

Problèmes de choix dans l'établissement d'une fiche terminologique

Anne Judge and Patricia Thomas

Volume 33, Number 4, décembre 1988

Symposium AILA 1987, Sydney

URI: <https://id.erudit.org/iderudit/003273ar>

DOI: <https://doi.org/10.7202/003273ar>

[See table of contents](#)

Publisher(s)

Les Presses de l'Université de Montréal

ISSN

0026-0452 (print)

1492-1421 (digital)

[Explore this journal](#)

Cite this article

Judge, A. & Thomas, P. (1988). Problèmes de choix dans l'établissement d'une fiche terminologique. *Meta*, 33(4), 520–534. <https://doi.org/10.7202/003273ar>

PROBLÈMES DE CHOIX DANS L'ÉTABLISSEMENT D'UNE FICHE TERMINOLOGIQUE

ANNE JUDGE ET PATRICIA THOMAS
University of Surrey, Guilford, U.K.

The comments which follow are the result of work carried out at the University of Surrey by a team of research workers studying the problems posed by the development of a multilingual terminological data bank with English as the source language. The team comprises a terminologist, several linguists as advisers, and computer scientists. This paper explores the subject in depth and attempts to solve the problems encountered when establishing the terminological record format.

0. INTRODUCTION

Travailler pour une banque de données terminologiques c'est d'abord faire partie d'une équipe *multidisciplinaire*. Le pivot de l'équipe est le *terminologue* qui travaille en association d'une part avec un certain nombre de *spécialistes* dans un domaine donné : ceux-ci fournissent au terminologue les termes dont il a besoin ainsi que les données qui s'y rapportent ; et il travaille d'autre part avec des *informaticiens* qui ont pour tâche de mettre en réserve les données fournies par les spécialistes et réinterprétées par le terminologue en fonction du problème ; ces données doivent être ensuite rendues accessibles au public. Pour cela les informaticiens doivent également travailler avec des *linguistes*.

Le savoir auquel il est fait appel se situe donc à trois niveaux différents : il s'agit, au niveau de la conceptualisation de la banque de données terminologiques, de *la représentation du savoir et de son déploiement*, ce qui relève et des Sciences humaines et de l'Intelligence artificielle ; il s'agit de l'interprétation de ce savoir au niveau *linguistique* ; et finalement, au niveau de la réalisation, il s'agit du problème de mise en réserve et re-tracage de l'information, ce qui implique des relations réciproques entre linguistique et informatique.

Il va de soi qu'une telle coopération présente des problèmes, car il y a forcément des malentendus qui peuvent se produire entre spécialistes appartenant à des domaines si différents. Cela dit, il faut néanmoins qu'il y ait coopération parce qu'il est impossible qu'une seule personne, ou qu'un groupe homogène de personnes, puisse arriver au niveau requis dans toutes les branches du savoir entrant en jeu pour la mise en place d'un tel projet.

Ayant donc situé la matière, nous allons maintenant traiter les questions suivantes :

1. Quels sont *les buts* que se donnent les différentes banques de données terminologiques ?
2. Quels sont les principes fondamentaux sur lesquels repose la terminologie ?
3. Quelles sont *les informations* que les utilisateurs pourraient s'attendre à trouver ?

Ces questions serviront à établir les choix qui doivent être faits lors de la mise en place d'une nouvelle banque terminologique ; en l'occurrence il s'agit d'une banque de données terminologiques ayant l'anglais comme langue source, banque qui est, en ce moment, au stade de projet.

1. LES BUTS QUE SE DONNENT LES DIFFÉRENTES BANQUES DE DONNÉES TERMINOLOGIQUES

Définition : Une banque de données terminologiques ou, plus simplement, une banque terminologique, est une collection de vocabulaires spécialisés mis sur ordinateurs et comprenant les nomenclatures et les termes — normalisés ou non — en même temps que l'information requise pour identifier ces éléments.

1.1 LES DIFFÉRENTES BANQUES TERMINOLOGIQUES ET LEURS BUTS

Celles-ci se sont multipliées depuis 1960 (voir Hvalkof, 1984). La première conférence de ces banques eut lieu à Vienne en Autriche en 1979, sous l'égide d'INFOTERM, établissement créé en 1971 par l'Institut de normalisation autrichien en vue d'encourager une coopération internationale dans la recherche en terminologie. Ce développement rapide a eu lieu à cause de besoins divers précis ; ceux-ci n'ayant pas été toujours les mêmes, les banques terminologiques créées diffèrent de par leurs buts. On peut ainsi distinguer celles qui sont nées pour résoudre des problèmes de traduction et celles qui sont nées pour aider les spécialistes à s'y retrouver dans un domaine précis.

1.2 LES BANQUES DE DONNÉES TERMINOLOGIQUES ORIENTÉES VERS LA TRADUCTION

Les premières banques créées dans le cadre du Marché commun devaient fournir un outil pour aider le traducteur au niveau de la recherche de mots équivalents dans les différentes langues du Marché commun et pour aider le réviseur à obtenir des traductions cohérentes au niveau du vocabulaire, même lorsque les textes sont traduits par des personnes différentes. Il s'agit essentiellement d'un dictionnaire sur ordinateur. EURO-DICAUTOM fut fondé en 1973 dans le but d'aider tous les traducteurs travaillant dans le cadre des institutions du Marché commun, mais les renseignements sur les quelque 310 000 termes contenus dans cette banque en 1984 sont maintenant à la disposition des usagers. Dans un autre contexte, Siemens, en Allemagne, est une firme s'occupant d'électronique : elle a un million de pages de documentation à traduire par an, d'où la création de sa banque de données terminologiques TEAM pour aider ses traducteurs dans ce domaine bien précis. LEXIS est une autre banque de données terminologiques allemande ; elle a été mise sur pied par le *Bundessprachenamt* de la République fédérale et elle compte à l'heure actuelle 1 433 000 termes en allemand ; 62% de ceux-ci ont des équivalents en anglais, 22% en russe, et 15% en français.

