



Rint
Réseau
international
de néologie
et de terminologie

Banques de terminologie

Actes de la table ronde
(Québec, 18 et 19 janvier 1996)

15

Afrique centrale
et de l'Est
Afrique de l'Ouest
Canada
Communauté
française de
Belgique
France
Haïti
Madagascar
Maroc
Québec
République
centrafricaine
Suisse
Tunisie
Union latine



Revue semestrielle
coéditée par l'Agence
de la francophonie
et la Communauté
française de Belgique.

N° 15
juin et décembre 1996.

Terminologies nouvelles

Allocutions d'ouverture

Allocution de Madame Nicole René,
Présidente de l'Office de la langue
française.
Page 4

Allocution de Madame Diana Monet,
Présidente-directrice générale du
Bureau de la traduction.
Page 6

Allocution de Monsieur
Louis-Jean Rousseau,
Secrétaire général du Rint.
Page 7

Thème 1: La gestion des banques de terminologie

A - Les problèmes juridiques

Communications

*La terminotique sans contrefaçon est-elle
possible?*
par Gérard Snow,
Université de Moncton, Canada.
Page 10

*Le cadre contractuel des banques de
terminologie,*
par Isabelle de Lamberterie,
CNRS, France.
Page 14

Comptes rendus des débats Page 18

B - L'organisation et le fonctionnement des banques de terminologie

Communications

Termdat, un partenariat réussi,
par Françoise Parc,
Chancellerie fédérale suisse, Suisse.
Page 21

Ivanhoé,
par Roger Racine,
Bureau de la traduction, Canada.
Page 24

*La configuration des outils de gestion
de Termium,*
par Roger Racine,
Bureau de la traduction, Canada.
Page 26

Comptes rendus des débats Page 30

Thème 2: Les outils terminotiques et le poste de travail du terminologue

Communications

*Adepto-Nomino: un outil de veille
terminologique,*
par Jean Perron,
Office de la langue française, Québec.
Page 32

*Filtact®: un automate d'extraction
des termes complexes,*
par Pierre Auger,
Université Laval, Québec,
Patrick Drouin,
Université de Montréal, Québec,
et Alain Auger,
Université Laval, Université
de Neuchâtel, Suisse.
Page 48

*Termplus, système d'extraction
terminologique,*
par Jacques Ladouceur
et Guylaine Cochrane,
Université Laval, Québec.
Page 52

Le poste de travail du terminologue arabe
par Saadia Ait Taleb
Page 57

Sommaire

Le système SAMI-BTQ: un système intégré de production et de mise à jour des données de la Banque de terminologie du Québec,
par Sylvie Pelletier,
Office de la langue française, Québec.
Page 66

L'ATelier du TERminologue (Latter®),
par Louise Claude,
Bureau de la traduction, Canada.
Page 77

Comptes rendus des débats
Page 83

Thème 3: Le développement et la diffusion des banques de terminologie

Communications

La banque de données terminologiques TermRom, petite expérience d'une petite banque de terminologie,
par Mihaela Ursu,
TermRom, Roumanie.
Page 86

Étude comparée de dictionnaires électroniques et de banques de termes,
par Élisabeth Hombrouck
Institut Marie-Haps, Belgique.
Page 90

Les banques de terminologie au seuil d'une nouvelle dynamique,
par Mireille Lacasse,
Office de la langue française, Québec.
Page 107

Des banques de données terminologiques en Afrique francophone,
par Marcel Diki-Kidiri,
CNRS, France.
Page 116

Une base de données multimédia: le dictionnaire descriptif et visuel de Parcs Canada,
par Raymond Pepermans,
Université d'Ottawa, Canada.
Page 121

Les banques de terminologie de services linguistiques,
par Marielle Hébert,
Services linguistiques Hébert, Québec.
Page 125

L'inventaire des ressources terminologiques du Rint,
par Louis-Jean Rousseau,
Secrétaire général du Rint.
Page 128

L'inventaire français des banques de terminologie effectué en 1995 dans le cadre du Réseau international de néologie et de terminologie (Rint),
par Loïc Depecker,
Délégation générale
à la langue française
et Christophe Riondet,
Université de Paris IV, France.
Page 132

L'utilisation d'Internet dans la communication des produits de l'industrie de la langue,
par Nathalie Côté,
Les Logiciels Machina Sapiens,
Québec.
Page 135

Comptes rendus des débats
Page 140

Thème 4: Prospectives et avenues de collaboration

Communications

La Banque internationale d'information sur les États francophones,
par Lucie Lépine,
Banque internationale d'information sur les États francophones (BIEF).
Page 144

Les avenues de collaboration entre les banques de terminologie,
par Louis-Claude Tremblay,
Bureau de la traduction, Canada.
Page 147

Vers la fiche virtuelle en terminologie,
par Jean-Guy Meunier,
Université du Québec à Montréal,
Québec.
Page 151

Évolution des banques de données terminologiques,
par Daniel Gouadec,
Université de Rennes 2, France.
Page 154

Comptes rendus des débats
Page 160

Clôture de la table ronde,
par Louis-Jean Rousseau,
Secrétaire général du Rint.
Page 162

Bibliographie de la néologie,
Rubrique préparée par Chantal Girardin et John Humbley.
Page 164

Normalisation
Par Élisabeth Blanchon.
Page 169

En bref
Publications
Page 172

Actes de la table ronde sur les banques de terminologie

tenue à Québec les 18 et 19 janvier 1996

Textes rassemblés par
Louis-Jean Rousseau
Secrétaire général du Rint

Révision: Micheline Lapointe-Giguère

Allocution de Madame la Présidente de l'Office de la langue française du Québec

Madame la Présidente-directrice
générale du Bureau de la traduction,
Monsieur le Secrétaire général
du Rint,
Mesdames, Messieurs,

C'est avec grand plaisir que je vous souhaite la bienvenue à cette table

ronde organisée par le Réseau international de néologie et de terminologie et l'Office de la langue française, avec la collaboration du Bureau de la traduction et de l'Agence de coopération culturelle et technique.

En tant que membre fondateur du Rint, l'Office de la langue française est heureux d'accueillir à Québec pendant toute cette semaine les autres modules du réseau. Au cours de la table ronde qui s'ouvre aujourd'hui, vous allez traiter d'un sujet plus que jamais d'actualité: les banques de terminologie.

Puisque la Banque de terminologie du Québec a 23 ans cette année, l'Office fait sans doute figure de pionnier en ce domaine, et je peux affirmer que le BTQ reste, en 1996, au centre des activités terminologiques de notre organisme. C'est pour nous un formidable outil de recherche, de production, de gestion, de diffusion et d'implantation de la terminologie, tout à la fois. Dans l'aménagement terminologique qui fait partie intégrante de la raison d'être de l'Office, tout comme dans les activités langagières qui nous occupent, nous concevons notre banque de

terminologie comme la première de nos ressources, et nous avons à cœur de l'exploiter et de la développer au maximum.

C'est pourquoi la qualité et l'actualisation des données terminologiques doivent aller de pair avec la diffusion toujours plus large et toujours plus innovante de ces données. Il faut que cette double mission se réalise aussi bien dans les milieux de travail québécois, pour l'implantation réelle et durable du français, que chez les partenaires francophones, pour le partage du français, langue de la modernité. Aujourd'hui, ces deux objectifs semblent de plus en plus accessibles grâce aux progrès des industries de la langue et des technologies de l'information, dont il faut se faire des alliés. En effet, de même que les supports électroniques permettent une diffusion sans entrave, de même les inforoutes s'offrent comme une voie tout indiquée pour affirmer la présence de la langue française et rendre les contenus francophones disponibles universellement. Les banques de terminologie y ont naturellement leur place, vous le savez bien.

J'ajouterais en terminant que la mondialisation des échanges de toute nature doit comprendre celle des échanges et de la coopération entre les banques de terminologie. Pour l'essor de la langue française, il importe que

Allocutions d'ouverture

le plus grand nombre de francophones aient accès à leur patrimoine terminologique, selon les modes qui répondent le mieux à leurs besoins. C'est aussi une question à laquelle le Rint est sensible, et la table ronde qui commence ce matin en témoigne.

À tous et à toutes, je souhaite donc des communications éclairantes et des échanges de vues constructifs.

*Nicole René,
Présidente de l'Office de la langue
française du Québec.*

Allocution de Madame la Présidente-directrice générale du Bureau de la traduction du gouvernement du Canada

Madame la Présidente de l'Office
de la langue française,
Monsieur le Secrétaire général
du Réseau international de néologie
et de terminologie,
Mesdames, Messieurs,
Chers collègues,

J'aimerais remercier
l'Office de la langue
française du Québec et
le module québécois du
Réseau international de
néologie et de
terminologie de nous
accueillir à la table

ronde sur les banques de
terminologie.

Cette manifestation, qui se
déroule sous les auspices de l'Agence
de coopération culturelle et technique
au cœur même de la plus vieille cité
française d'Amérique, revêt un intérêt
primordial pour le Bureau de la
traduction du gouvernement
canadien.

En effet, comme vous le savez, le
Bureau exploite Termium®, l'une des
plus grandes banques de données
linguistiques au monde, et, de ce fait,
est particulièrement préoccupé par
tous les thèmes qui seront abordés au
cours des deux prochaines journées,
soit la gestion des banques de
terminologie, les outils terminotiques,
le développement et la diffusion des
banques de terminologie ainsi que les
prospectives et les avenues de
collaboration.

À ce dernier chapitre et à titre
d'exemple, j'aimerais souligner les
excellents rapports qui existent depuis
une vingtaine d'années entre le
Bureau de la traduction et l'Office de
la langue française du Québec,
rapports qui se sont concrétisés par la
signature d'un protocole d'entente où
la consultation de nos banques de
terminologie respectives occupe une
place importante.

