

La place des reformulations dans les textes scientifiques

Philippe Thoiron and Henri Béjoint

Volume 36, Number 1, mars 1991

La terminologie dans le monde : orientations et recherches

URI: <https://id.erudit.org/iderudit/004478ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

[Explore this journal](#)

Cite this article

Thoiron, P. & Béjoint, H. (1991). La place des reformulations dans les textes scientifiques. *Meta*, 36(1), 101–110.

LA PLACE DES REFORMULATIONS DANS LES TEXTES SCIENTIFIQUES

PHILIPPE THOIRON ET HENRI BÉJOINT
Université Lumière — Lyon 2, Lyon, France

INTRODUCTION

Dans une publication récente (Chukwu et Thoiron 1989), notre équipe s'intéressait aux reformulations — définies comme des reprises, aux fins d'explicitation, d'un concept C par une formulation Y différente de la formulation sous forme d'un terme X — et à leur utilisation dans les textes scientifiques. L'article montrait qu'on peut voir les reformulations² comme des symptômes de la présence de termes, et qu'on peut donc les utiliser comme des outils de repérage. C'est ainsi que les auteurs proposaient une typologie des reformulations fondée sur la présence de marques formelles telles que certains éléments lexicaux (*c'est-à-dire*, *appelé*, etc.) ou signes de ponctuation (parenthèses, tirets, etc.), relateurs explicites ou implicites entre le terme reformulé et la séquence syntagmatique qui le reformule.

Le présent article a pour objet de poursuivre l'examen des reformulations dans les textes scientifiques en prenant pour champ d'investigation le texte en tant qu'unité. Il ne s'agira pas ici d'examiner dans un corpus les divers types de reformulations pour eux-mêmes ou pour leur valeur symptomatique. On cherchera plutôt à évaluer la place des différentes occurrences de reformulations à l'intérieur de chaque texte. On cherchera également à étudier dans quelle mesure une typologie des textes scientifiques est envisageable à partir d'une typologie des reformulations et d'un examen de la place qu'elles occupent dans chaque texte.

Un tel travail ne peut être que programmatique dans l'état actuel de la recherche. On s'appuiera donc sur un nombre de textes limité, en anglais et en français, dans un seul secteur, celui de la recherche sur le SIDA³. Les observations que nous serons amenés à faire devront ensuite être confrontées à des corpus plus volumineux, à des textes plus variés, dans d'autres langues, et à d'autres domaines de la connaissance.

I. LA TOPOGRAPHIE DES REFORMULATIONS DANS L'ENSEMBLE DU TEXTE

«Dans le discours scientifique, qui est par essence rhétorique, la progression va du connu vers une découverte [...]» (Vidalenc 1989 : 128). La reformulation étant une aide à la présentation de l'inconnu, il est logique, dans un premier temps, de penser qu'elle interviendra surtout à la fin des textes. C'est effectivement ce qui se passe dans un certain nombre de cas. Les textes qui fonctionnent ainsi sont, au moins dans le domaine limité qui nous concerne, ceux qui commencent par une partie expérimentale — la description de l'expérience ou du protocole utilisant en général des éléments connus — ou par une partie descriptive à haut degré de généralité.

Néanmoins, certains des textes de notre corpus échappent à ce type d'organisation : ils utilisent des reformulations dès les premières lignes. C'est donc qu'une autre logique y est à l'œuvre. L'auteur se base sur ce qui est connu de la communauté scientifique pour «asseoir» sa démonstration en début de texte, mais utilise un certain nombre de reformulations pour s'assurer que le lecteur le suit. Ces reformulations peuvent être

nécessaires soit parce que le savoir du lecteur potentiel risque d'être en-deçà du savoir de la communauté, soit parce que l'auteur utilise des termes dont l'usage n'est pas accepté de la même manière par tous.

Si la deuxième partie du texte ne nécessite pas la mise en œuvre d'autres termes que ceux qui ont été introduits et reformulés au début, on obtient un schéma inversé : davantage de reformulations en début de texte qu'en fin de texte. Le texte n° 2, par exemple, comprend 828 lignes, et contient 34 reformulations, soit une moyenne de 3,4 reformulations pour chaque tranche de 83 lignes. Or, le décompte du nombre de reformulations par tranche donne les chiffres suivants : 7, 8, 3, 2, 3, 4, 2, 3, 2, 0. De la même façon, le texte n° 5, avec en moyenne 1,5 reformulation par tranche, présente pour ses 10 tranches de longueur égale la distribution suivante : 2, 5, 3, 0, 0, 2, 1, 2, 0, 0 ; les deux tiers des reformulations se trouvent dans le premier tiers du texte. Si l'on ajoute à cela le fait que la reformulation de la 7^e tranche est aberrante (l'utilité de gloser *HIV positive serums par samples known to have antibodies to HIV* est contestable compte tenu du niveau général de complexité conceptuelle du texte) et que la 8^e tranche contient un rappel de termes utilisés dans la 1^{re} tranche⁴, on aura une meilleure image du déséquilibre statistique du nombre de reformulations au profit du début du texte.