1.3 LES BANQUES TERMINOLOGIQUES ORIENTÉES VERS LES SPÉCIALISTES

DANTERM, la banque danoise, et TERMDOK, la banque suédoise, sont spécialement intéressantes parce qu'il s'agit de banques tout à fait différentes de celles mises sur pied uniquement sous forme de dictionnaires informatisés. Dans le cas de TERMDOK, mise en place en 1968 par le Centre suédois de nomenclature technologique, *Tekniska Nomenklaturcentral* (TNC), il s'agit de fiches informatisées se référant à des *concepts* ; c'est donc un genre de *thesaurus* dans le sens anglais du terme. Ces fiches peuvent également contenir des détails tels que des adresses utiles, et des références bibliographiques. Il s'agissait au départ d'aider le gouvernement suédois à introduire dans sa propre langue les néologismes scientifiques, mais ces fiches sont également utilisées à cause de

leur contenu non strictement linguistique par des *spécialistes* dans un domaine précis, TERMDOK entretenant de bons rapports avec le monde industriel et ayant un personnel comprenant des ingénieurs. Le côté original de cette banque est qu'elle cherche à analyser et à gérer une terminologie d'origine multilingue, aussi une fiche peut-elle inclure jusqu'à dix langues. Cela est d'une importance vitale dans un pays fortement industrialisé et exportateur, mais parlant une langue peu connue.

DANTERM part aussi du concept plutôt que du mot ; dans ce cas-ci l'aspect thésaurus est lié aux données linguistiques associées aux concepts. DANTERM adhère en principe, c'est-à-dire toutes les fois que cela est possible, aux principes de l'École terminologique de Vienne (voir 2.2). Chaque fiche donne des détails fort nombreux sur le terme étudié, renseignements qui vont au-delà de la simple définition. Ces renseignements peuvent fournir une aide précieuse au traducteur se trouvant devant un terme qui lui est inconnu, ainsi qu'aux spécialistes.

Le Canada étant un pays officiellement bilingue, tous les documents officiels doivent forcément être rédigés dans les deux langues. Cela a entraîné la création de la Banque de terminologie du Canada (TBC) ou Termium, dont le but est la recherche d'équivalents anglais-français. Il s'agit donc essentiellement d'un outil de traduction, mais Termium inclut aussi une classification thésaurale de ses termes qui sont regroupés en 55 champs. Selon les derniers chiffres, Termium compte plus de 1 320 500 termes (voir 1.4 pour la Banque de terminologie du Québec).

1.4 LEUR RÔLE AU NIVEAU DE LA NORMALISATION DES TERMES

Pour que la communication internationale puisse être possible au niveau scientifique, il est essentiel qu'il y ait une normalisation des termes qui se rapportent à ces domaines, d'où l'établissement d'organismes tels que DIN qui est l'organisation visant à normaliser les termes en Allemagne fédérale. En France, l'Association française de normalisation et en Grande-Bretagne, la *British Standards Institution* jouent le même rôle. L'Association française de normalisation (AFNOR) a créé à cet effet NORMATERM au début des années 70. Au début il s'agissait surtout de rendre disponibles toutes les décisions en matière de terminologie publiées dans le Journal officiel ainsi que les recommandations de l'ISO (voir 2.3). NORMATERM contient également le vocabulaire de la Commission internationale d'électronique en français et en anglais. Celui-ci est important, représentant à la fois un vocabulaire normalisé et international. Cette banque contenait en 1984 environ 150 000 termes ainsi que leur définition. Les fiches comprennent entre autres, pour chaque terme, le terme générique s'il y a lieu et la numération du concept. Là où les fiches sont bilingues, sont inclus les synonymes du terme, par ordre de préférence, pour chaque langue. L'AFNOR produit également des glossaires soit par ordre alphabétique soit par matière, donnant le terme, la définition, les relations hiérarchiques et les termes associés. C'est donc là une création très importante.

Les banques de terminologie peuvent donc servir de réservoir pour tous les termes normalisés et d'outil permettant la comparaison de ceux-ci ainsi que leur mise à jour. Cela n'empêche pas d'ailleurs ces banques d'inclure également des termes qui n'ont pas été approuvés ou qui sont devenus archaïques, mais cet aspect du problème est toujours explicite.

Notons que malgré ces organismes, le problème de la normalisation des termes est loin d'être résolu ; il suffit de penser au problème des nombres : le billion et le trillion ne se réfèrent pas au même nombre de zéros en France et en Grande-Bretagne ; de plus, les valeurs attachées à ces mots ont changé, en France en 1948, et en Angleterre dans les années soixante. La standardisation au niveau international n'est toujours pas à l'ordre du jour dans ce domaine.

Il semblerait donc que le besoin de normaliser les termes soit évident, mais que ce soit loin d'être fait malgré une relative bonne volonté internationale à cet égard. Le premier organisme de ce genre à vocation internationale fut l'ISA (*International Federation of National Standards Association*) ; cette fédération fut fondée en 1937, par Eugen Wüster, le père de la terminologie moderne. L'ISA fut remplacée après la Deuxième guerre mondiale par l'ISO. Dans tous ces cas il s'agit essentiellement de standardisation de termes scientifiques et techniques.

Le cas de la Banque terminologique du Québec (BTQ), créée en 1974 par l'Office de la langue française à Québec, est légèrement différent puisque son but est essentiellement de franciser tous les termes de la vie courante. On a donc affaire à une normalisation faite dans un but politique et culturel interne au pays plutôt qu'à un effort au niveau de la communication internationale. Mais les travaux de recherche entrepris par cette banque sont spécialement intéressants vu que tous les domaines sont considérés aussi bien du point de vue pratique que théorique.

D'une manière générale, le niveau de normalisation dépend du degré de spécialisation du domaine considéré, ainsi que de son évolution géographique et temporelle. Ainsi le domaine de la virologie, par exemple, a connu une croissance rapide depuis le XIX^e siècle dans le monde occidental. Le résultat est une terminologie qui a évolué à un rythme comparable et de manière semblable dans tous les pays scientifiquement avancés. La virologie est donc aujourd'hui une science dont la terminologie est à un stade avancé de normalisation. Dans d'autres disciplines, cependant, il y a des exceptions à ce principe ; certains termes de l'art dentaire, par exemple, sont très difficiles à traduire, surtout lorsqu'il s'agit de domaines très spécialisés comme celui des soins des gencives.

Il y a un manque de normalisation dans les domaines qui ont évolué au niveau national à des rythmes différents. Dans ce cas c'est en général la langue dominante qui produit le néologisme. Mais celui-ci peut se retrouver plus tard en concurrence avec un terme d'origine locale. Ainsi l'informatique, en français, admet dans certains cas plusieurs termes se référant au même concept : à un certain moment, par exemple, les termes *software* et *logiciel* ont été employés simultanément. De même, dans un autre domaine, si la traduction officielle de *bulldozer* est *bouteur*, le terme anglais continue à être employé en français, francisé sous la forme *bouledozeur* d'après les instructions de la commission.