En terminant, j'aimerais
souhaiter à tous les participantes et à
tous les participants à la table ronde

sur les banques de terminologie de
fructueuses discussions qui
permettent de faire avancer la
réflexion sur la question.

Je formule ainsi le vœu qu'il se
dégage des pistes intéressantes pour
l'avenir des banques de terminologie
quant à leur contribution à la
communication internationale en
cette ère de mondialisation des
échanges où les marchés, en pleine
expansion, vont de plus en plus
intéresser ces banques à titre
d'entrepreneurs.

*Diana Monnet,
Présidente-directrice générale,
Bureau de la traduction,
Canada.*

Allocution de Monsieur le Secrétaire général du Rint

Madame la Présidente de l'Office
de la langue française,
Madame la Présidente-directrice
générale du Bureau de la traduction,
Mesdames, Messieurs,

Au nom du Comité d'orientation du Réseau international de néologie et de terminologie, permettez-moi de vous souhaiter la plus cordiale bienvenue à Québec. Je vous remercie d'être venus si nombreux à cette rencontre sur les banques de terminologie organisée par le Rint et l'Office de la langue française du Québec, en collaboration avec le Bureau de la traduction du Canada et l'Agence de coopération culturelle et technique.

Depuis sa création en 1986, le Rint, conformément aux objectifs qu'il s'est donné en matière de production et de diffusion de la terminologie, s'est efforcé de conduire des travaux de réflexion théorique et méthodologique afin de pouvoir définir des pratiques reposant sur l'application de principes de travail rigoureux et sur la mise en œuvre d'outils de travail permettant d'améliorer sans cesse la qualité des travaux terminologiques. Les journées d'étude que nous entreprenons aujourd'hui s'inscrivent dans cette continuité et s'intéressent cette fois à la collecte, le traitement et la diffusion des données terminologiques et, plus particulièrement, aux banques de terminologie.

On connaît le rôle qu'ont joué et que jouent encore les banques terminologie dans la diffusion du vocabulaire scientifique et technique. De plus, on assiste à la prolifération des banques de terminologie, grandes moyennes ou petites. Le répertoire des banques dont vous avez reçu un

exemplaire démontre, en effet, qu'à côté des grandes banques de terminologie qui touchent la plupart des domaines et qui contiennent des centaines de milliers de fiches, il existe un foisonnement de petites banques spécialisées, le plus souvent à diffusion restreinte ou privée.

Par ailleurs, des changements importants sont en train d'intervenir dans le marché terminologique et les innovations technologiques amènent des bouleversements dans les modes de production et de diffusion des données terminologiques.

Comment coordonner les travaux terminologiques dans un souci d'efficacité?

Comment accélérer la mise à jour et le développement du contenu des banques?

Avec quels outils?

Sous quels formats et sur quels supports faut-il diffuser la terminologie pour atteindre les clientèles cibles dont les besoins sont sans cesse plus nombreux et diversifiés?

Quels choix technologiques faut-il faire pour adapter les banques au marché actuel et futur à l'heure des autoroutes de l'information?

Faut-il fusionner les banques existantes?

Faut-il les mettre en réseau? Comment?

Pourquoi tant de banques demeurent-elles à diffusion restreinte, alors qu'il y a une clientèle croissante pour les données terminologiques?

Comment s'assurer que les usagers de la terminologie puissent avoir accès à l'ensemble des ressources disponibles?

Quels sont les obstacles techniques et juridiques à l'échange et à la diffusion des données?

Toutes ces questions que chacun d'entre nous se pose ne peuvent trouver des réponses satisfaisantes sans la mise en commun des problèmes, des expériences, des idées et des solutions. En l'absence d'un lieu pour débattre en commun de ces questions, nous avons pensé convoquer cette première réunion sur le thème des banques de terminologie en espérant réussir à trouver réponse à quelques-uns des problèmes que nous avons évoqués, en réunissant des chercheurs, des concepteurs et des producteurs d'outils informatisés, des exploitants de banques de terminologie, des consommateurs de données terminologiques et des gestionnaires de services linguistiques.

*Louis-Jean Rousseau,
Secrétaire général du Rint.*

La gestion des banques de terminologie

A — Les problèmes juridiques

Présidente de séance:
Tina Célestin
(Office de la langue française, Québec)

Thème 1

La terminotique sans contrefaçon est-elle possible?

Le droit d'auteur confère des droits exclusifs sur la publication, la reproduction et l'adaptation d'œuvres originales. Il n'existe à ce régime que quelques exceptions, par exemple lorsqu'il s'agit de courts extraits copiés pour des fins de recherche privée.

Le terminographe s'emploie notamment à attester les usages à l'aide de citations qu'il reproduit sur des fiches. Grâce à la micro-informatique, ces données sont facilement transmissibles dans leur entièreté. Y a-t-il, dès lors, atteinte au droit d'auteur?

Termes-clés :
droit d'auteur ; terminotique ; terminographie ; banques de terminologie.

(1) Plus exactement, la contrefaçon est l'atteinte aux droits patrimoniaux, qui découlent du monopole d'exploitation de l'auteur sur son œuvre, par opposition aux droits moraux, qui protègent l'intégrité et la paternité de l'œuvre. Il y a contrefaçon chaque fois qu'est reproduite sans permission ou justification la totalité ou une partie substantielle d'une œuvre protégée.

Il est clair que l'avènement de la micro-informatique a grandement facilité l'élaboration et la diffusion des fichiers terminologiques. Alors qu'autrefois, la diffusion passait presque nécessairement par la publication d'ouvrages lexicographiques, l'information devant au préalable être épurée, triée et condensée, il est aisé aujourd'hui de diffuser son fichier terminologique en entier, sous la forme d'une banque de données.

Le problème de l'infraction au droit d'auteur, ou contrefaçon⁽¹⁾, n'est certes pas nouveau en terminographie, car toute transcription non autorisée, même faite à la main, peut donner lieu à la contrefaçon (Tamaro 1994 : 92). Cependant, si la nouvelle technologie rend plus facile le travail du terminographe, elle amplifie en même temps les risques de contrefaçon.

Mon propos n'est pas de répondre aux nombreuses questions qui se posent à ce sujet, mais de formuler quelques hypothèses et proposer des moyens d'action.

Rappelons au départ que, même si le droit d'auteur varie, à certains égards, d'un pays à l'autre, la plupart des législations nationales, à la suite d'accords internationaux, partagent sensiblement les mêmes objectifs et dispositions fondamentales, en tout

cas pour ce qui a trait aux matières que je traiterai.

1 Incidences du droit d'auteur sur l'activité terminotique

Le travail du terminographe comporte différentes opérations, chacune soulevant des problèmes particuliers par rapport au droit d'auteur. En voici quelques exemples.

1.1 Adoption d'un arbre de domaine ou d'un thésaurus existants

La loi reconnaît un monopole d'exploitation à l'auteur de toute œuvre qui satisfait aux critères d'originalité et de fixation. Or, tant la loi elle-même que la jurisprudence ont donné à l'œuvre originale une définition extrêmement large, allant du poème au simple bottin (Nabhan 1992 : 17). Il est clair, à ce point de vue, qu'un arbre de domaine ou un thésaurus peuvent donner prise au droit d'auteur. Dans ce cas, le terminographe n'aura pas le droit de les reproduire sans permission, mais cela ne veut pas dire qu'il ne pourra pas s'en inspirer, bien au contraire. C'est, en effet, la reproduction de

l'expression, non l'utilisation des idées, que prohibe la loi.

1.2 Dépouillements

On ne peut pas faire de terminographie sans dépouiller des ouvrages. Les extraits qu'on consignera sur des fiches serviront à attester l'existence du terme et à illustrer son usage en contexte. En terminographie bilingue, on utilise à cette fin des ouvrages dans la langue source, dans la langue cible, ou dans les deux langues à la fois. Or, ces ouvrages de référence font généralement l'objet d'un droit d'auteur. Le terminographe est-il tenu, dans ce cas, de demander la permission?

Il paraît évident que dans la mesure où, pour attester un terme ou illustrer son emploi, le terminographe emprunte quelques lignes d'un ouvrage scientifique, journalistique ou littéraire, il ne fait guère concurrence au titulaire des droits sur l'ouvrage.

Est-il raisonnable, en effet, de penser que le consommateur intéressé par le contenu de tel ouvrage ou de tel article se contentera de consulter sa banque de terminologie plutôt que d'acheter l'ouvrage ou la revue? Puisque l'objet fondamental de la loi,

en matière de droits patrimoniaux, est d'encourager l'industrie du livre et du spectacle en réprimant la piraterie, ne serait-on pas fondé à conclure que le terminographe, ne nuisant en rien aux intérêts du titulaire du droit d'auteur, n'est pas coupable de contrefaçon?

Hélas! Ce n'est pas ainsi qu'a été interprétée la loi. Il faudrait, pour que le terminographe soit dispensé de l'obligation d'obtenir la permission de reproduire un passage⁽²⁾ d'une œuvre protégée, qu'il puisse invoquer la défense de l'utilisation équitable «pour des fins d'étude privée» ou «de recherche»⁽³⁾. Tant et aussi longtemps que le fichier terminologique reste à l'état de fichier de recherche, ou du moins qu'il n'est consulté qu'à l'interne, passe encore. Mais du moment que le fichier est diffusé (ce qui est le propre d'une banque de terminologie), il est douteux que le terminographe ait gain de cause en faisant valoir la défense de l'utilisation équitable.

