Dans ces textes, on trouvera une très haute densité de termes mais probablement très peu d'éléments définitoires. L'explication est simple: le lecteur est censé connaître et comprendre les termes utilisés.

Le deuxième type de rapports concerne la communication entre des experts et des gens qui ont déjà une certaine compétence dans le domaine en question. Je pense ici à des gens qui travaillent dans le même domaine mais qui n'ont pas le même niveau de formation. Cela pourrait être des communications entre ingénieurs et techniciens, entre médecins spécialisés et médecins généralistes. Ce qui distingue cette situation de la précédente est le niveau d'expertise. Les auteurs vont définir ou expliquer certains des termes qu'ils utilisent quand ils estiment que ces termes ne sont pas connus par leurs lecteurs.

Le troisième type de rapports concerne la communication entre les experts et les gens qui n'ont aucune formation dans le domaine en question mais qui ont besoin de connaître et de comprendre la terminologie du domaine. Dans ces communications, la densité de termes

va être nettement moins élevée que dans les deux catégories précédentes mais on peut s'attendre à trouver une forte densité d'éléments définitoires. Je pense par exemple à des livres d'école ou à des livres de formation professionnelle. Ce qui est intéressant ici est que ce genre de texte est rarement considéré comme étant intéressant dans les discussions de textes spécialisés, tout simplement parce que la fréquence de termes n'est pas suffisamment élevée. Comme nous le verrons plus tard, ce sont précisément ces textes-là qui sont les plus riches en éléments définitoires et les communications entre experts les plus pauvres. Ce qui est important à retenir dans le cas de chacun de ces rapports auteurs-lecteurs est que le cadre de communication dans lequel on trouvera des textes qui correspondent à ces rapports est un cadre professionnel ou éducatif.

Il y a un quatrième type de rapports auteur-lecteur que l'on aurait pu envisager mais qui s'est avéré ne pas convenir à nos besoins. C'est le cas des rapports entre des experts et des lecteurs occasionnels; là, il s'agit de communications spécialisées qui paraissent dans la littérature dite *d'intérêt populaire*, dans des revues comme *Scientific American* aux États-Unis ou *New Scientist* en Angleterre, par exemple. Le cadre n'est ni professionnel ni éducatif dans le sens où les lecteurs qui lisent ces textes ne le font pas dans le cadre de leur profession ou de leur formation.

4 Critères de sélection

Maintenant que nous avons décidé que nous voulons sélectionner des textes qui correspondent à un des cadres de communication énoncés précédemment, il nous reste à établir un jeu de critères pour sélectionner les textes que nous voulons inclure dans nos corpus. Il ne suffit pas de

dire que tel ou tel texte correspond à un de nos cadres de communications. Il y a d'autres critères qui entrent en jeu. Les critères les plus importants sont les suivants :

1. Les textes doivent avoir été publiés. Cela veut dire, d'après la définition de Biber, que « they are printed in multiple copies for distribution, they are copyright registered or recorded by a major indexing service » (1993: 245). Nous envisageons donc d'analyser des textes qui ont été rédigés pour être lus.

2. Les textes doivent être des textes entiers. Traditionnellement, les corpus contiennent des échantillons d'une longueur précise de textes différents.

3. L'auteur peut être un individu, un groupe d'individus, une association ou une fédération. Dans tous les cas, l'auteur doit être compétent pour écrire dans le domaine en question. Sa formation professionnelle doit correspondre au sujet dont il traite.

4. Le cadre doit être professionnel ou éducatif. Les textes sont destinés à être utilisés dans un cadre professionnel ou dans un cadre d'enseignement.

5. La date de publication d'un texte peut être importante, surtout dans des domaines en pleine évolution, tels que l'informatique.