Il y a en fait, à l'heure actuelle, un très fort mouvement d'internationalisation des néologismes. Un terme emprunté à l'anglais tend à être traduit en français, très souvent par un calque, tel que *mole* et *taupe* employés dans le domaine des travaux publics ; ou bien par une traduction mot à mot respectant l'ordre habituel dans la langue d'arrivée comme c'est le cas pour *acid rain* et *pluie acide*. Ces solutions tendent ensuite à être adoptées dans les autres langues romanes qui deviennent, de ce fait, non plus des langues sœurs mais des langues jumelles. Ce phénomène n'est pas nouveau. Ainsi l'anglais *sleeping-car* donne *wagon-lit* en français, *waggonne letti* en italien, et *coche cama* en espagnol. Ce phénomène tend même à s'étendre aux langues européennes non romanes telles que l'allemand, avec *Schlafwagen* et le suédois *sovvagn*.

1.5 LA NORMALISATION DES TERMES SE RÉFÉRANT À DES RÉALITÉS NATIONALES

Il existe un autre problème de normalisation important au niveau et de la communication et de la traduction ; c'est celui de la traduction des termes employés dans les domaines administratifs, juridiques, politiques et sociaux, propres à des pays particuliers. C'est là un problème qui affecte spécialement les pays parlant une langue autre que leur langue maternelle pour communiquer avec le monde extérieur. Les pays nordiques en sont un exemple : comment parler en anglais à quelqu'un d'un autre pays au sujet

d'une réalité danoise ? Pour donner un exemple, est-ce que « *The Copenhagen School of Economics* » se réfère au même établissement que « l'École des hautes études commerciales de Copenhague » ou s'agit-il de deux établissements ? En fait il s'agit d'un seul établissement mais ces deux traductions du même terme créent une ambiguïté au niveau de la communication.

C'est pour cela que les pays nordiques font des recherches approfondies dans ce domaine : le Danemark avec sa banque de terminologie DANTERM créée par l'École des hautes études commerciales de Copenhague déjà citée ; la Suède avec sa banque de terminologie TERMDOK (voir 1.3) à Stockholm ; la Norvège avec le « *Råddet for teknisk terminologi* » (RTT) à Oslo, ainsi que sa banque de terminologie ; la Finlande (*Vaasa Högskola, TSK, Forsknings centralen för de inhemska spraken, and Sjöfartsstyrelsen*) ; et l'Islande (*Islensk Málefud*). Depuis 1976, NORDTERM s'efforce de regrouper toutes les activités terminologiques de ces pays.

Le travail effectué dans ce contexte a un intérêt spécial pour les pays du Tiers Monde qui doivent le plus souvent faire face aux mêmes problèmes. Il s'agit d'ailleurs d'un travail particulièrement important au niveau de la théorie de la terminologie parce qu'il tient compte non seulement des sciences techniques mais aussi des sciences « douces », c'est-à-dire les sciences sociales qui présentent des problèmes tout autres au niveau de la conceptualisation.

On pourrait s'attendre dans ce contexte à ce que le droit constitue un domaine spécialement difficile. Cela est vrai en ce qui a trait aux institutions et aux administrations particulières à chaque état. Mais ce n'est pas le cas pour ce qui va au-delà de leurs propres institutions : le commerce est international et les contrats de vente et d'achat aussi. La terminologie pour ce genre de transactions est habituellement normalisée au niveau national et les équivalences de traduction ne sont pas trop difficiles à trouver.

1.6 CONCLUSION

EURODICAUTOM cherche à couvrir tout le vocabulaire employé dans le contexte des langues du Marché commun tandis que Termium ne s'occupe que du vocabulaire français et anglais. La banque de terminologie TEAM de l'entreprise Siemens à Munich s'intéresse surtout à la langue de l'électronique ; il y a également LEXIS qui fait partie du ministère de la Défense en Allemagne ; cette banque est spécialisée dans le vocabulaire militaire. La Banque terminologique du Québec est orientée vers la francisation de tout le vocabulaire ; NORMATERM en France s'occupe de la normalisation du vocabulaire scientifique et technique. VNIKI en U.R.S.S. s'occupe surtout de termes d'ingénierie, tandis que les banques nordiques sont tout autant intéressées par le problème du découpage du savoir et par celui de sa représentation. Les banques terminologiques représentent donc un ensemble très éclectique : ce sont à la fois leurs buts et leurs domaines qui diffèrent. Aussi, ce qui se fait sentir en ce moment, c'est un grand besoin d'harmonisation.

Ceci dit, on peut néanmoins les regrouper en deux genres de banque : celles qui prennent le *mot* comme point de départ et celles qui prennent le *champ conceptuel*. Commençons par celles qui partent du *mot*. C'est l'exemple classique du dictionnaire sur ordinateur, c'est-à-dire un outil de travail pour le traducteur et le réviseur. Ces dictionnaires sont *multilingues* dès le départ. EURODICAUTOM en est un exemple. Le problème associé avec ce genre de banque est que l'on a affaire le plus souvent à un mélange de mots venant de la langue générale et des langues spécialisées. Or un mot peut correspondre à un sens dans la langue générale et à un ou plusieurs concepts dans la langue scientifique, si le terme est employé différemment dans différents domaines. Comment inclure ces nuances dans toutes les langues incluses, vu qu'elles ne correspondent pas de langue à langue ? D'une manière générale, on tend à aboutir à de longues listes

contenant de nombreux termes dont le traducteur n'a nul besoin pour une question précise.

Il nous semble donc qu'il vaut mieux commencer par un *concept* et non pas par un mot, d'où la deuxième orientation qui prend le *champ conceptuel* comme point de départ. Dans ce cas il s'agit davantage d'un *thésaurus*. La fiche terminologique est donc monolingue au départ, puis des fiches analogues sont rédigées dans les autres langues requises (voir 3.6). Dans ce cas il y a forcément une langue de départ et des langues d'arrivée, principe opposé à celui qui guide l'élaboration d'EURODICAUTOM. Il s'agit cette fois d'un outil de classement pour le spécialiste. Il faut ensuite passer au problème des équivalences, en s'appuyant cette fois sur les données conceptuelles, langue par langue. Ainsi sont combinés les avantages des deux approches.