On pourrait opposer un autre argument à l'obligation pour le terminographe d'obtenir la permission du titulaire du droit d'auteur, qui serait fondé sur la nature de l'objet emprunté. La loi ne reconnaît pas de droit d'auteur sur les idées ou les faits en tant que tels, mais sur l'expression. Du coup, ne pourrait-on pas soutenir que le terminographe qui relève un mot ou un passage agit, non pas pour s'exprimer (en empruntant, légitimement ou non, les mots d'une autre personne), mais dans le seul but de constater un fait de langue, de la même manière qu'un psychologue reproduirait, dans une étude scientifique, les premiers mots d'un enfant. Dans l'un et l'autre cas, il ne s'agirait que d'une simple observation, non pas d'une expression. Force est de reconnaître, toutefois, que ni la jurisprudence ni la doctrine ne semblent jusqu'ici avoir distingué entre l'expression et le fait de langue, s'agissant de citations.

1.3 Emprunts aux dictionnaires

Le terminographe est-il tenu, en droit, de demander la permission pour citer une définition d'un dictionnaire existant?

Que les dictionnaires puissent donner lieu au droit d'auteur ne fait aucun doute. Au regard de la loi sur le droit d'auteur, ce sont, en tant que compilations ou recueils, des œuvres littéraires protégeables... pour autant, naturellement, qu'ils satisfont au critère d'originalité.

Or, il est bien clair que les définitions des dictionnaires ne sont pas toutes ni entièrement originales; on n'a qu'à comparer les dictionnaires pour constater leurs ressemblances. Rien de plus normal, d'ailleurs, considérant qu'il n'y a pas toujours trente-six façons pour le lexicographe de décrire une chose.

Dans la mesure, donc, où le terminographe emprunte à un dictionnaire une définition dont la formulation n'est pas originale, il ne devrait pas être tenu d'obtenir la permission. Mais cette affirmation ne règle pas vraiment le problème en pratique, car il faudrait qu'il s'astreigne chaque fois à comparer plusieurs dictionnaires pour vérifier le degré d'originalité. Et même s'il constatait que la première définition n'est pas originale, la deuxième ou la troisième, que lui-même ou ses collègues voudront citer, le seront peut-être.

Il est donc plus pertinent de se demander s'il n'est pas de la nature même des dictionnaires de se faire citer. N'y aurait-il pas, à ce point de vue, une permission implicite, de la part de l'auteur du dictionnaire, qu'on le cite, pourvu évidemment qu'on indique sa source? Je n'ai pas la réponse à cette question, et je ne connais aucune cause à ce sujet, mais il serait éminemment utile, il me semble, qu'un organisme comme le Rint invite les terminologues, les lexicographes et les éditeurs de

(2) Il est vrai que la loi n'interdit que des emprunts dits «importants» ou «substantiels», mais il faut savoir que ce critère a été interprété de façon extrêmement large, du moins dans les pays de common law. Tout ce qui mérite d'être cité est important, est-on même allé jusqu'à dire.

(3) Voir, par exemple, l'al. 27 (2) A de la loi canadienne.

dictionnaires⁽⁴⁾ à se rencontrer pour en discuter et tenter de déterminer dans quelle mesure et à quelles conditions les dictionnaristes acceptent d'être cités par les terminographes. Si nécessaire, il faudrait même envisager la gestion collective par le truchement d'un organisme de perception des droits, comme il en existe dans d'autres domaines, car on imagine difficilement le terminographe demander des permissions pour chaque occurrence⁽⁵⁾. Cela dit, gestion collective ou non, ce serait tout de même bien dommage de devoir en arriver là.

1.4 Constats d'usage

Le terminographe qui élabore des fiches bilingues ou multilingues ne peut pas se permettre d'ignorer ce que les autres terminologues et lexicographes ont constaté ou proposé comme usage. Il est donc obligé de recenser les traductions trouvées dans les banques et dictionnaires bilingues ou multilingues, mais les auteurs et que les autres terminologues et lexicographes ont constaté ou proposé éditeurs visés accepteraient-ils de lui accorder une licence générale? Sinon,

(4) Ce problème ne concerne pas exclusivement les dictionnaristes francophones, bien évidemment. Pour le terminographe qui travaille en contexte bilingue ou multilingue, les définitions empruntées peuvent être dans n'importe quelle des langues en cause.

(5) Voir, à ce sujet, ALAI 1990: 323-325.

(6) Ici, encore, il y a lieu de distinguer le cas du fichier terminologique interne de celui de la banque de terminologie. Il est certain que la gravité de la faute, le cas échéant, s'amplifie dans ce dernier cas.

doit-il donc, chaque fois, demander la permission?

En pratique, il faut sans doute distinguer les emprunts isolés des emprunts systématiques. Ce dernier cas peut fort bien se produire surtout lorsque la recherche terminographique porte sur un domaine précis et qu'il n'existe qu'une ou deux autorités lexicographiques en la matière (v. g. le dépoussiérage industriel). Le risque de poursuite pour contrefaçon est alors plus grand. Et même dans le cas d'emprunts isolés, les emprunts finissent par s'accumuler, principalement quand les terminographes sont plusieurs à fournir la même banque.

Si la traduction proposée par cette autre source est originale, il y a contrefaçon aux yeux de la loi, si légère soit la faute⁽⁶⁾. Mais comme nous l'avons vu plus haut, les informations contenues dans une banque ou un dictionnaire ne donnent pas toujours prise à un droit d'auteur. Ainsi donc, quand l'auteur de l'ouvrage ne fait que répéter une traduction qu'on peut trouver dans quantité de dictionnaires bilingues, aurait-on raison de soutenir que son effort n'est pas suffisant au sens de la loi pour répondre au critère de l'originalité? Je ne le crois pas. Même si la traduction n'est pas originale en soi, le fait qu'elle ait été recensée (et à plus forte raison recommandée) par l'auteur en question suffit amplement pour lui conférer un droit d'auteur.

Je ne vois pas d'autres solutions que celle que je viens de proposer pour les emprunts de définitions, c'est-à-dire une concertation entre les parties en cause.

2 Conclusion

On peut analyser le droit tant qu'on veut, on n'arrivera jamais à des solutions satisfaisantes. *Primo*, la législation actuelle est dépassée,

particulièrement en regard des moyens de communication modernes. *Secundo*, même si la loi était réformée, il n'est pas certain qu'elle tiendrait compte des usages établis. Aussi y a-t-il lieu de faire enquête sur les pratiques actuelles des terminographes et des lexicographes, afin de mieux comprendre la façon dont se passent réellement les choses. Le Rint serait ici l'opérateur tout indiqué.

L'enquête pourrait être l'amorce d'un dialogue utile entre lexicographes et terminographes. Il se peut qu'on arrive à la conclusion que les emprunts mutuels sont si fréquents qu'il n'y ait pas lieu, dans ce domaine, d'invoquer le droit d'auteur; c'est, personnellement, ce que je souhaiterais.

Il serait illusoire, en tout cas, de croire que nous n'avons qu'à nous arranger entre nous, tout comme il serait naïf de penser que les réformes législatives prendront en compte les préoccupations et les intérêts particuliers des langagiers. À cet égard, le Rint est en mesure d'intervenir avantageusement auprès des pouvoirs politiques, notamment dans des forums internationaux, pour les sensibiliser aux effets nocifs d'une législation qui se révélerait défavorable au libre échange des informations terminologiques.

*Gérard Snow,
Centre de traduction et de terminologie
juridiques,
Université de Moncton,
Canada.*

Bibliographie

ALAI (Association littéraire et artistique internationale), 1990: L'informatique et le droit d'auteur. Actes du 57^e Congrès de l'ALAI (Québec, septembre 1989), Cowansville (Québec), Éditions Yvon Blais.

Nabhan (Victor), 1992: Droit d'auteur et banques d'information dans l'Administration, Québec, Publications du Québec.

Tamaro (Normand), 1994: Le droit d'auteur: fondements et principes, Montréal, Presses de l'Université de Montréal.

Le cadre contractuel des banques de terminologie

La constitution, la gestion et l'utilisation des banques de terminologie impliquent une multiplicité d'acteurs qui interviennent à chaque étape du processus : de la création à l'utilisation de la banque. Certains peuvent même être amenés à jouer plusieurs rôles et être à la fois fournisseurs de données et utilisateurs de la banque. Seront examinés en premier lieu les points communs aux différentes opérations contractuelles, pour esquisser ensuite quelques-unes des facettes de chacune des opérations en faisant ressortir la place fondamentale qui devra être réservée à l'échange d'informations, préalable indispensable pour établir des bases solides aux différentes conventions.

Termes-clés :
convention ; droits d'auteur ;
information réciproque.

Les juristes ont aussi leurs spécialistes de la terminologie et il me semble opportun d'introduire cet exposé par une définition de ce que l'on peut entendre par contrat afin de resituer ce que nous présenterons sur le cadre contractuel des banques de terminologie.

Dans le *Vocabulaire Cornu* le sens précis du terme contrat renvoie à : « espèce de convention ayant pour objet de créer une obligation... ». Cette première définition nous renvoie à celle de convention qui est : « le nom générique donné – au sein des actes juridiques – à tout accord de volonté entre deux ou plusieurs personnes destiné à produire un effet de droit quelconque : créer une obligation... ».

Quand deux ou plusieurs personnes, physiques ou morales, s'entendent sur un projet commun elles passent une convention, un contrat qui est générateur de droits et d'obligations pour chacune d'entre elles.

Pourquoi « cadre contractuel » et non pas contrat ou convention ? La « gestion » des banques de terminologie, telle qu'elle est entendue (au sens large), implique l'intervention d'une multiplicité d'acteurs qui n'ont pas tous les mêmes rôles et qui interviennent dans des opérations successives ou concomitantes. Celles-ci n'ont pas toutes les mêmes finalités même si elles ne peuvent être dissociées du projet global.

Avant de décomposer les conventions relatives à ces différentes opérations, on s'attachera à attirer

l'attention sur un certain nombre de points communs à chacune de ces opérations, points essentiels pour qu'une convention soit correctement formée et remplisse son plein effet.