6. La fonction du texte doit être informative, didactique ou normative.

5 Les corpus

Pour élaborer ma méthodologie, j'ai choisi d'utiliser trois corpus différents. Le premier, le corpus *Nature*, est une collection d'articles publiés dans la revue *Nature*. *Nature* est une revue spécialisée anglaise qui contient des articles rédigés par des experts pour des experts. Le cadre de communication correspond donc à

notre première catégorie, la communication entre experts. J'ai choisi d'étudier ce corpus parce que mon intuition me suggérait que bien que ces textes sont on ne peut plus spécialisés, il y avait de fortes chances pour qu'il y ait très peu d'éléments définitoires, ce qui s'est avéré être le cas. Le corpus *Nature* contient environ 230 000 mots.

Le deuxième corpus, le corpus UIT, rédigé en anglais, français et espagnol, a été publié par l'Union internationale des télécommunications et est disponible sur CD-Rom. Le cadre de communication correspond à notre deuxième catégorie ; il s'agit de communication entre des experts et des gens qui travaillent dans le même domaine mais qui n'ont pas nécessairement le même niveau de compétence que les auteurs. Le corpus UIT comprend des textes normatifs et contient environ 4,7 millions de mots.

Le troisième corpus contient des livres d'enseignement pour des matières sur le programme d'études GCSE en Angleterre. Les élèves passent l'examen GCSE deux ans avant de passer l'équivalent du baccalauréat français. Le cadre de communication correspond à notre troisième catégorie, la communication entre des experts et des gens qui n'ont aucune ou très peu de connaissances dans le domaine en question, mais qui ont besoin de connaître et comprendre les notions du domaine. Le corpus comprend des textes didactiques et contient environ un million de mots.

6 Des éléments définitoires

Maintenant que nous avons défini ce que nous voulons dire par *texte spécialisé* et, plus particulièrement, ce que nous entendons par un *corpus spécialisé* à

but terminographique, je vais diriger mon attention vers la notion d'éléments définitoires. Je pars du principe que dans les cadres de communication mentionnés précédemment, les textes vont contenir des éléments définitoires. Les auteurs vont nous fournir des explications de certains des termes qu'ils utilisent.

Je crois qu'il y a plusieurs moyens d'accéder à ces éléments définitoires : par l'intermédiaire du terme, par les structures grammaticales, etc. Je propose d'abord d'élaborer une description des éléments définitoires du point de vue de leur structure grammaticale. Je vais partir de structures très formelles pour arriver ensuite à des structures moins formelles et donc plus difficiles à identifier. Je propose d'utiliser les descriptions élaborées par Trimble en 1985 et développées ensuite par Flowerdew en 1992, deux chercheurs qui ont analysé des textes réels. D'après ces chercheurs, il y a trois catégories de définition : la définition formelle, la définition semi-formelle et la définition non formelle. D'après Trimble, la définition formelle est « *of course, the well-known equation-like "Species = Genus + Differentia", usually called "formal" because of its rigidity of form* » (Trimble 1985: 75-76). La définition formelle a une structure rigide qui est généralement représentée par la formule suivante : $X = Y + caractéristique$. Dans cette formule, x représente le terme, y représente un hyperonyme et la caractéristique sert à distinguer x de tous les autres termes de la même catégorie. Voici quelques exemples que l'on trouve dans nos corpus :

- *A control circuit is a telephone-type circuit between the point of origin of the programme and the point where it terminates.* (UIT)
- *An ISDN connection is a connection established between ISDN reference points* (UIT)

- *A spore is a single cell which by itself can grow into a new organism* (GCSE)

En ce qui concerne la définition semi-formelle, Trimble dit « *By definition, a semi-formal definition contains only two of the three basic defining elements: the term being defined and the statement of differences* » (Trimble 1985: 77). La définition semi-formelle est généralement représentée par la formule: $x = \text{caractéristique}$, comme dans les exemples suivants:

- *Geometric elements are used to construct drawings of various types by a succession of overlay of points, straight lines, arcs etc.* (UIT)
- *Photographic elements are used to render an image by the transmission and display of an array of picture elements (pixels)* (UIT)
- *Arachnids have four legs and their bodies are made of two parts* (GCSE)
- *Amylase breaks up starch into sugar* (GCSE).