2. LES PRINCIPES FONDAMENTAUX DE LA TERMINOLOGIE

Si le chercheur part du mot, il fait de la lexicologie ; s'il part du terme il fait de la terminologie. Pour comprendre les principes fondamentaux de la terminologie, il est nécessaire de commencer par des *concepts* et de les représenter par des *termes*.

2.1 L'UNITÉ DE BASE : LE TERME

Le point de départ pour définir le terme est la distinction saussurienne entre le signifiant et le signifié : toute unité terminologique a pour fonction la dénomination, c'est-à-dire la relation entre la forme linguistique et une réalité (abstraite ou concrète) par le biais d'un concept.

Mais comment distinguer entre un « terme » et un « mot » ? Le mot est une entité individuelle qui a une vie à lui, c'est-à-dire une étymologie : son extension sémantique peut s'accroître ou se restreindre, son sens peut varier d'une personne à une autre et ne représente pas un niveau de connaissances.

Le terme a un *sens très précis limité par sa place dans le système*. La valeur d'un terme est définie de manière *synchronique* soit par les associations de normalisation, soit par les organismes professionnels. Il n'a un sens précis que dans le cadre de son domaine particulier. Les termes ne sont maniés que par des spécialistes et représentent le niveau de connaissances atteint dans un domaine. Il peut arriver que le sens d'un terme change, par exemple un « atome » était autrefois une unité indivisible alors que maintenant on sait l'atome divisible, cependant le terme lui-même demeure inchangé malgré l'évolution de son contenu sémantique.

Le terme peut être composé de *un ou plusieurs mots* (par exemple « une touche » et « une touche de piano ») ; il peut également être un sigle ou une abréviation. Mais comment distinguer entre *un terme* et *un syntagme* ? Pour qu'il y ait un seul terme il faut qu'il y ait un seul concept. Par exemple, le groupe *trieuse pondérale automatique de classement pour les œufs* (en anglais *egg grading machine*) est-il un terme ou un syntagme ? Si cette machine est *unique* pour cet emploi, nous avons affaire à un terme ; si ce n'est pas le cas, nous avons affaire à un syntagme. L'allemand sert souvent de test dans ces cas-là : l'équivalent allemand est *Eiersortiermaschine*. La version française constitue donc bien un *terme* malgré le nombre de mots qui le compose. Il ne s'agit pas bien entendu d'un test infaillible, ainsi *appareil balistique* donne *ballistisches Gerät*, mais il s'agit d'un seul concept et par conséquent d'un seul terme dans les deux langues.

Le terme peut correspondre à des *concepts différents dans différents domaines*, ce qui peut prêter à confusion. Ainsi les termes « *nucleus* » en anglais et « *noyau* » en français sont employés à la fois dans le domaine de la biologie et de la physique. Dans ce genre de cas les déplacements de sens se font souvent soit par métaphore soit par méto-

nymie. Ceci n'est pas gênant, vu que le terme est défini par rapport à un domaine spécifique.

2.2 LA TERMINOLOGIE

Les précurseurs furent tous des scientifiques. Les premiers créateurs furent des hommes comme Lavoisier et Berthollet, qui semblent avoir été conscients de la distinction entre concept et terme ; ils semblent également avoir été conscients de l'importance de l'expressivité du terme et de son degré de communication. Puis au 19^e siècle c'est le besoin de standardiser objets et mesures qui semble avoir été à l'ordre du jour. Le 20^e siècle, lui, a été marqué par le besoin de standardiser les termes.

C'est Eugen Wüster qui a, le premier, mis en place une théorie de la terminologie. C'est un scientifique qui s'est passionné pour les problèmes de langue qui en découlent. Il s'est inspiré de la linguistique de Saussure et de celle de l'École de Prague qu'il a d'abord appliquées à son domaine particulier. Pour lui la terminologie est à la frontière entre la langue et la matière étudiée. Il s'agit donc d'une matière multidisciplinaire. Ses composantes sont les suivantes : (1) la linguistique (sémantique, lexicologie et lexicographie) ; (2) la logique pour ce qui est du problème de la classification des connaissances ; (3) les principes d'information et de documentation (thésaurus) ; (4) la standardisation ; (5) la discipline concernée. De ses idées est née d'abord l'École de Vienne puis celle de Moscou. L'École de Vienne est devenue en 1971 INFOTERM issu lui-même du programme UNISIST de l'UNESCO. L'œuvre de Wüster a été continuée par Felber (1984) à INFOTERM, par H. Picht et J. Draskau à DANTERM (1985), et Sager *et al.* à UMIST (1981, 1983, 1986).

2.3 LA DÉLIMITATION DES CONCEPTS BASÉE SUR LES THÉSAURUS

Le point de départ étant le concept, comment le délimiter ? Il faut d'abord distinguer le *concept individuel* représenté par un nom tel que « Pierre Durand », du *concept général* tel que « homme », celui-ci étant composé d'un certain nombre de caractéristiques.

On peut associer à la notion de concept celle de *caractéristiques*. Elles sont de deux sortes : les caractéristiques intrinsèques (forme, couleur, matière) et les caractéristiques extrinsèques (origine : l'inventeur, comme Diesel, par exemple ; ou but : la télévision).

Certaines de ces caractéristiques peuvent être pertinentes dans certains contextes et pas dans d'autres : la couleur d'une rose peut être intéressante pour un jardinier mais pas pour un botaniste par exemple. D'une manière générale, ces caractéristiques sont souvent plus utiles dans le domaine des sciences pures que dans le domaine des « sciences douces ». Il en est de même pour les deux autres concepts d'extension et d'intension que nous allons maintenant définir.

L'intension du terme est la *somme de toutes les caractéristiques qui constituent le concept* (définition donnée par ISO/R1087, 1969). Ainsi un véhicule motorisé est un véhicule qui a un moteur et un mécanisme de guidage ; un camion possède les mêmes caractéristiques plus la caractéristique « pour le transport », et « camion-citerne » a de nouveau les mêmes caractéristiques plus la caractéristique « liquide, gaz ou poudre » (ciment, farine, produit chimique).

L'extension du terme est la *somme de tous les objets individuels recouverts par le concept*. Ainsi le terme « pays nordiques » recouvre le Danemark, la Norvège, la Suède, l'Islande et la Finlande. Le terme extension peut également s'appliquer aux parties formant un tout. Il va de soi que plus l'extension du terme est grande, moins il a de caractéristiques.