1 Les caractères communs aux opérations contractuelles relatives à la gestion des banques de terminologie

On ne reprendra ici que quelques-uns de ces caractères en privilégiant ceux qui ont une implication dans le secteur qui nous concerne. On traitera donc de ce qui relève :

- De la formation des conventions (la nécessité d'un consentement éclairé) ;
- De la nature de l'objet (une œuvre de l'esprit) ;
- De l'interdépendance entre les différentes opérations ;
- Du caractère international des conventions.

1.1 La formation des conventions

Pour qu'un contrat soit formé, il faut qu'il y ait une rencontre des volontés sur les obligations respectives. Pour que les droits et obligations soient connus des parties, il faut que celles-ci soient informées sur ces conditions.

Afin d'éviter toute source d'ambiguïté, il est donc souhaitable que soient déterminés tant l'objet du contrat (sa finalité) que les conditions

de son exécution ou les conséquences de son inexécution.

Mais la question de savoir s'il est possible de tout prévoir à l'avance ne peut manquer de se poser. Les parties ne sont-elles engagées que par ce qui est écrit et seulement par ce qui est écrit (ou a été décidé)? Ou peut-on mettre à la charge des parties les obligations que les usages ou encore la bonne foi peuvent impliquer?

Dans les contrats civilistes, on laisse une large part à l'interprétation tout en prenant soin de déterminer, le plus souvent comme en *common law*, expressément les droits et obligations. Dans les contrats de *common law*, il est souvent question de caractère raisonnable. Ce qui importe aussi c'est que soit prévue, dans l'un et l'autre cas, une possibilité d'aménager ou de modifier le contrat en fonction des circonstances.

1.2 La nature de l'objet

Le travail du terminologue est de « définir », celui du juriste de « qualifier ». Qu'est-ce qu'une banque de terminologie au regard du droit?

Une banque de terminologie est un recueil de données qui sont disposées de manière systématique ou méthodique et qui sont individuellement accessibles⁽¹⁾. Il faut donc distinguer le contenu (les données) du contenant (le recueil, la base ou banque de données).

La banque de terminologie (contenant) est considérée par le droit comme une œuvre de l'esprit susceptible d'être protégée, les droits sur cette œuvre pouvant appartenir à ceux qui l'ont conçue et créée ou à ceux qui l'ont commanditée. Il faudra donc tenir compte dans les montages contractuels de l'existence de ces droits.

(1) Cette définition s'inspire de celle de la directive européenne sur la protection des bases de données du 11 mars 1996.

Le contenu (à savoir les fiches terminologiques) doit faire l'objet d'une analyse juridique. Il ne fait pas de doute qu'elles soient une œuvre (travail, activité et, par extension, résultat de ce travail ou de cette activité). Mais sont-elles toujours des œuvres de l'esprit, empreintes de l'originalité de leur auteur, susceptibles d'être protégées? Même si elles ne sont que le résultat d'un travail sans être original, le réalisateur de ce travail peut revendiquer le droit d'être identifié comme étant « l'auteur » de ce travail. Peut-on parler d'originalité quand l'auteur a repris une définition existante ou a relevé un emprunt?

Bien que ces problèmes ne puissent être tranchés *a priori*, car les réponses à ces questions dépendent du contexte et de chaque cas d'espèce, les différentes conventions auront à tenir compte de l'existence et de la gestion des droits sur ces œuvres.

1.3 L'interdépendance entre les différentes opérations

Les différentes opérations contractuelles relatives aux banques de terminologie s'inscrivent dans le cadre général de la politique commune des partenaires qui sont à l'origine de cette initiative. Même si elles peuvent ne pas concerner les mêmes acteurs, elles restent étroitement imbriquées les unes par rapport aux autres. Le cahier des charges qui préside à la création de la banque de terminologie s'imposera aux personnes qui alimenteront la base et à ceux qui l'utiliseront. Ce cahier des charges servira de référence pour les contrats passés avec les personnes chargées de la gestion technique.

Cette interdépendance va plus loin et chacune de ces opérations contractuelles, même si elles lient des personnes différentes, pourra subir les

conséquences de la disparition d'un des liens contractuels.

Si les partenaires se désengagent de la création cela aura une incidence sur les conventions relatives à l'utilisation ou encore à la gestion technique. Mais il ne s'agit que d'une possibilité et il est souhaitable de déterminer les limites de l'interdépendance. Un changement de gestionnaire peut n'avoir aucune incidence sur l'alimentation ou l'utilisation de la banque à condition qu'il soit fait référence au même cahier des charges.

Le départ d'un partenaire modifie-t-il la survie de la banque, son alimentation et son utilisation?

1.4 Le caractère international des conventions

L'origine des acteurs mais aussi l'espace géopolitique dans lequel s'inscrivent les banques de terminologies rendent nécessaire la prise en compte de la diversité des droits.

Toutefois, en matière de contrats, ces frontières ne sont pas toujours un obstacle. Si les législations nationales divergent d'un pays à l'autre ou d'un système juridique à l'autre, les parties peuvent aménager comme elles l'entendent les droits et obligations qu'elles veulent se donner. Elles peuvent aussi aménager la gestion de leurs conflits (renvoi à une juridiction d'un pays, à un droit applicable, à un arbitrage...).

L'utilisation des nouveaux médias, et particulièrement l'usage d'Internet, impose malgré tout une attention particulière sur la détermination de règles du jeu communes, détermination qui ne pourra se faire sans l'intervention des acteurs concernés et qui aura à prendre en compte les spécificités de chacun des réseaux existants.

Après ces quelques remarques relatives au cadre contractuel dans son ensemble, on esquissera quelques-unes des facettes des différentes opérations.

2 Les questions particulières aux différentes opérations contractuelles

Celles-ci seront distinguées en tenant compte des finalités poursuivies, à savoir :

- La création de la banque ;
- L'alimentation de la banque ;
- L'utilisation de la banque ;
- La gestion technique de la banque.

2.1 La création de la banque

Il s'agit d'une opération contractuelle qui peut être qualifiée de « contrat cadre » qui définit les objectifs de la politique poursuivie et détermine pour les différents acteurs concernés le rôle respectif de chacun.

Les parties aux contrats sont les « initiateurs » de la banque. Ils s'entendent à travers ce qu'ils peuvent appeler un cahier des charges sur les modalités de la mise en œuvre de leur politique commune. Ce contrat cadre est formalisé par un écrit et est le résultat d'une négociation.

2.2 L'alimentation de la banque

Certains pourraient penser qu'il n'y a pas de contrat propre en ce qui concerne l'alimentation de la banque et pourtant l'opération de remise d'une fiche au gestionnaire de la base est l'exécution d'une convention entre l'auteur de la fiche (ou titulaire des droits sur la fiche) et le gestionnaire. Cette convention renvoie aux conditions fixées dans le cahier des charges du contrat cadre, mais elle a

aussi son autonomie car elle engage le gestionnaire et l'auteur de la fiche à respecter les conditions de l'alimentation.

Comment se forme cette convention ? Le plus souvent la convention sera formée lors de la remise de la fiche et le contenu des droits et obligations de chacun sera précisé à ce moment là. On ne saurait trop insister sur l'importance de l'information réciproque : les conditions du cahier des charges ont-elles été portées à la connaissance des auteurs de fiches ? Comment le titulaire des droits sur la fiche fait-il savoir ce qu'il autorise ou n'autorise pas ?

2.3 L'utilisation de la banque

La finalité d'une banque de terminologie, c'est son utilisation. La question de cette utilisation doit être traitée différemment dans le cas où elle est accessible à un réseau fermé ou à un réseau ouvert.

Mais qu'il s'agisse d'un réseau ouvert ou fermé, l'utilisation d'une fiche ou d'un groupe de fiches recouvre différentes opérations. Il peut s'agir de l'utilisation intellectuelle, mais aussi de la reproduction de la fiche, de l'utilisation de la fiche dans un nouveau travail, de la traduction ou de l'adaptation. De quels droits dispose le futur utilisateur ? Quels droits sont cédés par le titulaire ? Quelle sera la finalité de l'utilisation ? Là encore, il s'avère nécessaire qu'il y ait une information réciproque afin que soient respectés les droits et obligations de chacun des cocontractants.

Si l'on peut parler de convention entre l'utilisateur et le gestionnaire de la base, ce dernier agit pour partie comme mandataire du titulaire des droits sur la fiche.

Dans le cadre d'un réseau fermé, les utilisateurs potentiels peuvent, en

adhérant au réseau, s'engager explicitement à respecter les conditions d'utilisation. Pour un réseau ouvert, ce ne sera plus l'adhésion au réseau mais l'accès à la banque de données qui sera considéré comme une reconnaissance implicite des conditions d'utilisation. Cela implique, bien entendu, une information préalable sur les conditions d'utilisation pour que celles-ci puissent être opposables aux utilisateurs.

2.4 La gestion technique

Pour mettre en œuvre le cahier des charges relatif à la gestion de la banque, il faut faire appel à des prestataires qui, d'une part, réaliseront les outils nécessaires et, d'autre part, assureront les prestations de service nécessaires. Sur la base du cahier des charges, des conventions seront passées entre les initiateurs de la banque et les prestataires choisis. Pour éviter toute ambiguïté, il est de l'intérêt de chacun que soient précisées :

- La nature de l'ouvrage à réaliser et des services à assumer ;
- Les conditions d'exécution de l'ouvrage et du service ;
- Les conditions de réception ;
- Les conséquences en cas d'inexécution.

Le résultat dépend de la façon dont les deux parties au contrat ont collaboré à son exécution. Ce principe est aujourd'hui rappelé par la doctrine et les tribunaux quand il y a des litiges relatifs aux contrats d'entreprise ayant pour objet des « nouvelles technologies ». Il ne peut manquer de s'appliquer au secteur qui nous concerne.