La structure discursive de ce texte est d'ailleurs bien corrélée avec la structuration quantitative décrite ci-dessus. L'auteur annonce lui-même ainsi le mouvement général de son exposé :

«[...] to learn more about viruses related to HIV and so understand how HIV has evolved the unique and deadly properties that lead to AIDS».

Les concepts fondamentaux — parmi lesquels certains sont supposés inconnus, donc reformulés — doivent être introduits dès le début. Dans le premier tiers, on trouve *HIV*, *HTLV-I*, *HTLV-II*, *STLV*, *immortality*, *cross-reactivity*, *cross-reactive epitopes*, *simian AIDS*, et *simian immunodeficiency virus*. La suite du texte est essentiellement un développement concernant l'histoire de la découverte des relations entre plusieurs de ces concepts. On assiste, à partir de là, à l'introduction très épisodique d'un petit nombre de nouveaux termes (avec leurs reformulations) renvoyant à des concepts périphériques au débat (*exogenous* ou *myxoma*) et à la reprise des termes reformulés au début. Dans ce texte d'une longueur de 4 410 mots, les termes reformulés au début sont ensuite amplement ré-utilisés : ainsi, on retrouve 111 fois *HIV*, 53 fois *SIV*, 34 fois *AIDS*, 18 fois *HTLV* et 13 fois *STLV*, tous très bien répartis tout au long du texte.

On pourrait donc penser qu'il existe au moins deux types de textes : l'un avec reformulations au début, l'autre avec reformulations à la fin. Mais les choses ne sont pas si simples. Ainsi, le texte n° 1 : il commence par un exposé très général sur le fait que le SIDA peut être transmis au fœtus par la mère. Le mode de transmission est ensuite présenté à l'aide de quelques concepts, puis une nouvelle étude est décrite, en termes généraux. La nécessité d'une modélisation est évoquée : la présentation minimale de l'anatomie du cortex rend inévitable le recours à des termes (*cortical columns*, *germinal matrix*) qui sont reformulés. Vient ensuite un développement sur les résultats des études et leurs implications. Le processus d'infection est décrit de nouveau, plus précisément, à la lumière de ces expériences : de nouvelles hypothèses sont émises, impliquant une nouvelle description anatomique. Le texte s'achève sur des considérations d'ordre sociologique. On voit, à travers cette brève description de l'organisation du texte, que le mouvement général peut être caractérisé par un apport ponctuel d'information suivi d'une discussion, de type logique, à partir de cette information nouvelle. Les termes nouveaux, ceux qui sont reformulés, ne sont donc pas regroupés au début ou à la fin du texte, mais introduits par vagues, les développements intermédiaires restant vides de reformulations.

Les observations ci-dessus montrent bien que les modes de présentation de la connaissance sont multiples, même si l'on s'en tient à un critère classificatoire aussi peu élaboré que celui de la topographie des reformulations, appuyé sur l'hypothèse de la progression linéaire, dans le texte, du connu vers l'inconnu.

Il convient, au stade où nous sommes, de nous interroger sur ces concepts de *découverte* et d'*inconnu* évoqués plus haut. S'agit-il d'une mé-connaissance des termes, ou des concepts, ou des deux? Cet *inconnu* dont on parle est-il un inconnu conceptuel ou seulement terminologique?

Nous verrons plus tard, en examinant le contenu des reformulations, qu'un texte de vulgarisation peut être informatif sans être fortement terminologisé. On conçoit bien que l'évaluation de la quantité d'information nouvelle susceptible d'être assimilée par le lectorat doit être une préoccupation constante du scripteur. Il est donc vraisemblable qu'il cherchera à jouer sur les deux tableaux (diffusion d'information «conceptuelle» et mise à disposition d'une terminologie), avec le souci de n'en privilégier aucun, tout en sachant que des interactions entre les deux sont inévitables.

II. LA TOPOGRAPHIE DES REFORMULATIONS PAR RAPPORT AUX DIFFÉRENTES OCCURRENCES D'UN MÊME TERME

La distinction entre inconnu conceptuel et inconnu terminologique permet aussi de rendre compte de la topographie des reformulations par rapport aux différentes occurrences d'un même terme. Si un terme doit faire l'objet d'une reformulation, celle-ci doit logiquement apparaître à sa première occurrence, seule garantie que le terme sera suffisamment compris dès ce moment-là, quel que soit l'endroit du texte où il apparaît pour la première fois. C'est effectivement ce qu'on observe dans la très grande majorité des cas. Si l'on s'intéresse aux seuls termes reformulés les plus fréquents du texte n° 2 (*retrovirus, HTLV, HIV, LAV, T4, T8, reverse transcriptase, lymphadenopathy, T-lymphocytes, macrophage*), on verra que la totalité des reformulations intervient lors de la première occurrence du terme.