La troisième définition est une définition non formelle, décrite par Trimble comme suit « *The function of a non-formal definition is to define in a general sense so that a reader can see the familiar element in whatever the new term may be... Most non-formal definitions are found in the form of synonyms* » (Trimble 1985: 78). Elle fournit au lecteur le nom correct du terme et une autre unité lexicale qui a plus ou moins le même sens que le terme mais qui n'est pas obligatoirement un terme. Flowerdew (1992: 211) appelle cette définition *une définition par substitution* et suggère que la définition par substitution peut être exprimée de trois façons différentes: par un synonyme, par paraphrase ou par dérivation. Voici quelques exemples que l'on trouve dans nos corpus:

- *a signal of limited duration, known as a «measuring signal»* (UIT)
- *the aggregate of time during which the speech in question is present (called the active time)* (UIT)

- *A bulge called a pseudopodium...* (GCSE)
- *Peas have small shoots called tendrils* (GCSE)
- *...an extended region of homology called the POU domain* (Nature).

Ce qui est intéressant ici est que l'on trouve, outre des paraphrases, non seulement des rapports de quasi-synonymie mais aussi des rapports hyperonyme-hyponyme.

6.1 Les définitions formelles

6.1.1 Les définitions formelles simples

En étudiant les structures grammaticales utilisées dans les corpus, j'ai pu constater qu'il y en a certaines qui correspondent généralement aux différents types de définitions énoncés précédemment. J'ai donc conçu un ensemble de conditions qu'une structure grammaticale doit remplir pour être considérée comme un élément définitoire. Regardons d'abord les conditions que doivent remplir les définitions formelles simples, c'est-à-dire les définitions qui correspondent à la structure $x = y + \text{caractéristique}$ et qui apparaissent à l'intérieur d'une seule phrase.

1. X doit être un terme. Pour être considéré comme terme, il doit avoir une structure qui correspond aux structures terminologiques identifiées pour le corpus en question. Si x apparaît au début de la phrase, il doit être précédé ou de l'article indéfini ou ne pas être précédé d'article du tout. Ceci est très important parce que, en anglais, si le terme est précédé de l'article défini (quand il est en début de phrase), il n'a pas nécessairement un statut générique. Si le terme est précédé de l'article défini ou d'un adjectif démonstratif, cela peut vouloir dire deux choses: ou que la phrase est une continuation d'une définition déjà élaborée en partie dans la phrase

précédente, ce qui est le cas pour les définitions complexes qui ne nous concernent pas ici, ou que la phrase concerne un cas précis qui ne peut pas être considéré comme étant valable pour le terme en général. Par contre, si x apparaît à la fin de la phrase, il peut être précédé soit de l'article défini, soit de l'article indéfini ou il peut apparaître sans aucun article.

2. Y doit être un terme ou doit être un hyperonyme générique tel que *technic, method, process, function*, etc. Le tableau 1 indique les hyperonymes génériques que l'on trouve dans les trois corpus.

Hyperonyme générique	UIT	GCSE	Nature
<i>technic</i>	341	7	9
<i>method</i>	2,374	149	32
<i>process</i>	1,143	201	63
<i>function</i>	3,105	54	133
<i>property</i>	39	71	9
<i>system</i>	7,396	712	135
<i>class</i>	1,598	416	81
<i>device</i>	746	9	4

Tableau 1:
liste d'hyperonymes génériques

3. Il y a un certain nombre de verbes ou de syntagmes verbaux qui peuvent lier x et y . On les appelle *hinges* ou *connective verbs*. Ils servent de lien entre les deux éléments de la phrase. Quelques exemples de liens: *is/are, is/are called is/are known as, is/are defined as, denote(s)*. Le verbe ou la phrase verbale qui sert de lien entre x et y doit être au présent de l'indicatif et ne peut pas être modifié par un verbe modal. Le verbe ne doit être modifié par aucune particule négative, telle que *not, never*. La phrase définitoire ne doit pas être modifiée par un «adverbe de focus» (*focussing adverb* en anglais). Quelques exemples de ces adverbes: *commonly, frequently, usually, specifically, generally, mainly,*

primarily. Le tableau 2 indique les adverbes de focus que l'on trouve dans les trois corpus.