En conclusion, seules quelques pistes de réflexion ont pu être ouvertes, mais elles ont eu pour but d'attirer l'attention sur le rôle du contrat comme moyen d'éviter des échecs préjudiciables et d'assurer la

sécurité juridique dans les relations
entre les partenaires qui partagent le
même objectif de diffusion de la
langue française.

*Isabelle de Lamberterie,
CECOJI (Centre d'études
sur la coopération juridique
internationale),
CNRS,
France.*

M. André Clas

fait remarquer que le problème soulevé est capital et, bien que n'étant pas nouveau, il est plus aigu compte tenu des nouvelles technologies. Les propriétaires de bases de données se protègent comme ils le peuvent de la contrefaçon, notamment en insérant des « sous-marins » révélateurs du plagiat éventuel. Par ailleurs, les contrats sont rendus nécessaires, notamment en raison des codes de déontologie, mais ils peuvent paralyser l'action en restreignant le droit de reproduction des données, mêmes transformées.

Faisant référence à l'exposé de M. Snow, M. Jean-Guy Meunier demande ce que l'on entend ici par *expression*? M^{me} de Lamberterie soutient que l'on doit distinguer, en matière de propriété intellectuelle, le fond et la forme, cette dernière étant la seule protégée en vertu du principe général de la libre circulation des idées. S'agissant d'une œuvre de l'esprit, c'est la manière dont elle est formalisée qui peut être protégée et c'est là le sens qu'il faut donner au terme *expression*. Par contre, s'il s'agit de protéger le procédé qui a mis en œuvre, c'est-à-dire le contenu de ce qui est communiqué, on est sur le terrain de la propriété industrielle et la protection par le brevet protège le fond. Mais quand on protège quelque chose par le brevet ou qu'on obtient la protection de la création d'un

procédé par le brevet, on est obligé de communiquer tous les éléments qui ont permis de réaliser ce brevet. Pour sa part, M. Snow répond que la forme verbale est protégeable; théoriquement, du moins, mais en pratique elle ne l'est pas étant donné que la loi exige, en plus, que l'expression fasse l'objet d'une fixation sur un support matériel qui, traditionnellement, était du papier. Tant que ce n'était pas sur papier, aussi longtemps que ce n'était que verbal on pouvait donc la copier. Par conséquent, quelqu'un qui donnait une conférence, par exemple, qui n'était pas rédigée, risquait de la voir utilisée par quelqu'un d'autre parce que elle n'était pas protégée. C'est pourquoi maintenant on utilise des appareils (magnétophones, etc.) pour la fixation, mais cela n'est pas clair.

M. Louis-Jean Rousseau s'interroge sur la question du dépouillement terminologique des textes en faisant remarquer que, contrairement à ce qui se passait à l'époque du travail artisanal, les logiciels d'aujourd'hui permettent de balayer la totalité des corpus, et il est à prévoir que des pourcentages importants d'ouvrages seront susceptibles d'être reproduits sous forme de citations dans les banques de données. M. Snow fait remarquer que les textes juridiques ne parlent

Comptes rendus des débats

pas de pourcentage, mais font référence à la notion de partie substantielle sans la définir. Il est donc évident que pour le balayage systématique d'un ouvrage, il faut obtenir l'autorisation du propriétaire de l'œuvre. M^{me} de Lamberterie fait remarquer qu'il faut tenir compte de la finalité de la citation. Les exceptions prévues à la limitation de l'utilisation d'un contenu sont les suivantes: des fins d'illustration dans l'enseignement ou la recherche scientifique à la condition qu'on indique la source, dans la mesure où elle est justifiée par le but non commercial à atteindre et enfin, lorsqu'il s'agit d'une extraction, de réutilisation à des fins de sécurité publique ou du bon déroulement d'une procédure administrative ou juridictionnelle.

Enfin, à une question de M^{me} Diane Michaud, M. Snow fait remarquer que le fait que les citations soient étalées dans le temps ne change rien aux dispositions des textes juridiques.

La gestion des banques de terminologie

B — L'organisation et le fonctionnement des banques de terminologie

Présidente de séance:
Nycole Bélanger
(Québec)

Termdat, un partenariat réussi

L'accord de coopération conclu entre le Conseil fédéral suisse et les Communautés européennes a donné naissance à Termdat et a permis de mettre en commun les efforts de développement des deux parties dans le domaine de la terminologie. Le succès du partenariat entre Termdat et Eurodicautom est lié au juste équilibre entre concertation et autonomie.

Termes-clés :

Termdat; Eurodicautom;
banque de terminologie;
échange de données; traitement
terminologique automatisé.

La banque de terminologie de l'Administration fédérale suisse (Termdat) est née d'un accord de coopération conclu en novembre 1987 entre le Conseil fédéral suisse et ce qui s'appelait alors les Communautés européennes (représentées par la Commission). Cet accord prévoit que chacune des parties, par la mise en commun des efforts de développement dans le domaine de la terminologie, et en particulier du traitement terminologique automatisé (on parlerait aujourd'hui de *terminotique*), que chaque partie donc bénéficie au maximum des moyens affectés aux travaux respectifs, évite les doubles emplois, arrive à un niveau optimal d'alimentation et d'utilisation de la banque de données terminologiques Eurodicautom.

Concrètement cette mise en commun signifie pour la Suisse, que celle-ci acquiert une copie du progiciel et une copie des données d'Eurodicautom; elle peut en disposer à sa guise pour l'usage interne suisse, c'est-à-dire pour la consultation non commerciale à l'usage de la Confédération, des Cantons, des communes et de tout organisme ou institution de droit public, telles les universités, les centres de recherche, les bibliothèques; par ailleurs le droit d'accès aux données terminologiques existantes ou à développer est réciproque (à l'exception des données classifiées secrètes). Il est également prévu que chaque partie veille à assurer l'information mutuelle régulière, la concertation et une totale réciprocité, qu'elle gère et traite ses

données en toute autonomie, prend à sa charge ses frais de développement mais prête son concours à titre gracieux à l'autre. Huit ans après sa signature l'accord n'a rien perdu de son attrait, et nul ne songe à sortir de l'oubli l'article 9 qui règle les modalités de résiliation.

1 Le fonds Termdat/Eurodicautom

La banque de terminologie de l'UE (Union européenne), qu'il n'est plus besoin de présenter, est une de ces « grandes banques », dont la fin prochaine a déjà été annoncée à plusieurs reprises; pourtant elle ne fait que croître et embellir. Après de nécessaires actions d'épuration – qui se poursuivent toujours et qui vont être suivies d'une action de consolidation – le fonds se situe actuellement aux alentours de 840 000 fiches de terminologie et de 197 000 fiches d'abréviation. Ces chiffres en disent évidemment long sur les domaines abordés, de tous ordres, avec des pointes marquées dans les domaines traditionnels des grandes administrations publiques (économie, finances, administration, etc.), mais aussi des percées importantes dans de nombreux domaines techniques et scientifiques de pointe, lesquels reflètent ainsi les préoccupations de l'heure (environnement, biotechnologies...) sans parler du domaine strictement politique (Traité de Maastricht par exemple).

Multidisciplinaire, la banque Eurodicautom/Termdat est aussi

remarquablement multilingue, puisqu'elle s'efforce de traiter à parité toutes les langues de l'UE. On en compte neuf actuellement et l'ajout du finnois et du suédois est imminent; quant au latin, précieux pour les nomenclatures scientifiques, il vient de conquérir son autonomie et est désormais directement interrogeable après avoir été longtemps confiné à la partie française de la fiche. Les restrictions de domaine, les choix et changements de langue sont des opérations simples et rapides, comme d'ailleurs la consultation de la banque en général. La Suisse alimente surtout la banque dans ses trois langues officielles (à ne pas confondre avec langues nationales), soit l'allemand, le français et l'italien, cet ordre étant non seulement l'ordre alphabétique, mais aussi l'ordre de grandeur des trois communautés linguistiques en question.

2 Gestion de la banque

L'accord entre l'UE et la Suisse prévoit que chaque partie génère et traite ses données en concertation certes avec l'autre, mais pourtant en toute autonomie. La concertation porte évidemment sur la méthodologie et le format de la fiche, tous deux parfaitement conformes d'ailleurs aux normes internationales en la matière. L'autonomie quant à elle permet de gérer le réseau des collaborateurs et utilisateurs de Termdat à l'échelon régional, c'est-à-dire à celui de l'Administration fédérale suisse et des administrations ou organismes d'intérêt public qui s'y rattachent. Le lien de proximité est bel et bien existant et permet un contact quasi personnel avec les utilisateurs (environ 500). À l'heure de la mondialisation des échanges sur les autoroutes de l'information, ce souci peut faire sourire, et pourtant le

small is beautiful garde paradoxalement beaucoup de son sens; un traitement personnalisé, un échange direct avec l'équipe de gestion de la banque peut, dans un milieu où la terminologie est implantée et pratiquée relativement de fraîche date, raffermir la pratique de certains et susciter des collaborations spontanées ou plus longuement réfléchies. L'entente CH/UE permet de concilier harmonieusement le proche et le plus lointain, formule qui à coup sûr accroît considérablement la convivialité de Termdat.

3 Développement du fonds terminologique

Chaque partie développe ses collections terminologiques en fonction de ses propres priorités. La priorité pour la Suisse est de rendre compte de la terminologie contenue dans sa législation trilingue, et la plupart des travaux entrepris le sont dans cette optique. Il serait trop long de donner ici ne serait-ce qu'un aperçu de ces travaux, nous nous contenterons de signaler celui qui a été récemment réalisé sur le code civil suisse, plus précisément sur le registre foncier et qui vient de faire l'objet d'une publication. La grande autonomie dont jouit Termdat par rapport à Eurodicautom favorise la constitution de mini-réseaux locaux, plus faciles à établir et à animer.