Mais comment appliquer ces principes à une banque de données terminologiques ? Pour définir la valeur d'un terme, il est souvent utile d'établir quels sont les ter-

mes qui sont synonymes (voir 3.2.1) et quels sont ceux qui ont un sens plus large ou moins large, soit dans une hiérarchie distinguant le générique du spécifique, soit dans une hiérarchie distinguant la partie du tout. Par exemple, le terme générique *bétail* inclut le terme spécifique *bœuf* tandis que *la Bretagne* fait partie de *la France* (voir 3.2.3, « liens conceptuels »).

Une autre caractéristique du terme peut être ses *affinités associatives* avec un autre terme. Par exemple, le terme *oiseau* fait penser au terme *ornithologie* et le terme *insecte* peut également faire penser à *insecticide*. Parmi ces relations on compte l'antonymie (par ex. *dur/mou*), les liens génétiques (*père/fils*), les relations objet-but (*voiture/transport*), les relations cause-effet (*enseigner/apprendre*). La liste de ces relations est évidemment vague, mais elles sont néanmoins importantes lorsque l'on veut situer le terme dans le système. En virologie, deux virus peuvent avoir des affinités associatives s'ils partagent un certain nombre de caractéristiques mais occupent des « hôtes » différents (voir 3.2.3 b). Dans les systèmes préconisés par ISO (ISO 5964, 1985), ces termes sont précédés de VA, c'est-à-dire *Voir aussi* en français et de *Related term* en anglais. « Terme associé » semble plus transparent, ou du moins plus facile à employer dans le contexte du discours, d'où son emploi ici.

Ces distinctions sont fort importantes et il se trouve que les concepts de terme générique et de spécifique sont spécialement commodes en virologie, étant donné sa structure strictement hiérarchique (voir 3.2.3 a) ; c'est le cas d'ailleurs pour tous les domaines taxonomiques. De même les concepts de « terme associé » (*related term*) ou de « liens par affinité » sont essentiels (voir 3.2.3 b). Mais d'une manière générale ces relations sont souvent difficiles à établir. Ce qui n'empêche que de nombreuses compagnies cherchent à l'heure actuelle à établir des thésaurus des termes qu'elles emploient (par exemple le thésaurus norvégien/anglais qui est en train d'être rédigé par l'industrie du pétrole et du gaz de la mer du Nord). Un des exemples d'entrée donnés dans ISO est comme suit :

AÉROSTAT
TG AÉRONEF
TS DIRIGEABLE

(TG = terme générique ; TS = terme spécifique) (page 38)

3. ÉTABLISSEMENT D'UNE FICHE TERMINOLOGIQUE EN VIROLOGIE

3.1 LES COMPOSANTES DE LA FICHE DE TERMINOLOGIE

Elle se compose de quatre genres différents de cases : (1) les cases de données générales (nom du programmeur, de la date d'entrée du terme, etc.) mais il s'agit là de la gestion de la banque ; or celle-ci ne sera pas examinée dans notre contexte ; (2) les cases de données relatives au concept qui servent à lui donner un sens précis (synonymes, définition, liens conceptuels avec d'autres termes et antonymes s'il y a lieu), voir 3.2 ; (3) les données supplémentaires relatives au concept servant à élargir le sujet (exemples d'emploi du terme et sources bibliographiques) voir 3.3 ; (4) des renseignements linguistiques sur le terme en question, voir 3.4.

Il semble que le mieux soit de représenter dès le départ le champ conceptuel de la matière envisagée. Pour la virologie, voir schéma 1. Il est important que ce schéma figure dans la banque de terminologie.

Ayant donné cette vue d'ensemble du sujet, nous allons passer à l'étude de la terminologie de la virologie, et surtout des virus. Un aspect important de celle-ci est de situer le virus dans le cadre de la virologie.

3.2 LES CASES DÉFINISSANT LE CONCEPT

3.2.1 Les synonymes

Sous la rubrique « synonymes » sont inclus :

a) *Les synonymes absolus*

Dans ce cas il s'agit de termes représentant le même concept dans des registres différents ; ces synonymes peuvent être des variantes en latin, en langue populaire ou, dans le cas de la virologie, un équivalent médical ; pour le *virus de la fièvre aphteuse*, par exemple, il faudrait inclure *aphthovirus* qui est le nom latin donné à ce virus puisque celui-ci est employé au niveau international. Cette case peut également inclure d'autres créations terminologiques recouvrant le même concept dans le cas où la normalisation des termes n'a pas eu lieu ou est mal acceptée. Ainsi « *knowledge engineer* » peut être traduit soit par *ingénieur de la connaissance* ou *ingénieur de la cognoscence*.

b) *Les abréviations*

Celles-ci peuvent être des *sigles* ou des *acronymes* (initiales qui forment un nouveau mot : *laser* et *sida* en sont des exemples), des *initialismes* (dans ce cas le concept d'initiales est retenu, comme dans *FMD* pour « *foot-and-mouth disease* » et les mots à *apocopes* (comme dans *labo.*), mais ceux-ci n'existent pas dans tous les domaines.

c) *Les termes déconseillés*

Les termes déconseillés (*deprecated terms*) sont soit des termes démodés, soit des termes familiers ou bien des termes spécifiques à une entreprise. Il s'agit alors de jargon. Le terme *cocotte* correspond à « *fièvre aphteuse* » mais c'est là un terme familier et démodé.

3.2.2 La définition

Le contenu de la définition est déterminé selon le domaine traité et comprendra des données différentes selon les besoins des experts. En virologie, elle comprend (1) les caractéristiques intrinsèques (voir 2.3) comme le poids moléculaire, (2) les caractéristiques extrinsèques comme les signes et symptômes, le traitement et l'épidémiologie. C'est-à-dire que la « définition » comprend à la fois la définition du terme et la description du virus. Elle comprend donc tout ce que le spécialiste veut y trouver, c'est-à-dire tout ce qui a trait directement au virus tandis que les autres cases se réfèrent à des faits ou idées associés (voir schéma 2 pour un exemple de définition). En principe, ce système devrait avoir l'avantage de permettre le regroupement des connaissances ; si, par exemple, un chercheur voulait connaître tous les virus qui attaquent le bétail en Afrique, il devrait être possible de taper les mots « Afrique » et « bétail » pour que tous les virus ayant « Afrique » et « bétail » dans leur définition apparaissent sur l'écran. Mais la manière de procéder à ce genre d'opération appartient au domaine des techniques de l'intelligence artificielle.