Il existe néanmoins, à ce niveau aussi, des cas de déviance où ce n'est pas la première occurrence du terme dans le texte qui supporte la reformulation.

Certaines contraintes syntaxiques peuvent empêcher l'insertion de la reformulation dans la phrase qui supporte la première occurrence du terme : si la reformulation risque de «distendre» exagérément les liens syntaxiques entre les éléments qui doivent rester liés, elle peut être reportée à la phrase suivante. C'est ce qui se passe pour *opportunistic* dans le texte n° 7 :

«For instance, the lethal opportunistic infection *Pneumocystis carinii* pneumonia (PCP), which has been a hallmark of AIDS, can now actually be prevented with medication given early in the course of HIV disease. (Opportunistic infections are ones that occur because the immune system has broken down.)»

Le même phénomène se produit pour le terme *protozoan* dans le texte n° 6 :

«an extremely rare type of pneumonia caused by the protozoan *Pneumocystis carinii* had been diagnosed in the Los Angeles area. (Protozoans are a type of primitive microorganisms.)»

Dans des structures syntaxiques moins contraignantes, l'usage de parenthèses ou de tirets immédiatement après le terme aurait permis d'insérer la reformulation dans l'entourage immédiat du terme, évitant ainsi la reprise (Chukwu et Thoiron 1989). Par exemple, dans le même texte n° 6, *lymphadenopathy* est reformulé instantanément :

«[...] two unexplained conditions: chronic lymphadenopathy (a condition characterized by enlarged lymph nodes) and a relatively rare malignancy called diffuse, undifferentiated non-Hodgkin's lymphoma.»

Mais d'autres reformulations sont beaucoup plus nettement différées. Ces cas ne peuvent s'expliquer que par le jeu de contraintes rhétoriques : certains concepts sont présentés d'abord par l'intermédiaire de leur seule dénomination, puis développés ultérieurement. La reformulation et la mise en œuvre du concept peuvent éventuellement intervenir beaucoup plus loin dans la séquence discursive. Ainsi, dans le texte n° 8, le terme *ELISA* apparaît-il dans la première partie du texte alors qu'il sera reformulé plus d'une page plus loin dans le cadre d'un paragraphe consacré aux tests de dépistage. Dans le texte n° 4, *lymphocytes T4* est utilisé dès la première page, sans reformulation, alors qu'une distinction détaillée entre *lymphocytes B* et *T* sera faite deux pages plus loin.

Si certaines reformulations sont retardées, d'autres sont tout simplement absentes, certains termes n'étant jamais reformulés dans le cours du texte. C'est vraisemblablement le contenu sémantique du terme en question et son adéquation avec la compétence supposée du lectorat qui peuvent expliquer ces absences. Dans le texte n° 2, *AIDS*, *DNA* et *RNA* ne sont pas décryptés, et *virus* n'est pas reformulé : leur connaissance peut être raisonnablement supposée. Les raisons pour lesquelles des termes tels que *lymphoma* et *lymphotropic* ne sont pas reformulés alors qu'on propose aux mêmes lecteurs une reformulation de *genome* sont plus difficiles à entrevoir. Y a-t-il manque de cohérence chez le scripteur, ou, plus vraisemblablement, souci d'économie et d'efficacité par rapport au but poursuivi ? Ces questions pourront être mieux posées après un examen du contenu sémantique des reformulations.

III. LE CONTENU DES REFORMULATIONS

Outre la classification basée sur la nature des relateurs proposée dans Chukwu et Thoiron (1989), on peut esquisser une typologie des reformulations qui se baserait sur les informations qu'elles apportent par rapport à la «simple» désignation d'un concept par le terme correspondant.

A. REFORMULATIONS ET DÉFINITIONS

Si l'on prend pour référence la définition du terme telle qu'on la trouve dans un dictionnaire spécialisé, on s'aperçoit que, dans notre corpus, la plupart des reformulations donnent une quantité d'informations sémantiques inférieure à celle que donnerait une définition⁵. Il existe certes des reformulations qui se présentent comme des définitions en extension (cf. les *reformulations énumératives par liste exhaustive continue* dans Chukwu et Thoiron 1989), mais elles se trouvent dans des textes didactiques d'initiation à un domaine. Parmi les reformulations de notre corpus, on peut distinguer entre reformulations approximatives et reformulations parcellaires.

Nous appelons *approximative* une reformulation dans laquelle l'hyperonyme servant de mot générique est délibérément tiré vers le haut de la taxinomie. On trouve dans le texte n° 2 un exemple-type de reformulation approximative, celle de *retrovirus* : *a class of infectious agents*⁶.