4 Échange des données

L'échange de données entre Termdat et Eurodicautom est solidement établi depuis 1988 et a fait ses preuves. L'identité des systèmes et des logiciels à Luxembourg et à Berne fait que tout se déroule rapidement et sans problèmes.

Les données sont échangées sur bandes magnétiques en vertu du principe de l'identité de contenu des deux bases: les fiches préparées pour Termdat sont versées dans Eurodicautom et vice-versa, et leur intégrité est garantie. L'autonomie de chaque «bureau émetteur» est donc totale. Le nombre de transactions effectuées au cours de la trentaine de mises à jour annuelles atteint environ 200 000, ce qui englobe non seulement les créations, en augmentation régulière et importante, mais aussi les modifications et l'épuration du fonds. L'importance des échanges de données témoigne de l'actualité du fonds ainsi que de la vitalité de cette coopération, qui se reflète également dans le nombre croissant d'utilisateurs et utilisatrices en Suisse.

5 Développement technique

La coopération s'étend également au développement du logiciel d'interrogation. Pour des raisons de droits d'auteur, celui-ci est essentiellement développé à Luxembourg, mais sur la base d'un cahier des charges conçu à Berne et mis au point conjointement par les gestionnaires des deux banques partenaires. Le but principal est d'accroître la convivialité du logiciel en raison du nombre toujours plus important d'utilisateurs et de la diversité toujours croissante des domaines représentés. C'est ainsi qu'on s'est attaché à diversifier les possibilités d'interrogation et de paramétrage, sans rien perdre de la rapidité du temps de réponse ni de l'efficacité générale du système. La recherche s'effectue comme auparavant dans toute la banque, mais au besoin de façon beaucoup plus ciblée.

Les principales améliorations peuvent être résumées de la façon suivante:

– L'anglais n'est plus l'unique langue de dialogue avec le système. L'utilisateur peut choisir n'importe quelle langue de l'UE;

– Il est désormais possible d'éliminer les réponses partielles ou d'exiger strictement la chaîne de caractères demandée;

– L'interrogation peut se faire «à la carte»: inclusion ou exclusion de certains bureaux émetteurs, de certaines collections, ce qui a pour effet de faire baisser le bruit;

– Il est également possible de choisir les champs à afficher ou ne pas afficher;

– L'interrogation des abréviations est améliorée quant au choix des langues.

À court terme d'autres améliorations sont prévues, toujours destinées à accroître la convivialité. Les efforts portent actuellement sur l'interrogation par bureaux émetteurs et par collection selon le même principe que l'interrogation par domaines. Enfin le développement

d'une surface graphique sous *WWW* (W3) est amorcé pour Termdat et Eurodicautom, afin que les utilisateurs et utilisatrices puissent bénéficier des avantages courants aujourd'hui tels que menus déroulants, touches de fonction, technique des fenêtres.

En conclusion nous aimerions simplement rappeler que le partenariat entre Termdat et Eurodicautom tire son efficacité d'un juste équilibre entre concertation et autonomie – un peu à l'image de ce qui fait le succès du fédéralisme à la suisse – et que ce modèle nous semble pouvoir être recommandé à quiconque songe à développer sa propre banque de terminologie sans vouloir partir de zéro.

*Françoise Parc,
Chancellerie fédérale suisse,
Services linguistiques centraux,
Section de terminologie,
Suisse.*

Ivanhoé

Outil d'assistance au dépouillement, Ivanhoé permet de consigner sur une fiche modèle, dont certaines valeurs sont définies par défaut, les termes relevés dans des documents électroniques. Grâce à cet outil, le terminologue est libéré des tâches rébarbatives; il peut se consacrer à la sélection des termes et laisser à Ivanhoé le soin de transcrire les données.

Termes-clés:
Ivanhoé; aide au dépouillement;
macros; Bureau de la traduction;
Gouvernement du Canada.

1 Origine – Macros de dépouillement

L'avènement de logiciels dans la vie des terminologues a éveillé chez eux la fibre terminotique. Yvan Cloutier, terminologue à la Division de terminologie militaire et technique de la Direction de la terminologie et de la documentation du Bureau de la traduction de Travaux publics et Services gouvernementaux Canada, a développé à l'aide du langage macro de WordPerfect® 5.1 une série de macros permettant d'extraire de documents électroniques des termes (délimités à l'aide de certains symboles) et de les consigner dans un fichier pouvant être importé dans Latter®. Les macros Ivanhoé (ainsi nommées d'après leur auteur) éliminent de cette façon toute l'étape de la frappe.

Toutefois, l'arrivée de la version 6.0 de WordPerfect® a posé un problème sérieux. Le langage macro de cette dernière version était complètement différent de celui de la version 5.1. Deux solutions furent envisagées: redévelopper l'outil à l'aide du langage macro de WordPerfect® 6.0 (en sachant à l'avance qu'il faudrait reprendre les travaux dès que WordPerfect® modifierait son langage macro) ou redévelopper l'outil à l'aide d'un langage de programmation indépendant de tout logiciel de traitement de textes.

La dernière solution fut retenue. Ainsi, le développement du logiciel

fut d'abord entrepris à l'aide du langage Turbo Pascal®(1) et poursuivi par la suite avec Delphi®(2), dès l'arrivée de ce dernier sur le marché.

2 Ivanhoé – version Delphi

Dorénavant, les terminologues peuvent dépouiller à l'aide du logiciel de traitement de textes de leur choix, la seule contrainte étant que le logiciel choisi permette de sauvegarder les fichiers en format Ascii.

Pour un fonctionnement optimal du logiciel, certaines conventions ont été adoptées:

- Le terminologue doit délimiter les termes à l'aide des symboles << >> lorsqu'il veut importer comme contexte ou comme exemple la phrase dans laquelle figure le terme marqué, ou encore les symboles [[]] lorsqu'il désire importer le terme seul, sans contexte ou exemple d'utilisation. Les deux types de marqueur peuvent être utilisés indifféremment dans le même contexte;
- L'indicatif de page, accompagné du symbole @@, doit figurer en début de page, par exemple @@100;
- Pour les documents bilingues, des indicatifs numériques facultatifs permettent d'associer un terme et son équivalent s'ils n'apparaissent pas dans le même ordre dans les deux textes. Cette caractéristique prend toute sa valeur dans le cas, par exemple, de documents comportant un glossaire et dont les termes sont classés par ordre alphabétique, en fin de document.

(1) Turbo Pascal® est une marque de commerce de Borland International, Inc.

(2) Delphi® est une marque de commerce de Borland International, Inc.

2.1 Fonctionnement

En bref, le terminologue marque les termes dans les documents électroniques, les sauvegarde en format Ascii et lance le programme Ivanhoé dans lequel il aura préalablement inscrit, dans les cases appropriées, les valeurs par défaut de la fiche modèle. Ivanhoé extrait les termes marqués, les indicatifs de pages et, selon le cas, les contextes. Il regroupe les données en fiches qu'il verse ensuite dans un fichier conforme à un format d'importation Latter[®]. Le terminologue récupère le fichier dans Latter[®]. Les fiches feront l'objet de recherches complémentaires en vue de leur éventuelle exportation vers Termium[®] ou vers Publiciel[®], ou vers les deux.

La figure ci-dessous montre la fenêtre qui apparaît à l'écran dès que le programme est lancé. Dans l'exemple ci-dessous, on présume que le terminologue dépouille en anglais et en français. Le terminologue a déjà inscrit les valeurs appropriées.



Document anglais

@@100

With «1 Nato countries» no longer feeling the threat of massed Warsaw Pact armour, and Middle Eastern countries who had purchased «2 tank» s from the former Soviet Union finding that it is easier to replace equipments with Western-built technology rather than depend on the erratic supply of critical spares, the business of retrofitting armoured

vehicles is becoming increasingly lucrative.

Document français

@@67

Depuis que les pays du Moyen-Orient qui avaient acheté leurs «2 char» s en Union Soviétique se sont rendus compte qu'il était plus facile de remplacer les équipements de ces chars par des matériels de technologie occidentale plutôt que de tabler sur un approvisionnement hasardeux en rechanges, la modernisation des véhicules blindés est devenue pour les industriels une activité de plus en plus lucrative. De plus, les [[1 pays membres de l'OTAN]] ont...

Le fichier d'importation se présente comme suit :

_A, S = 7

!10 = TER

!13 = B

!16 = EN-FR

!30 = BBB#AAA

!60 = Nato countries*a*COR

!63 = EX*With Nato countries no

longer feeling the threat of massed

Warsaw Pact armour, and Middle

Eastern countries who had purchased

tanks from the former Soviet Union

finding that it is easier to replace

equipments with Western-built

technology rather than depend on the

erratic supply of critical spares, the

business of retrofitting armoured

vehicles is becoming increasingly

lucrative.*a

!70 = pays membres de

l'OTAN*b*COR

!140 = a*TexteAnglais*1995****100#

b*TexteFrançais*1995****67

!400 = projet_LATTER

!150 = RACINE

_A, S = 7

!10 = TER

!13 = B

!16 = EN-FR

!30 = BBB#AAA

!60 = tank*a*COR

!63 = EX*With Nato countries no

longer feeling the threat of massed

Warsaw Pact armour, and Middle

Eastern countries who had purchased tanks from the former Soviet Union finding that it is easier to replace equipments with Western-built technology rather than depend on the erratic supply of critical spares, the business of retrofitting armoured vehicles is becoming increasingly lucrative.*a

!70 = char*b*COR

!73 = EX*Depuis que les pays du Moyen-Orient qui avaient acheté leurs chars en Union Soviétique se sont rendus compte qu'il était plus facile de remplacer les équipements de ces chars par des matériels de technologie occidentale plutôt que de

tabler sur un approvisionnement

hasardeux en rechanges, la

modernisation des véhicules blindés

est devenue pour les industriels une

activité de plus en plus lucrative.*b

!140 = a*TexteAnglais*1995****100#

b*TexteFrançais*1995****67

!400 = projet_LATTER

!150 = RACINE

3 Conclusion

L'exemple ci-dessus montre bien le fonctionnement du programme dans le cas le plus complexe, c'est-à-dire lorsqu'il y a utilisation des indicatifs numériques. En conclusion, on peut donc dire que le logiciel Ivanhoé est très utile à titre d'aide au dépouillement, car il laisse au terminologue le soin d'identifier les termes, de façon simple, tout en lui permettant d'accélérer les étapes liées au transfert des données dans le poste de travail du terminologue Latter[®].