Adverbe	UIT	GCSE	Nature
<i>Chiefly</i>	5	9	1
<i>commonly</i>	91	19	4
<i>especially</i>	176	142	21
<i>exceptionally</i>	88	1	4
<i>exclusively</i>	61	3	6
<i>frequently</i>	93	59	10
<i>generally</i>	496	140	35
<i>mainly</i>	98	125	20
<i>mostly</i>	11	85	18
<i>occasionally</i>	18	13	1
<i>often</i>	213	574	37
<i>only</i>	4,504	1,457	385
<i>on the whole</i>	2	13	1
<i>predominantly</i>	12	5	6
<i>primarily</i>	118	4	19
<i>principally</i>	9	0	2
<i>purely</i>	61	10	3
<i>rarely</i>	22	31	1
<i>solely</i>	93	5	5
<i>sometimes</i>	111	252	11
<i>specifically</i>	175	2	27
<i>usually</i>	354	408	17

Tableau 2:

Adverbes de focus dans les trois corpus

Bien que l'on puisse penser que ces adverbes sont souvent utilisés pour accentuer la validité d'une phrase définitoire, ils ont souvent l'effet opposé. Quand un auteur dit qu'un terme est *souvent* défini de telle ou telle façon, on ne peut pas conclure que c'est *toujours* le cas.

4. Il ne suffit pas de préciser quels mots peuvent se trouver à la place de x , y et $=$. Il faut aussi regarder et la forme de la phrase, et la situation de l'élément définitoire dans la phrase en général. Tout d'abord, la phrase définitoire doit constituer la proposition principale de la phrase et ne doit pas être précédée de phrases subordonnées. Si la phrase définitoire est précédée d'une phrase subordonnée, celle-ci peut atténuer la force et la validité de la phrase définitoire. Par contre, la phrase

définitoire peut être suivie d'une autre phrase à condition que ces deux phrases soient liées par la conjonction *and*.

5. Y doit être suivi directement par la caractéristique qui distingue x de tous les autres membres de la même catégorie. La caractéristique peut être introduite par une proposition, un participe passé ou un pronom relatif. Y ne doit pas être immédiatement suivi de la conjonction *and* parce que cela signale généralement la fin de l'élément définitoire.

Voici quelques exemples de définitions formelles simples:

- *A robot is a machine that tries to copy one or more human functions (GCSE).*
- *A videotex service centre is a computer used by the videotex service provider to authorize access to a videotex service (ITU).*
- *INTERLEUKIN-1 (IL-1) is a cytokine produced by mononuclear phagocytes (Nature).*

6.1.2 Les définitions formelles complexes

Bien qu'il y ait un nombre considérable de phrases définitoires qui correspondent à une définition formelle simple, c'est-à-dire une définition formelle exprimée dans une seule phrase, il y en a beaucoup d'autres qui sont exprimées dans deux phrases. Ce sont ce que Trimble appelle *des définitions complexes*. En général, une définition formelle complexe est exprimée d'une des deux façons suivantes: a) le terme est introduit à la fin d'une phrase et expliqué au début de la phrase suivante, ou b) le terme est nommé au début d'une phrase et expliqué dans la phrase précédente. Voici quelques exemples, d'abord du premier cas, et ensuite du deuxième:

Premier cas (le terme apparaît à la fin de la phrase précédente):

- *...mean holding time. This is the total holding time divided by the total*

number of seizures and can be calculated on a circuit group basis or for switching equipment. (UIT)

- *...digital speech interpolation (DSI). This is a technique whereby advantage can be taken of the inactive periods during a conversation, creating extra channel capacity. (UIT)*

• *...a ring mains system. This is a loop of cable that runs from the consumer unit round the house and back to the unit. (GCSE)*