Nous appelons reformulation *parcellaire* une reformulation où le mot générique est celui qui pourrait figurer dans une définition de dictionnaire mais où les différences spécifiques sont trop peu nombreuses pour constituer une véritable définition. Un exemple nous est fourni, pour la reformulation de *HIV*, par le texte n° 1 : «the virus that causes AIDS», et par le texte n° 5 : «the AIDS virus — HIV».

Dans tous ces cas de reformulation, il est clair que ce qui importe pour l'auteur du texte, ce n'est pas de pourvoir son lecteur immédiatement d'une connaissance solide du

concept, mais de lui transmettre la quantité d'information strictement nécessaire pour qu'il puisse tirer parti du texte. Le choix des éléments d'information présents dans la reformulation dépend donc étroitement du contexte, et c'est seulement dans cette optique qu'on pourrait se livrer à une évaluation de la qualité des reformulations : certes les reformulations sont fragmentaires, mais la question pertinente est de savoir si elles permettent d'avancer, de continuer à lire, de comprendre. Ce qui compte, c'est l'efficacité du message, et non la précision dans l'absolu des équivalents fournis par les reformulations.

Le décryptage des sigles fournit à cet égard une série d'exemples intéressants : la reformulation se borne en général à donner en clair chacun des mots à l'origine du sigle (par exemple, dans le texte n° 2, «the human immunodeficiency virus (HIV)»). Elle est donc exclusivement tributaire de la quantité d'information contenue dans le syntagme d'origine. Si le syntagme est peu éclairant sur le sens de l'entité qu'il désigne, le décryptage du sigle qui le résume n'en dira pas plus. On peut penser que l'auteur éprouvera d'autant plus le besoin de décrypter un sigle qu'il estimera ce décryptage utile pour le lecteur. Il y aurait donc surtout des décryptages de sigles lorsqu'ils apportent une véritable information sémantique.

Dans le texte n° 2, *AIDS* n'est pas décrypté : il s'agit du thème de l'article, et la notion est supposée connue, ne serait-ce que de manière minimale, suffisante en tout cas pour comprendre l'ensemble du texte. *RNA* et *DNA* ne le sont pas non plus, peut-être parce que, n'étant pas spécifiques du sous-domaine, ils ne sont pas indispensables au développement général de l'argument, peut-être parce que leur décryptage n'apporterait guère d'information supplémentaire sur leur sens pour un lecteur qui ne les connaîtrait pas. En revanche, tous les autres sigles — principalement des noms de virus — sont décryptés : leur décryptage met en évidence le fait qu'ils procèdent d'un mode de dénomination en partie systématique, et permet au lecteur d'ébaucher l'exploration de leur taxinomie.

B. REFORMULATION ET STRUCTURE DU DOMAINE

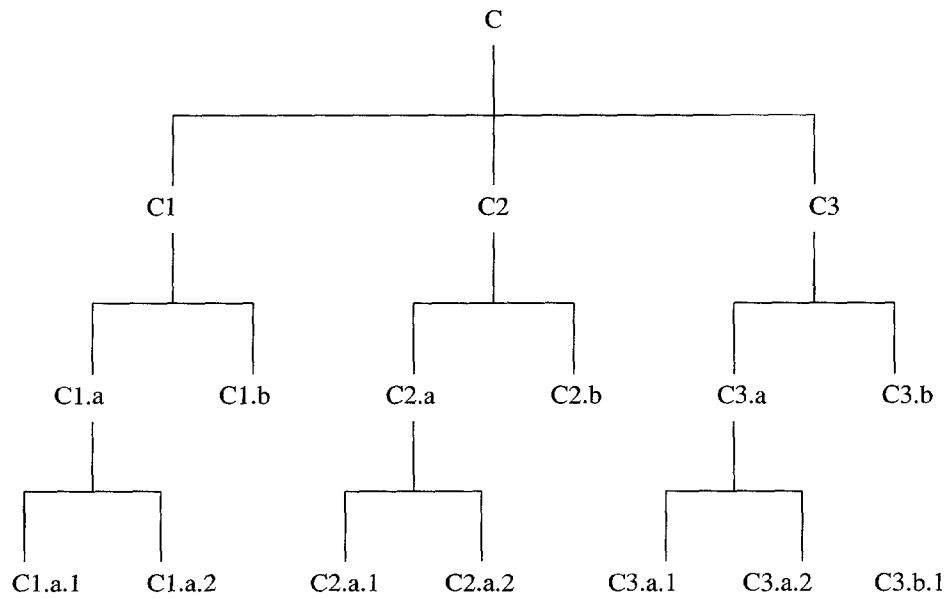
Toute reformulation implique que les termes utilisés dans la séquence reformulatrice soient connus au moment de leur utilisation, soit qu'ils fassent partie du connu supposé, soit qu'ils aient été eux-mêmes reformulés précédemment. Si deux termes reformulés dans un même texte entretiennent des rapports taxinomiques tels que l'un est l'hyperonyme de l'autre, alors celui qui a la plus grande extension doit être reformulé d'abord. C'est ce qu'on constate dans le texte n° 4, où *lymphocyte* est reformulé avant *lymphocyte B* et *lymphocyte T*.