3.2.3 Les liens conceptuels

(a) *Les liens hiérarchiques*

La classification se fait en familles ou groupes de virus, sous-familles, genres et espèces (voir schéma 1). La case se référant aux liens conceptuels contient donc pour la virologie ce qui, pour d'autres sujets, serait classé comme des hyperonymes (termes plus larges) et des hyponymes (termes plus étroits). Ainsi pour le virus de la rougeole (*measles*) peut-on lire :

family (famille) = *Paramyxoviridae* ; *genus* (genre) = *morbillivirus*.

Si un chercheur voulait savoir quels sont les virus appartenant au genre *morbillivirus*, il lui suffirait de taper *Morbillivirus* pour qu'il puisse les retrouver.

D'une façon plus générale, le terme peut avoir des liens conceptuels précis avec d'autres termes dans un système hiérarchique. Pour la virologie, par exemple, la hiérarchie du domaine est bien définie (voir ci-dessus). Dans certains cas, cependant, ce genre d'analyse n'a pas de sens ; par exemple, l'eau se compose d'oxygène et d'hydrogène, mais il serait ridicule de parler de l'hydrogène comme étant un hyponyme ou un terme plus étroit que l'eau.

(b) *Les liens par affinité : les termes associés (related terms)*

Ce terme s'applique en virologie à des virus semblables mais occupant un « hôte » différent. Ainsi lit-on, pour le virus de la rougeole :

full synonyms : rubeola virus

contextual synonyms : distemper virus (affects dogs), rinderpest virus (affects cattle), Kata virus (affects sheep and goats).

Un terme classé « terme associé » partage donc un certain nombre de caractéristiques-clés avec le terme décrit. Ce qu'il faut pour qu'un terme soit classé « terme associé » varie selon le domaine ; la décision appartient aux spécialistes et non aux linguistes.

Dans le cas où un même mot peut avoir une signification différente dans deux domaines très rapprochés, cette information est également incluse dans ces cases. En anglais, par exemple, les médecins se servent et du terme *mucosae* et du terme *mucous membrane* de manière interchangeable. Cependant il semble que certains médecins préfèrent employer *mucous membranes* pour les organes externes et *mucosae* pour les organes internes.

(c) *Quasi-synonymes*

Le système de classification des virus a l'avantage de rendre inutile le concept des « quasi-synonymes » qui présente toujours des problèmes au niveau de l'interprétation. Ce concept est couvert à la fois par « terme associé », terme qui peut être défini de manière précise comme en virologie, par exemple, et par les termes à sens soit plus large, soit plus étroit.

(d) *Antonymes*

De nombreux traducteurs désirent voir l'inclusion d'antonymes, lorsque cela est possible. La raison en est qu'il est parfois plus élégant de traduire un terme par un antonyme accompagné d'une négation. De plus, il est possible que face au mot *polluant*, un traducteur préfère *environmentally unacceptable* en anglais, plutôt que *polluting*, même s'il ne s'agit pas de termes absolument équivalents.

3.2.4 Commentaires sur les cases relatives au concept

Nous voyons donc que les cases ont été définies de manière à être adaptées à la description des virus et que ces mêmes cases seraient interprétées différemment pour d'autres matières. Le principe essentiel, c'est qu'il faut d'abord délimiter le sujet de manière conceptuelle, et que le reste suit.

3.3 LES CASES ÉLARGISSANT LE DOMAINE

3.3.1 La case « exemples d'emplois »

Elle contient à la fois des indications quant aux contextes dans lesquels le terme est employé, comme par exemple « médecine humaine et vétérinaire », et des exemples

de l'emploi du terme dans des phrases. Peuvent être ajoutés à ces exemples des commentaires quant au registre correspondant à l'emploi cité.

3.3.2 La case « références »

Elle comprend des références bibliographiques telles que les revues, les articles et travaux divers qui sont intéressantes à la fois pour le spécialiste à cause des informations qu'elles contiennent et pour le traducteur parce qu'elles montrent le terme employé dans différents contextes. Exemple, toujours pour le virus de la rougeole :

Références : Morgan, E.M. and Rapp, F. (1977), *Bact. Rev.* 41,636

3.4 RENSEIGNEMENTS LINGUISTIQUES

(a) Les renseignements grammaticaux

La séquence des concepts exprimés par des termes peut être importante lorsqu'il s'agit de termes formant des mots composés. Dans le terme « *combine harvester* », le premier concept est celui de « *harvest* » (moisson), le deuxième est celui de « *-er* » qui indique l'agent et finalement « *combine* » qui s'ajoute aux deux autres dans un ordre déterminé (exemple emprunté à H. Picht et J. Draskau 1985). Ceci est spécialement important lorsque l'on passe du terme allemand aux termes français ou anglais, vu que le terme recouvrant un concept composé en allemand sera écrit en un mot alors qu'il sera écrit en mots séparés dans les deux autres langues mentionnées. ISO 5964 (1985) donne l'exemple suivant : le terme *Lehrerbildungsgesetz* peut être traduit par *Loi sur la formation des enseignants* ; ceci n'est pas cependant un terme satisfaisant pour la recherche documentaire. Pour rendre cette dernière possible, il faut que le terme allemand soit analysé en *Bildung*, *Gesetz* et *Lehrer* dans l'index. Ceci permet que des recherches soient effectuées sur les mots *Gesetz* — ce qui donne *Loi*, *Lehrer* — ce qui donne *enseignant* et *Bildung* — ce qui donne *Formation*. Il s'agit là de termes « post-coordonnés » en ce sens que la mise en facteur s'est faite dans le système et non avant ; ce genre d'analyse ne s'applique pas aux termes qui ont été « précoordonnés », c'est-à-dire là où la combinaison d'éléments n'entre pas en ligne de compte, appartenant généralement au domaine de l'étymologie, comme dans *trade wind* qui n'a plus rien à voir avec le commerce et qui se traduit par *vents alizés*.

Pour ce qui est du nombre, ISO recommande l'emploi du pluriel pour le terme générique, et du singulier pour ses propriétés et pour les procédés qui peuvent lui être associés ; en virologie, toutefois, la norme dans les dictionnaires est de donner le singulier ; c'est ce qu'il nous a semblé préférable de faire, le pluriel étant indiqué là où le sens est changé par le nombre.