• *...a device called a proboscis. This is a hollow tube which is normally tightly coiled beneath the head of the butterfly when it is not feeding. (GCSE)*

• *an exoskeleton. This is a hard outer protective covering made of chitin.*

Deuxième cas (définition dans une première phrase, terme nommé dans la phrase suivante):

• *There are millions of compounds containing just hydrogen and carbon. These are called hydrocarbons.*

• *Some bacteria get their food from the dead bodies of plants and animals. These are called saprophytic bacteria.*

Ces définitions n'ont pas la même structure formelle que les définitions formelles simples mais tous les éléments des définitions formelles simples sont présents. Ces définitions sont relativement faciles à repérer. Dans le cas a) cité ci-dessus, il suffit de préciser que les mots que l'on trouve à la fin de la première phrase doivent correspondre à une structure terminologique déjà précisée et que la phrase définitoire doit commencer avec l'expression *This is...* ou *These are...* La phrase définitoire doit remplir les mêmes conditions précisées pour les définitions formelles simples à l'exception de x qui est remplacé par un pronom (*this* ou *these*). Dans le cas b) cité ci-dessus, il faut préciser que les mots qui suivent l'expression *These are* ou *This is* doivent être un terme. Les définitions complexes sont plus difficiles à identifier que les définitions simples, mais, dans le

corpus UIT en particulier, elles sont beaucoup plus fréquentes.

6.2 Les définitions semi-formelles

Plus fréquentes encore sont les définitions semi-formelles. Ce sont là des structures qui correspondent à la formule: $x = \text{caractéristique}$. Dans une définition semi-formelle, l'hyperonyme est absent de la phrase. En réalité, quand on analyse les définitions semi-formelles que l'on trouve dans le corpus, on découvre que l'hyperonyme est souvent précisé dans la phrase précédente et que la définition semi-formelle sert à fournir des informations supplémentaires. Les conditions pour les définitions semi-formelles sont un peu différentes des conditions précisées pour les conditions formelles. X doit être un terme qui peut être précédé de l'article défini ou de l'article indéfini. Le verbe qui lie x avec la caractéristique doit être au présent de l'indicatif; la phrase définitoire semi-formelle doit constituer la proposition principale de la phrase. Voici quelques exemples de nos corpus:

- *Expanded polystyrene is made by blowing a gas (such as carbon dioxide) into the liquid polymer. (GCSE)*
- *One of the most important applications of neutralization is in making fertilizers like ammonium sulphate ($((NH_4)_2SO_4)$) and ammonium nitrate (NH_4NO_3). Ammonium sulphate is manufactured by neutralizing sulphuric acid (H_2SO_4) with ammonia solution ($NH_3(aq)$). (GCSE)*
- *Whenever a change in the status of a signalling link, route or point occurs, the three different signalling network management functions (i.e., signalling traffic management, link management and route management) are activated. The signalling traffic management function is used to divert signalling traffic from a link or route to one or*

more different links or routes, to restart a signalling point, or to temporarily slow down signalling traffic in the case of congestion at a signalling point... (UIT).

6.3 Les définitions non formelles

Les définitions non formelles sont peut-être les plus intéressantes à étudier. Rappelons que d'après Flowerdew, ces définitions, qu'il appelle *définition par substitution*, peuvent être exprimées de trois façons différentes: par un synonyme, par paraphrase ou par dérivation. En cas de substitution, l'auteur peut fournir un synonyme pour un terme (ce cas est assez rare) ou il va introduire un terme en utilisant un mot équivalent de langue générale qui est déjà connu du lecteur. En cas de paraphrase, l'auteur peut expliquer un terme en utilisant une paraphrase. Dans ce cas, il fournira les mêmes éléments que l'on trouve dans les définitions formelles et semi-formelles mais emploiera d'autres méthodes pour le faire.