Mais, là encore, il arrive que ce ne soit pas le cas. Le contexte suivant est tiré du texte n° 4 :

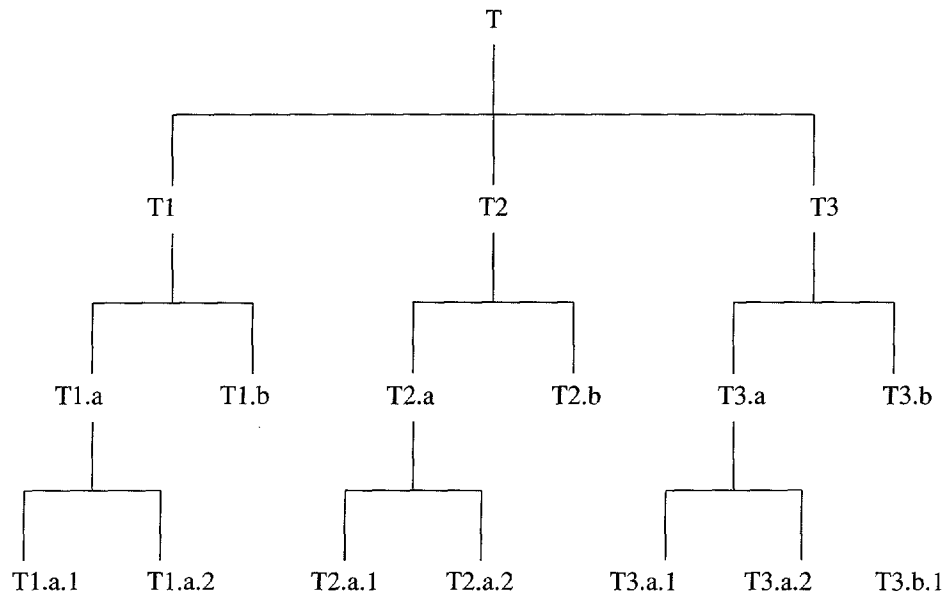
«Les globules blancs qui jouent un rôle clef dans la réponse immunitaire, les lymphocytes, se répartissent en 2 classes principales : les lymphocytes B, ainsi appelés parce qu'ils se développent dans la moelle osseuse (en anglais *bone* signifie os), et les lymphocytes T, qui naissent également dans la moelle osseuse mais dont le développement s'achève dans le thymus.»

Le terme *lymphocyte T4* a été utilisé précédemment et sans définition. Voilà donc un exemple d'incohérence didactique, ou, à tout le moins, de rupture entre le mode d'explication et la structure hiérarchique du domaine. On pourrait imaginer qu'il soit plus simple de reformuler en suivant, dans la chronologie du texte, la hiérarchie conceptuelle, c'est-à-dire selon l'ordre : lymphocyte, lymphocyte B/lymphocyte T, lymphocyte T4.

À ce stade, une tentative de formalisation peut être utile. La structure conceptuelle d'une partie du domaine est représentée comme suit :



À cette structure conceptuelle correspond une structure terminologique qui peut être représentée de la même façon:



À titre d'exemple de correspondance entre les deux structures, le concept C est dénommé par *infectious agent* (T), le concept C1 par *virus* (T1), le concept C1.a par *retrovirus* (T1.a), le concept C1.a.1 par *HIV-1* (T1.a.1).

En règle générale, le texte scientifique n'a pas pour but principal de présenter une terminologie complète, c'est-à-dire de faire correspondre à chaque concept, à chaque niveau de la hiérarchie conceptuelle, un élément en T. On vient de voir que la précision des équivalences n'est pas déterminante. Une autre observation va dans le même sens. Il s'agit de la présence de nombreux vides terminologiques qui ne sont pas nécessairement associés à des vides conceptuels. À titre d'exemple, voici un extrait du texte n° 4 :

«[...] il [le lymphocyte B] donne naissance à un clone de lymphocytes B dont la membrane porte des molécules d'anticorps qui agiront comme récepteurs spécifiques de l'antigène à combattre. Certaines de ces cellules, les plasmocytes, sécrètent des anticorps; d'autres, dont la durée de vie est plus longue, sont en partie responsables d'une immunité durable et empêchent la récurrence de nombreuses infections. Ces cellules constituant la *mémoire* du système immunitaire [...]».