Roger Racine,

Direction de Technologie, recherche et

développement,

Bureau de la traduction,

Travaux publics et Services

gouvernementaux Canada,

Ottawa,

Canada.

La configuration des outils de gestion de Termium

Beaucoup de professionnels du milieu langagier connaissent Termium® sur CD-ROM.

Cependant, peu d'entre eux sont au fait qu'il existe une version en direct de Termium® qui est à l'origine de la version sur CD-ROM. La présentation porte sur Termium® en direct, plus spécifiquement sur la couche logicielle qui se situe entre la base de données et l'interface de consultation de la banque.

Mots-clés:

Termium®, base de données linguistiques, Bureau de la traduction, Gouvernement du Canada; CD-ROM.

Beaucoup de professionnels du milieu langagier connaissent Termium® sur CD-ROM. Cependant, peu d'entre eux sont au fait qu'il existe une version en direct de Termium® qui est à l'origine de la version sur CD-ROM. Commençons par voir ce qui distingue les deux versions.

Termium® en direct est le point de destination de la majorité des travaux terminologiques à la Direction de la terminologie et de la documentation du Bureau de la traduction de Travaux publics et Services gouvernementaux Canada, c'est-à-dire que toutes les fiches terminologiques produites sont versées dans Termium® à un moment ou l'autre. De plus, l'accès à Termium® en direct est réservé aux terminologues et traducteurs du Bureau, ce qui explique la faible visibilité de Termium® en direct à l'extérieur de l'organisme. Contenant des fiches multilingues, la version en direct tourne sur un mini-ordinateur VAX® (1) de Digital Equipment.

Medium de diffusion de Termium® à l'extérieur du Bureau de la traduction, Termium® sur CD-ROM contient uniquement les données des volets anglais et français de Termium® en direct. Les données nécessaires à la production du CD-ROM sont extraites du système en direct, de sorte que la version CD-ROM est en pour ainsi dire un instantané des données de Termium® en direct au moment où l'extraction des données a lieu.

La présentation portera uniquement sur Termium® en direct,

système qui sert à la gestion des données terminologiques. Tout au long de la présentation, j'établirai une comparaison entre l'ancienne version de Termium® en direct (Termium® III) et la version la plus récente (Termium® IV) pour mettre en lumière les innovations. Cette comparaison permettra aussi de montrer que les limites d'une application sont souvent étroitement liées au logiciel utilisé, c'est-à-dire le SGBD (système de gestion de base de données) dans le cas de Termium®. La présentation se terminera par un bref survol du contrôle de la qualité.

1 Comparaison Termium® III — Termium® IV

La comparaison portera sur les éléments suivants:

- SGBD;
- Fiche;
- Saisie des données;
- Mise à jour;
- Consultation par les utilisateurs.

1.1 SGBD

Termium® III a été développé à l'aide du logiciel Basis®(2) de la société américaine IDI (Information Dimensions, Inc. Apparu au début des années quatre-vingt, ce SGBD se situe entre les types hiérarchique et réseau et est qualifié comme étant un système d'extraction de textes (*Text-Retrieval System*). La caractéristique de Basis® demeure son orientation textuelle et sa souplesse pour ce qui est des champs de type texte. Comme

(1) VAX® est une marque de commerce de Digital Equipment Corporation.

(2) Basis® est une marque de commerce de Information Dimension, Inc.

beaucoup de SGBD de l'époque, il permet l'ouverture d'une seule base de données à la fois.

Basisplus[®] (3) est le rejeton relationnel de la société IDI. Mis au point à la toute fin des années quatre-vingt et au début des années quatre-vingt-dix, il s'agit d'un SGBD relationnel qui conserve toutefois son orientation textuelle. Cependant, l'affiliation avec son prédécesseur Basis[®] étant quasi inexistante, la mise au point de Termium[®] IV a bel et bien consisté en un redéveloppement au lieu d'une simple mise à niveau. Si on le compare aux autres logiciels de même calibre, Basisplus[®] permet des enregistrements comptant un plus grand nombre de caractères, en raison de son orientation textuelle. De plus, comparativement à son ancêtre Basis[®], il offre la possibilité d'ouvrir simultanément plusieurs bases de données. Nous verrons à la rubrique suivante les avantages inhérents à cette dernière caractéristique en ce qui concerne la fiche.

1.2 Fiche

Termium[®] III pouvait compter au maximum deux volets linguistiques sur quatre volets possibles. La fiche comportait des champs pour les vedettes, synonymes et blocs justification anglais, français, espagnol et allemand. Tous les renseignements devaient figurer en clair sur la fiche du fait qu'il n'était possible d'ouvrir qu'une seule base de données à la fois.

Grâce à sa structure relationnelle évoluée, la fiche Termium[®] IV est constituée de plusieurs volets dispersés dans des tables différentes. Tous les volets sont regroupés en une fiche terminologique grâce à une clé

d'affiliation numérique. Dans la structure actuelle, la fiche Termium[®] IV peut compter des données en 13 langues. Comme il est possible d'ouvrir plusieurs tables et bases de données en même temps, la fiche terminologique peut comporter des codes qui sont convertis au moment de l'affichage. Termium[®] IV permet aussi d'associer à la fiche des dossiers de synthèse qui sont des articles donnant des éclaircissements ou des précisions quant à l'utilisation d'un terme ou d'une expression.

1.3 Saisie

Dans Termium[®] III, toutes les fiches modifiées et les nouvelles fiches dactylographiées par les opérateurs étaient versées dans un fichier de transactions. Ce dernier souffrait de temps à autre de corruption à la suite d'une erreur d'entrée-sortie sur un disque, problème très courant à l'époque avec les utilitaires de gestion de fichiers séquentiels ou indexés. La corruption du fichier de transactions se traduisait souvent par la perte des données, entraînant du même coup la perte du temps de saisie. Au point de vue de la validation des données, l'ensemble de cette étape très importante pour le contrôle de la qualité était effectué lors de la mise à jour, donc en différé. Si une erreur était détectée, la fiche demeurait dans le fichier de transactions jusqu'à ce que la correction soit apportée et que la fiche puisse passer la validation et être versée dans la base de données.

La saisie de données dans Termium[®] IV est effectuée dans une base de données jumelle de Termium[®] sur le plan de la structure. Ainsi, les créations et les modifications sont effectuées en direct dans cette base jumelle, permettant alors aux opérateurs de saisie de profiter en direct des utilitaires de validation. Par conséquent, lorsque

l'opérateur tente de verser une création ou une modification dans la base de saisie, toute erreur détectée par le mécanisme de validation est mise en évidence immédiatement, ce qui permet à l'opérateur d'apporter la correction sur-le-champ. Il va sans dire qu'une telle validation facilite grandement la correction des erreurs.

1.4 Mise à jour

Dans Termium[®] III, la mise à jour se résumait au versement dans la base de données des fiches contenues dans le fichier de transactions. La mise à jour avait lieu en soirée, période pendant laquelle le système demeurait inaccessible aux usagers. Un décodage rudimentaire était incorporé à la mise à jour, notamment pour ce qui est des paramètres. Les opérateurs pouvaient se contenter de taper une partie des paramètres des vedettes et, lors de la mise à jour, les paramètres étaient alors traduits dans leur forme complète avant l'introduction de la fiche dans la base de données.

Dans le cas de Termium[®] IV, la mise à jour a toujours lieu en soirée, à cette différence près que le système demeure accessible. La mise à jour est en fait une simple opération d'exportation à partir de la base de saisie et d'importation dans la base Termium[®]. L'importation se fait par table et non par fiche. Ainsi, tous les volets centraux des fiches sont d'abord importés, puis les volets anglais et ainsi de suite. Aucun décodage n'est effectué pendant cette opération car les données sont stockées avec des codes dans Termium[®] IV.

1.5 Consultation

Programmée en langage Pascal, l'interface de consultation de Termium[®] III effectuait la recherche

(3) Basisplus[®] est une marque de commerce de Information Dimension, Inc.

sur un seul fichier qui contenait tous les renseignements en clair. Il est important de noter que la fiche Termium® III, telle que vue précédemment, ne contenait que deux volets linguistiques. L'affichage était donc toujours bilingue. De façon simplifiée, le programme contenait une seule variable de type «enregistrement» («record» en Pascal). Il effectuait la recherche demandée par l'utilisateur, versait les données de la fiche dans la variable, puis affichait à l'écran le contenu de la variable. Il répétait l'opération pour chaque fiche.

L'interface de consultation de Termium® IV a été développée en langage C, langage reconnu pour sa grande vitesse d'exécution. Le programme, qui compte plus de 200 fonctions ou sous-programmes mis au point à l'interne, contient une variable pouvant accommoder le contenu de tous les volets linguistiques, c'est-à-dire 13 dans la configuration actuelle. Le programme effectue la recherche dans la table contenant le volet linguistique correspondant à la langue de départ. Il compare ensuite le profil de l'utilisateur, qui établit les langues permises, aux langues de la fiche. Il reconstitue chaque fiche terminologique en parcourant chacune des tables correspondant aux volets linguistiques inclus à la fois dans la fiche et dans le profil de l'utilisateur. Par exemple, si la fiche contient des volets anglais (EN), français (FR) et espagnol (ES), et que le profil de l'utilisateur ne permet que les langues EN et FR, seuls les volets EN et FR sont chargés dans la variable. Le programme affiche ensuite les données dans un format d'affichage bilingue ou multilingue selon le nombre de langues.