Lorsqu'on travaille dans un domaine taxonomique tel que celui de la virologie, tous les termes ou presque sont nominalisés. Les seuls renseignements grammaticaux qu'il convient d'inclure sont ceux qui se réfèrent au nombre pour les termes anglais, et au genre et au nombre pour d'autres langues telles que le français. Dans d'autres domaines, par contre, la situation est plus complexe à cause de l'inclusion des autres parties du discours. La syntaxe des mots composés présente des problèmes particulièrement ardu dans de nombreux domaines, notamment celui de l'ingénierie, à cause, en particulier, de l'omission des prépositions, qui devient de plus en plus répandue dans les textes techniques. La valence des verbes est également le genre de renseignement qui devrait être inclus.

(b) Les renseignements phonétiques

Ceux-ci sont utiles en particulier pour les spécialistes qui assistent à des conférences. Les dictionnaires donnent des transcriptions phonétiques « standard » ou « neutralisées », mais il serait peut-être intéressant de donner des alternatives, comme certaines

prononciations américaines par exemple, lorsque celles-ci varient au point de nuire à la compréhension du discours pour l'étranger n'ayant pas une maîtrise parfaite de la langue d'arrivée.

3.5 LE PROBLÈME DU CLASSEMENT

Il reste le problème du classement qui, lui, demeure le même quel que soit le domaine traité ; le système de classement universel (*Universal Decimal Classification*), qui est employé entre autres par la *British Standards Institution*, sert de base pour les thésaurus qui correspondent au principe des termes-clés. Il s'agit bien d'un système de classification conceptuel, mais il comprend trop de sous-divisions pour être employé facilement dans ce contexte. EURODICAUTOM a mis en place son propre système, d'après Lenocho (1977), mais il nous est difficile d'en évaluer l'efficacité vu que la case correspondant à la classification est souvent restée vide lorsque nous avons voulu nous en servir. Pour le moment dans notre banque terminologique, nous avons inclus le système de classification décimale universelle, mais nous cherchons toujours un système qui soit plus facile à manipuler.

Face à ce manque de système satisfaisant, les pays nordiques se sont mis au travail pour en élaborer un nouveau adapté aux besoins spécifiques des banques terminologiques. Il s'agit d'un système comprenant une majuscule suivie de quatre chiffres pour subdiviser chaque domaine ; mais l'introduction d'un nouveau système pose des problèmes à l'égard des systèmes existants. Il faudra attendre par conséquent un certain nombre d'années avant que ce système soit au point.

3.6 LE PROBLÈME DE LA TRADUCTION DES TERMES

Une possibilité serait de reproduire la fiche toute entière pour chaque langue ; c'est-à-dire que si l'on cherchait *rougeole* par exemple, et que l'on trouvait *measles*, on pourrait juger du degré d'équivalence entre ces deux termes en comparant les deux fiches. Mais ceci semble un procédé bien long vu que dans de nombreux cas, et en particulier dans les domaines où la terminologie a été non seulement normalisée mais internationalisée, il peut y avoir équivalence satisfaisante, même si elle n'est pas absolue. Que signifie « équivalence satisfaisante » ? Prenons le cas du mot *laboratoire* ; « *laboratory* » en anglais peut signifier non seulement le lieu de travail mais aussi une équipe de chercheurs, tandis qu'en français « *laboratoire* » signifie l'endroit seulement. Il faut bien sûr trouver le meilleur moyen de spécifier ces différences sémantiques — (souvent simples d'ailleurs comme dans cet exemple). Mais au niveau de la traduction c'est le contexte qui résout le problème : dans une phrase comme *His laboratory produced some outstanding work*, le verbe *produced*, réclamant un sujet qui soit animé, *laboratoire* ne peut être pris dans son sens de *lieu de travail*. Il semble que dans ce genre de cas une « note explicative » (ou *scope note*) serait suffisante pour mettre en garde le traducteur. Nous suggérons donc que la case « traduction » contienne simplement le terme équivalent, s'il existe, plus ses synonymes (synonymes absolus, acronymes, termes déconseillés) (voir 3.2.1) ; si le terme dans la langue d'arrivée n'est qu'un équivalent partiel, il est nécessaire d'indiquer tout de suite, peut-être par des initiales telles que TG ou TS qu'il s'agit d'un terme ayant un sens plus large ou plus étroit que le terme traduit. Dans ce cas il faut se reporter à une deuxième case contenant une « note explicative ». On aurait donc :

LABORATORY : la fiche suivrait, dûment remplie

TERME FRANÇAIS : LABORATOIRE (TS c'est-à-dire terme plus spécifique)

NOTE EXPLICATIVE : ne se réfère en français qu'au lieu de travail, tandis qu'en anglais ce terme peut se référer également à une équipe de chercheurs.

Il s'agit là, bien sûr, d'un exemple fort simple, mais cette partie de la fiche est encore au stade expérimental. Il va de soi qu'une case donnant les renseignements linguistiques est également incluse pour la langue étrangère.

4. CONCLUSIONS

Si l'on considère le format de la fiche remplie (voir schéma 3), celui-ci peut sembler compliqué. Mais ceci est uniquement au niveau de sa conception ; pour l'utilisateur, la banque de données terminologiques est très « conviviale » car il ne voit sur son écran que ce qu'il a envie d'y voir, un menu permettant à chacun de ne subvenir qu'à ses propres besoins.

De nos jours, les spécialistes et les traducteurs ont besoin d'avoir accès à une terminologie définie et à jour. Dans le passé les spécialistes l'obtenaient en consultant des livres, dont la matière est organisée de manière conceptuelle. Les traducteurs, eux, consultaient des dictionnaires techniques et aussi quelquefois des livres sur le sujet donné. Il s'agissait dans les deux cas d'un travail long et artisanal. Ce système ne résolvait pas d'ailleurs un certain nombre de problèmes, comme celui de la standardisation pour le réviseur, et comme le fait que chaque ouvrage de référence est suranné dès sa parution, ni d'autres problèmes comme celui résultant de l'absence d'un équivalent exact d'un terme dans une autre langue.