Le clone de *lymphocytes B* n'est pas dénommé; il n'est pas défini scientifiquement, mais certaines de ses propriétés sont décrites. Il n'y a donc pas de vide conceptuel : il existe un nœud en C. En revanche, le vide terminologique — au moins dans le texte — impose, lors de l'anaphore, la reprise par un terme placé assez haut dans la taxinomie (*cellule*).

Le cas de *plasmocytes*, bien que différent, est également révélateur. Le concept est, cette fois, dénommé : il existe une correspondance entre un C et un T. Mais ni l'hyperonyme ni les co-hyponymes ne le sont. Pourtant, là encore, les propriétés de ces objets sont énumérées, ou, plus précisément, celles des propriétés qui intéressent le thème développé à ce moment-là (durée de vie de ces cellules et, corollairement, durabilité de l'immunité). Pour reprendre ces concepts non dénommés, l'auteur doit de nouveau avoir recours à un terme du haut de la taxinomie, et, inévitablement, il emploie à nouveau le syntagme *ces cellules* qui renvoie ainsi, à très bref intervalle, à deux concepts différents. Néanmoins, ancré dans la logique discursive du texte et dans sa chronologie, le procédé fonctionne sans créer l'ambiguïté. Le vide terminologique ne semble donc pas être nécessairement un facteur disruptif au plan du développement conceptuel.

C. REFORMULATION, ÉCONOMIE ET EFFICACITÉ

Non seulement le vide terminologique n'est pas une gêne; il peut même constituer un avantage. Il peut être plus efficace, plus rapide, de ne pas dénommer, mais de décrire *la* propriété (parmi plusieurs) ou *le* mode de fonctionnement qui permettra d'avancer vers une meilleure connaissance des concepts. Un concept, de type C1.a par exemple, n'est pas dénommé dans le texte; il est évoqué, dans le cadre d'une reformulation concernant un terme de type T1.a.1 (par exemple, *HIV*), à l'aide d'un autre terme, de type T1 ou T (par exemple *virus* ou *infectious agent*), et de certains des spécifiques de T1 ou de T qui réfèrent à certaines propriétés de C1 ou de C.

Il convient toutefois de nuancer les avantages du vide terminologique, de deux manières au moins. Tout d'abord, certains concepts sont bel et bien dénommés. Pourquoi ne pas faire l'économie du terme si, comme on vient de le voir, le concept correspondant peut être décrit dans la reformulation a minima d'un autre terme? D'autre part, ce que nous appelons métaphoriquement «vide terminologique», et qui correspond en fait à un nœud sans étiquette dans la structure taxinomique est-il vraiment impossible à circonscrire par des moyens langagiers?

À propos du premier point, revenons sur l'exemple de *plasmocyte* : le terme est fourni au lecteur parce que le concept sera ré-utilisé dans la suite du texte. En revanche, les autres (*cellules*) ne sont plus évoquées. Nous avons donc le cas de figure classique où

seul le concept ré-utilisé est dénommé puis ré-employé. Cette procédure est évidemment fondée sur l'économie des moyens.

Il se peut aussi que la présentation du terme *T'* soit facilitée par la reformulation antérieure du terme *T* ou par la mise en évidence de l'ensemble de la taxinomie, y compris à l'aide de reformulations. Cette situation est illustrée par le cas de *lymphocytes T* dans le texte n° 4. L'auteur lui-même évoque le problème :

«Quand on considère la fonction des lymphocytes *T*, on peut les classer en 4 sous-groupes; en revanche, si l'on ne considère que les molécules présentes à la surface de ces lymphocytes [...] on ne distingue plus que 2 groupes.»

On comprend alors pourquoi, devant traiter prioritairement de la fonction des lymphocytes et sachant que les modes d'action de ceux-ci dépendent de la présence de telle molécule, une double classification doit être présentée. Cette double classification serait bien difficile à faire, et au mieux peu économique, sans le recours aux termes.

Pour le deuxième point, s'il est vrai qu'un texte scientifique de vulgarisation peut (et doit) se passer d'une partie de la terminologie du domaine qu'il traite (ce qui n'est pas le cas du texte pour spécialistes), il fait néanmoins appel à une série de procédures linguistiques qui aident le lecteur à se repérer, fût-ce approximativement. Ce dernier doit savoir, en effet, à quelle zone de la hiérarchie taxinomique se situe le concept abordé. C'est pourquoi on arrive à des présentations simplifiées des taxinomies où un terme est, en quelque sorte, le représentant de tout un niveau, de tout un groupe de co-hyponymes. L'exemple de la série des *HTLV* est révélatrice: l'un d'eux, généralement *HTLV-I*, est présenté et reformulé, les autres ne font l'objet d'aucune reformulation.