Il y a un certain nombre d'indicateurs qui nous signalent la présence de définitions non formelles. Ils comprennent entre autres l'utilisation de parenthèses, l'utilisation d'expressions telles que *i.e., e.g., called, known as*. En plus des fonctions élaborées ci-dessus, ces indicateurs peuvent aussi être utilisés pour indiquer des rapports hyperonyme-hyponyme. On trouvera ci-dessous quelques exemples de chacun des indicateurs qui révèlent combien les indicateurs sont multifonctionnels. On ne peut pas dire d'avance que tel indicateur va fournir telle information.

- *...but all microphones contain a disc called a diaphragm. (GCSE)*
- *...threads of pure cellulose known as rayon. (GCSE)*

- *Modulation with two-phase conditions, called bi-phase modulation (2-PSK)... (UIT)*
- *...a signal of limited duration known as a "measuring signal"... (UIT)*
- *Row crops (e.g. strawberries) can be protected from... (GCSE)*
- *cell types, e.g. root-hair cell, egg cell (ovum) sperm cell, muscle cell... (GCSE)*
- *some are steady state impairments (e.g. loss, noise, quantization, distortion, phase jitter...) (UIT)*
- *But how can liquids change to solids (i.e. melt) or liquids to gases (i.e. boil or evaporate)? (GCSE)*
- *an idle signalling terminal, i.e. a signalling terminal not connected to a signalling data link (UIT).*

7 Conclusion

La méthodologie proposée ici est une approche qui cherche à exploiter l'utilisation de certaines structures dans des circonstances précises. L'identification de ces structures nous permet de repérer des éléments définitoires qui peuvent être considérés dans la formulation de définitions terminologiques. On peut aussi aborder le problème d'une autre façon, en prenant les termes eux-mêmes comme point de départ. Dans ce cas, on produit des concordances pour chaque terme pour voir si on peut cerner des éléments définitoires dans les lignes de concordance. Cette approche est très proche de celle qui est utilisée par l'équipe lexicographique *Cobuild*. Bien que cette approche soit aussi très productive, elle demande beaucoup plus d'analyse manuelle. Je suis persuadée qu'en associant une approche grammaticale telle que celle décrite dans cet article avec une approche qui prend les termes comme point de départ pour l'analyse de textes spécialisés appropriés, nous

réussirons à réduire les coûts élevés de travail terminographique.

Jennifer Pearson,
Salis,
Dublin City University,
Irlande.

Références

Atkins (S.), Clear (J.) & Ostler (N.), 1992: «Corpus Design Criteria», dans *Literary and Linguistic Computing*, Oxford, Oxford University Press, vol. 7,1, p. 1-16.

Biber (D.), 1993: «Representativeness in corpus design», dans *Literary and Linguistic Computing*, Oxford, Oxford University Press, vol. 8 (4), p. 243-257.

Daille (B.) 1994. *Approche mixte pour l'extraction de terminologie: statistique lexicale et filtres linguistiques*, thèse de doctorat, Université Paris VII.

Flowerdew (J), 1992: «Definitions in Science Lectures», dans *Applied Linguistics*, vol.13 (2), p. 202-221.

Jacquemin (C.), Royaute (J.), 1994: «Retrieving Terms and their Variants in a Lexicalized Unification-Based Framework», dans *Proceedings of the Seventeenth Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval*, New York, Heidelberg: Springer-Verlag, p. 132-141.

Nkwenti-Azeh (Bl.), 1992: *Positional and Combinational Characteristics of Satellite Communications Terms*, Final Report, Eurotra Project, UK-CCL-UMIST.

Pearson (J.), 1998: *Terms in Context*, Amsterdam, John Benjamins Publishing Company

Sinclair (J.) 1994: «Corpus Typology: A Framework for Classification » EAGLES document. 1-18, disponible maintenant dans Sinclair (J.), 1995: «Corpus Typology: A Framework for Classification», dans Melchers (G.) & B. Warren (éd.) *Studies in Anglistics*. Stockholm, Almquist and Wiksell International, p. 17-34.

Trimble (L.), 1985. *English for Science and Technology: A Discourse Approach*, Cambridge, Cambridge University Press.