Les banques de données terminologiques peuvent aider à résoudre tous ces problèmes, et ceci pour de nombreuses raisons : (1) elles peuvent être conceptuellement organisées comme des livres ; (2) leur longueur est illimitée (on n'a plus le problème du nombre de pages) ; (3) le nombre des auteurs est également illimité puisqu'il n'y a pas de date de publication ; (4) elles peuvent être mises à jour à peu de frais ; (5) leur accès est facile puisqu'on peut en retirer uniquement ce que l'on veut ; (6) la description de chaque terme est complète et séparée pour chaque langue.

Mais l'utilité d'une banque dépend d'une part de son formatage (la fiche doit être complète du point de vue du spécialiste, et les cases ne doivent pas être ambiguës) et d'autre part de son degré d'harmonisation ou de compatibilité avec d'autres banques, puisqu'aucune banque ne saurait couvrir tous les domaines qui existent ; son succès dépend également de la mesure dans laquelle elle est sentie par l'utilisateur comme étant « amicale » ou « conviviale », et ceci dépend du succès de la coopération « informaticien/terminologue ».

Maintenant c'est dans ce domaine que nous concentrons nos efforts. Notre banque terminologique est à présent sur ORACLE, parce qu'il s'agit là d'un système commode, connu et adaptable puisqu'il peut être employé sur différents ordinateurs. Étant un SGBD (système de gestion des bases de données) relationnel, il nous est possible dès à présent de demander à l'ordinateur de mettre sur l'écran les cases voulues. Mais ce n'est pas suffisant. Nous voudrions que l'utilisateur puisse questionner directement la machine comme il l'entend. En virologie, par exemple, un spécialiste pourrait vouloir savoir les noms de tous les pays dans lesquels on trouve un certain virus. Cette information étant incluse dans la définition et ne faisant pas l'objet d'une case séparée, il nous est en ce moment impossible d'interroger directement la machine sur ce point précis.

Ceci étant regrettable, c'est là le problème que nous cherchons à résoudre. De plus nous voudrions que l'utilisateur puisse questionner la machine en langage naturel. C'est dans ce contexte que nous avons établi une liste du genre de questions qu'un spécialiste pourrait vouloir poser en virologie. Il s'agit maintenant de traduire ces questions de ce genre en SQL, la langue d'ORACLE. Ainsi, ayant établi le « format » de la fiche terminologique du point de vue du spécialiste et du linguiste, c'est aux informaticiens, et plus

spécialement aux spécialistes en systèmes experts et en intelligence artificielle à prendre le relais.

RÉFÉRENCES

- BENNETT, P.A., R.L. JOHNSON, J. McNAUGHT, J.M. PUGH, J.C. SAGER, H.L. SOMERS (1986) : *Multilingual Aspects of Information Technology*, London, Gower Press, pp. 53-63. (Référence indirecte.)
- FELBER, H. (1984) : *Terminology Manual*. INFOTERM, Paris, UNESCO (UNISIST).
- HVALKOF, S. (1985) : *Étude comparative des données terminologiques des banques de terminologie DANTERM, B.T.Q., EURODICAUTOM, NORMATERM, O.L.F. et SIEMENS*, Kbenhavn, Fagsprogligt Center, Handelshojskolen.
- ISO/R1087 (1969) : *Vocabulaire de terminologie*.
- ISO 2788 (1974) : *Principes directeurs pour l'étude et le développement de thesaurus monolingues*. (Référence indirecte.)
- ISO 5964 (1985) : *Principes directeurs pour l'étude et le développement de thesaurus multilingues*.
- LENOCH, H. (1977) : *System of Classification*, Luxembourg, E.E.C.
- PICHT, H. and J. DRASKAU (1985) : *Terminology : An Introduction*, Guildford, University of Surrey, Department of Linguistic and International Studies.
- SAGER, J.C. and R.C. CANDELAND (1986) : *The British Term Bank Project*, Report n° 86/1, Manchester, CCL/UMIST.
- SAGER, J.C. and J. McNAUGHT (1981) : *Feasibility Study of the Establishment of a Terminological Data Bank in the U.K.* British Library R. & D. Report n° 5642.
- SAGER, J.C. and L.E. PRICE (1983) : *The British Term Bank Project*, Report n° 83/14, Manchester, CCL/UMIST.
- THOMAS, P., K. AHMAD, A. JUDGE, S. BARBOUR and N. REEVES (1986) : « The British Term Bank », in *Terminologie et traduction*, 3, Luxembourg, E.E.C., pp. 63-100.

SCHEMA 3. Format de la fiche de la banque de données terminologiques de l'Université du Surrey, Royaume-uni.

- | | |
|----|--|
| | <i>Champs ou « cases »</i> |
| 0 | N° du terme |
| 1 | Lieu d'origine du terme |
| 2 | « Pool » (sous-partie du domaine) |
| 3 | Rédacteur |
| 4 | Date d'entrée du terme |
| 5 | Date de mise à jour |
| 6 | Nom du responsable pour la mise à jour |
| 7 | Source de mise à jour |
| 8 | Texte de mise à jour |
| 9 | Domaine |
| 10 | Code de la langue et du pays (deux lettres chacune selon l'ISO) |
| 11 | Terme-entrée |
| 12 | Source (3 lettres e.g. ISO) |
| 13 | Source-type (3 lettres e.g. DIC = dictionnaire) |
| 14 | Source n° : (dans dossier séparé) |
| 15 | Page dans la source |
| 16 | Synonymes (équivalent latin ; langage familier) |
| 17 | Abréviations (initialismes = SNCF ; acronymes = LASER ; apocopes = labo.) |
| 18 | Synonymes contextuels (dans le cas de la virologie, il s'agit du virus associé) |
| 19 | Terme(s) déconseillé(s) (archaïque, trompeur, jargon) |
| 20 | Définition (celle-ci comprend en particulier les caractéristiques du terme) |
| 21 | Références bibliographiques |
| 22 | Exemples d'emplois |
| 23 | Liens conceptuels (positions du terme dans le système des concepts et affinités associatives là où ces principes sont applicables) |
| 24 | Antonyme(s) (case qui reste le plus souvent vide) |
| 25 | Note grammaticale |
| 26 | |
| 27 | Classification |
| 28 | Copyright |
| 29 | Forme(s) phonétique(s) |
| 30 | Deuxième langue et pays |

- 31 Terme équivalent
- 32 Synonyme(s)
- 33 Terme(s) déconseillé(s)
- 34 Note explicative (sur le degré d'équivalence entre le terme français et anglais)
- 35 Référence bibliographique
- 40-45 Troisième langue, etc.