On voit, à la lumière de ces quelques exemples, que le processus de reformulation est employé aussi bien lorsqu'un terme devra être ré-utilisé que lorsqu'on souhaite se passer de l'usage d'un terme apparenté à celui qui est reformulé. Cette situation cesse d'être paradoxale lorsqu'on l'envisage du point de vue de l'économie et de l'efficacité discursive. Dénommer trop peu, c'est courir le risque d'être peu clair, vouloir tout dénommer, c'est encombrer inutilement le texte.

D. TYPOLOGIE DES REFORMULATIONS EN FONCTION DE LEUR CONTENU

L'examen du contenu des reformulations fait apparaître la possibilité d'établir une typologie fondée sur le décalage éventuel entre le niveau d'enchâssement du concept dans sa taxinomie et le niveau d'enchâssement, dans la hiérarchie terminologique, des termes utilisés pour renvoyer au concept.

À ce premier type de critère, qui est d'ordre purement hiérarchique, on peut ajouter un second type qui concerne la nature de l'information complémentaire apportée sous forme de descriptions ou d'énumérations de propriétés. On verra ainsi une distinction possible entre des ajouts de type classificatoire non terminologisés, et des ajouts de type «fonctionnaliste». Ces derniers sont très fréquents dans les textes de vulgarisation où c'est la présentation du mode de fonctionnement d'un concept, d'un objet, qui constitue le moyen le plus simple de familiariser un lecteur avec la connaissance de ce concept ou de cet objet. Les exemples tels que «*HIV*, the virus that causes *AIDS*» sont fréquents. La double classification que nous proposons permet de montrer que la reformulation de *HIV* inclut à la fois un terme ne dominant pas directement *HIV* dans la taxinomie, et l'énoncé d'une propriété concernant le mode d'action, de fonctionnement, de ce rétrovirus. Dans la reformulation de *protozoans* sous la forme «a type of primitive microorganisms», le saut dans la hiérarchie terminologique est encore présent. Cependant l'ajout informationnel n'est pas de type fonctionnaliste, mais classificatoire.

Il est donc possible de raffiner notre typologie des reformulations en fonction de leur contenu. L'étude de notre corpus nous permet de distinguer deux nouveaux types de reformulations : les reformulations «classificatoires» et les reformulations «fonctionnalistes». Les intersections entre ces deux nouveaux types et les deux types précédemment mentionnés (parcellaires et approximatives) restent à explorer.

L'auteur d'un texte scientifique se trouve ainsi placé devant un certain nombre de choix : utiliser un concept/ne pas l'utiliser, dénommer ce concept par un terme/ne pas le dénommer, reformuler le terme/ne pas le reformuler, utiliser une reformulation parcellaire/approximative, utiliser une reformulation classificatoire/fonctionnaliste. Mais ces options ne se présentent pas comme un simple enchaînement; chaque décision prise pour un terme T1 a une influence non seulement sur la décision suivante pour le même terme, mais aussi sur les décisions concernant les autres termes T2, T3, ..., Tn qui se trouvent dans la suite du texte.

CONCLUSIONS

L'étude des reformulations met en évidence la complexité du jeu rhétorique utilisé dans certains textes scientifiques, en particulier les textes de vulgarisation. L'auteur de tels textes vise avant tout une efficacité didactique qui passe par l'introduction des concepts, et éventuellement des termes, appropriés au domaine traité. Il est amené à évaluer à chaque instant, et sans perdre de vue la cohérence générale de son texte, le degré de difficulté de l'information à transmettre et la quantité d'effort qu'il peut demander à son lecteur : la reformulation est pour lui un moyen efficace de parvenir à ses fins.

Il semble bien, à la lumière de notre corpus en tout cas, que ce souci d'efficacité affecte la structure même des textes. Une typologie des textes scientifiques élaborée en fonction des reformulations est donc envisageable. Mais les critères à mettre en œuvre sont multiples : ils doivent être envisagés du triple point de vue de la nature, de la topographie et de la densité.

S'agissant de la nature des reformulations, la typologie fondée sur des critères formels mise en place dans Chukwu et Thoiron (1989) peut servir de point de départ. En y joignant des mesures statistiques concernant le nombre total de reformulations et le nombre de chaque type, il est possible de mettre en évidence l'existence de différentes catégories de textes. La distinction établie, selon le contenu des reformulations par rapport à la structure hiérarchique du domaine, entre les reformulations approximatives, parcellaires, classificatoires et fonctionnalistes peut, elle aussi, constituer un critère de classification.

Un autre critère concerne la topographie des reformulations. On a vu qu'il est possible de distinguer trois grandes familles de textes scientifiques selon que les reformulations sont concentrées au début ou à la fin, ou encore éparpillées tout au long du texte. Dans la mesure où nous savons maintenant construire un indice de topographie des items répétés (Thoiron et Serant 1989), cet indice peut servir à affecter précisément chaque texte à l'une ou l'autre des trois catégories.

Enfin, certaines données quantitatives, faciles à obtenir, ne doivent pas être négligées. Le taux de terminologisation d'un texte, défini simplement comme le rapport du nombre de termes au nombre total de mots-occurrences, même s'il ne fait pas intervenir directement la reformulation, est un bon indicateur du niveau de spécialisation. À partir de là, le taux de reformulation, défini comme le rapport du nombre de reformulations au nombre de termes, peut être calculé. On obtient alors un indice supplémentaire permettant de situer le texte sur un continuum dont les extrêmes sont occupés d'un côté par les textes pour spécialistes du sous-domaine, et de l'autre par les textes de vulgarisation pour grand public.

On a bien vu que la nature même de la communication de l'information à un public dont l'homogénéité ne peut être que supposée, et la gestion de procédures discursives visant à l'efficacité dans l'économie, conduisent à des situations de reformulation très diverses — et pas toujours évidemment cohérentes du strict point de vue de la logique. Il est assez naturel, dans ces circonstances, que la typologie des textes scientifiques ne puisse s'établir simplement en fonction d'un tout petit nombre de critères. Nous proposons plutôt, compte tenu des observations faites sur notre corpus, de recourir à la mise en place d'un réseau de critères qui permettrait de situer un texte donné dans un espace multidimensionnel. Le rôle joué par la reformulation dans la mise en place de ces critères est, nous croyons l'avoir montré, considérable.

NOTES

1. Un exemple tiré de notre corpus d'écologie des eaux continentales : «Wavelength and period are mainly determined by the fetch, or distance of open water over which the wind blows.»
2. Voir aussi l'essai de classification de quelques indices révélateurs dans Otman (1989 : 70).
3. Notre corpus comprend 8 textes désignés ci-après par les numéros correspondants :
 n° 1 : «The Infant Brain», *Discover*, August 1989, pp. 30-32;
 n° 2 : «AIDS in 1988», *Scientific American*, October 1988, pp. 41-48;
 n° 3 : «The Human Immune Deficiency Virus (HIV): An Update», *Annales de l'Institut Pasteur*, 1987, 138, pp. 34-11;
 n° 4 : «Le système immunitaire et le SIDA», *Pour la science*, février 1986, pp. 40-51;
 n° 5 : «The Origins of the AIDS Virus», *Scientific American*, October 1988, pp. 64-71;
 n° 6 : «The Epidemiology of AIDS in the U.S.», *Scientific American*, October 1988, pp. 72-81;
 n° 7 : «HIV Infection: The Clinical Picture», *Scientific American*, October 1988, pp. 90-98;
 n° 8 : «L'étiologie virale du SIDA et son impact en santé publique», *JAMA*, 1985, 10, pp. 414-417.
4. Ces reformulations concernent des concepts qui ont été présentés et dénommés dans la 1^{re} tranche, où le terme correspondant a fait l'objet d'une reformulation. On sait que la terminologie du SIDA a fluctué, reflétant la compétition entre équipes françaises et américaines. Dans la 8^e tranche, l'auteur aborde ces questions de conflit terminologique : il est donc conduit à reprendre des éléments de la 1^{re} tranche, tels que *HIV* et *HTLV*.
5. Sur la définition en terminologie, voir Clas (1985).
6. On peut rapprocher cette technique de celle qui est couramment utilisée dans les définitions spontanées dites *populaires*, où la varlope est un machin, ou un truc plutôt qu'un rabot, et l'escargot est une bestiole plutôt qu'un gastéropode (Béjoint 1989 : 341).

BIBLIOGRAPHIE

- BÉJOINT, Henri (1989)]: «Traditions et innovations dans les dictionnaires monolingues généraux de langue anglaise depuis 1960», Thèse de Doctorat d'État, Université de Franche-Comté.
- CHUKWU, Uzoma et Philippe THOIRON (1989): «Reformulation et repérage des termes», *La Banque des mots*, numéro spécial CTN, INaLF, CNRS, pp. 23-50.
- CLAS, André (1985): «Les problèmes de définition en lexicographie et en terminologie», *Guide de recherche en lexicologie et terminologie*, Paris, ACCT, pp. 77-78.
- OTMAN, Gabriel (1989): «Terminologie et intelligence artificielle», *La Banque des mots*, numéro spécial CTN, INaLF, CNRS, pp. 63-95.
- THOIRON, Philippe et Daniel SERANT (1989): «La topographie des termes», *Meta*, 34-3, pp. 435-442.
- VIDALENC, Jean-Louis (1989): «Thème et discours scientifiques», *Travaux, L'Anaphore*, C.I.E.R.E.C., Université de Saint-Étienne, pp. 111-129.