Au chapitre de la consultation, Termium® IV offre aussi des clés de recherche élaborées permettant de procéder simultanément à la recherche dans les champs vedette

principale, vedettes secondaires et justifications. Le programme contient aussi des raccourcis dans les menus et permet de réafficher les fiches avec les sources décodées en clair à partir de la base de données contenant les sources. Il est aussi possible de naviguer dans l'index de bas en haut et de haut en bas et d'y effectuer des sélections multiples (du type «toutes les fiches à l'écran», «les fiches 5 à 12», etc.) sur plusieurs écrans. Le programme offre aussi, sur demande de l'utilisateur, un bilan qui ventile par domaine les fiches répondant à la requête, ce qui s'avère utile dans le cas de recherches ciblées dans un domaine particulier.

Voilà pour la comparaison des versions III et IV de Termium®. Passons maintenant au contrôle de la qualité, autre élément important de la gestion des données.

2 Contrôle de la qualité

Vers la fin des années soixante-dix, de nombreuses fiches ont été chargées dans Termium® II par le truchement de la numérisation de fiches papier. À cette époque, la numérisation en était à ses débuts et ne faisait pas appel aux algorithmes performants connus aujourd'hui. Certaines erreurs que l'on pourrait qualifier de pernicieuses se sont alors infiltrées dans la base de données; on les qualifie de pernicieuses car il y avait quelquefois des triplets de lettres (par exemple «connaître») qui ont défié les relectures d'épreuves.

À l'été 1993, après avoir mis au point un processus réalisable sur le plan technique, la Direction de la terminologie et des services linguistiques du Bureau de la traduction a lancé un projet visant à corriger ces fautes de frappe (ou plutôt de numérisation) par l'utilisation du correcteur orthographique du logiciel de

traitement de textes WordPerfect® (4), version 5.1. Les données relues à l'aide du correcteur étaient d'abord extraites d'une version adaptée du CD-ROM qui versait les fiches en caractères Ascii dans un fichier séquentiel, dont le format de présentation était adapté à la relecture à l'aide du correcteur orthographique. Les terminologues qui relisaient les fiches ouvraient un document, identifiaient les erreurs, fermaient le document puis lançaient un programme d'extraction en Turbo Pascal® (sur leur micro-ordinateur) qui versait dans un fichier de correction les erreurs relevées et leurs coordonnées. Le terminologue apportait les corrections dans ce dernier fichier, puis le transmettait ensuite à l'équipe de saisie qui apportait les corrections sur le système en direct.

3 Conclusion

On peut voir que l'évolution entre les versions III et IV de Termium® a été très grande. Il est évident que plusieurs des améliorations dont j'ai fait mention sont attribuables à la grande expérience du Bureau de la traduction dans le domaine des bases de données linguistiques. Cependant, il ne faudrait pas minimiser l'importance de l'évolution des SGBD; il est très important de garder à l'esprit que toute évolution sur le plan du logiciel est toujours étroitement liée à l'augmentation sans cesse croissante de la vitesse d'exécution et de la puissance des nouveaux ordinateurs qui rendent maintenant possible la pleine exploitation du modèle relationnel, modèle bien adapté à la gestion des données mais reconnu en

(4) WordPerfect® est une marque de commerce de Novell Corporation.

contrepartie pour sa voracité en ressources informatiques.

*Roger Racine,
Direction de Technologie, recherche
et développement,
Bureau de la traduction,
Travaux publics et Services
gouvernementaux Canada,
Ottawa,
Canada.*

À diverses questions de la part des participants, M^{me} Françoise Parc précise que la banque Termdat n'est pas ouverte aux utilisateurs privés et qu'elle n'est pas commercialisée, mais que des échanges de contenus sont possibles. Dans tous les cas, l'alimentation se fait par la Section terminologie et non directement par des contributeurs externes. Selon l'entente établie, les partenaires sont autonomes en matière d'alimentation.

Eurodicautom ne peut modifier des fiches suisses et inversement, ce qui peut causer une certaine redondance, mais on peut limiter l'accès à certaines zones de la banque.

Pour sa part, M. Roger Racine informe les participants que tous les développements de Termium en ligne ne sont pas transférés sur la version CD-ROM, mais que la décision d'intégrer telle ou telle amélioration se fait en fonction des besoins ou de l'intérêt manifesté par les usagers.

Les outils terminotiques et le poste de travail du terminologue

Présidente de séance:
Monique Cormier
(Université de Montréal, Québec)

Thème 2

Adepte-Nomino : un outil de veille terminologique

Adepte-Nomino est un système d'aide au dépouillement terminologique de textes français ; en cela, il constitue un outil de base pour les activités de veille terminologique. Adepte-Nomino utilise l'analyseur morphosyntaxique Nomino pour l'analyse linguistique des textes et le dépistage automatique des unités complexes nominales ; il est doté de l'interface Adepte, un poste de travail dont nous présentons les fonctions : consultation des listes d'unités dépistées et accès aux contextes de ces unités, rédaction assistée de fiches et gestion de la base de fiches.

Termes-clés : dépistage de termes ; dépouillement terminologique assisté par ordinateur ; unité complexe nominale ; analyse de textes ; analyse morphosyntaxique ; veille terminologique.

La disponibilité d'outils informatiques adaptés à la recherche, à la production et à la diffusion terminologiques ou, mieux encore, développés spécifiquement pour ces activités, permet de renouveler certains procédés de travail terminologique ; grâce à ces outils, certains types d'activités naguère difficiles sinon impossibles à mener peuvent être entreprises. C'est le cas de la veille terminologique qui, tout en demeurant une activité intégrée à la chaîne de production terminologique en dehors de laquelle elle n'aurait pas de signification, peut, grâce à la disponibilité de ces outils, se dégager du rôle timide auquel elle était confinée.

La veille terminologique (sinon le terme, du moins sa réalité) a toujours fait partie de l'activité terminologique ; c'est à l'étape du dépouillement qu'elle s'est, jusqu'à présent, manifestée. Toutefois, sa portée a le plus souvent été circonscrite à la recherche thématique visant la description et l'uniformisation de terminologies de domaines ou de sous-domaines bien délimités. La veille ainsi canalisée ne produit pas ses propres résultats et n'a pas d'existence propre.

On pourrait, bien sûr, considérer le développement et la mise à jour continue d'une banque de

terminologie comme une vaste entreprise de veille qui rassemblerait et diffuserait l'information terminologique répertoriée dans l'immense production de lexiques, vocabulaires, glossaires... Ce serait, cependant, fausser un peu le sens originel de *veille*.

Que serait la veille terminologique dans l'ensemble des activités diversifiées ayant déjà cours en terminologie ? Sans vouloir en donner une description, forcément limitative, nous pourrions tenter de relever certaines caractéristiques. La recherche de la nouveauté constitue sans doute son objectif fondamental : la recherche de nouvelles terminologies mais aussi, par elles, celle de notions nouvelles, ainsi que la cueillette de données sur ces termes et notions ⁽¹⁾. La diffusion la plus rapide possible de ces termes et des données s'y rattachant, bien qu'ils ne soient pas nécessairement validés, accordera une certaine autonomie à cette activité de veille qui doit toutefois demeurer intégrée à la chaîne globale de la recherche et de la production terminologiques. Plutôt que de différer la diffusion de termes non établis ou non validés, c'est la finalisation de leur traitement que l'on remet à plus tard. Les données recueillies lors de la veille constituant un matériau brut ayant subi un traitement minimal et ne comportant pas nécessairement l'ensemble des

(1) Voir note page 33.

(1) La notion de « nouveauté » est, reconnaissons-le, assez floue : est un terme nouveau, un terme qui vient de surgir ; comment sait-on qu'il est nouveau si ce n'est en constatant qu'il n'apparaît dans aucune nomenclature terminologique. Et comme il serait extrêmement difficile, voire tout à fait impossible, de recenser de façon exhaustive l'ensemble des termes existants – leur nombre même suffit à nous en convaincre – cette hypothétique nomenclature, qui existe partiellement et qui constitue un « corpus d'exclusion », comporte des lacunes. Un terme qui ne se retrouve pas dans ce corpus d'exclusion n'est pas forcément nouveau, ce peut être un terme non recensé ou dont la diffusion est trop restreinte. Aussi faut-il considérer comme nouveaux, ces termes qu'on n'aurait pas déjà relevés (à la condition, toutefois, que ce ne soit pas des termes tombés en désuétude).

(2) Rappelons que le dépouillement terminologique consiste à relever dans des textes spécialisés des termes ainsi que des données relatives à ces termes ou aux notions qu'ils représentent, ces données étant fournies par les contextes des termes ; ces termes et données seront, par la suite, analysés et traités, mis en correspondance avec des termes d'autres langues par le biais de descriptions semblables de la notion désignée par les termes des langues étudiées. La grande difficulté du dépouillement réside dans le fait que les termes recherchés ne sont pas forcément connus au départ : lorsque l'on dépouille un texte, il ne s'agit pas de retrouver les occurrences des termes x, y ou z mais de trouver les termes (potentiels) compris dans la documentation pour ne retenir que ceux qui sont (ou peuvent être) pertinents au domaine de recherche, ce qui ne pourra souvent être déterminé de façon certaine qu'au moment de l'analyse des données. Dans un projet de veille terminologique, cependant, ce n'est plus aussi exclusivement le domaine qui oriente le choix des unités dépistées.

(3) Sur la problématique du dépouillement terminologique et, particulièrement, sur le dépistage automatique des unités complexes, voir (Perron : 1989) et surtout (David et Plante : 1991).

informations que l'on recueille généralement sur une fiche de terminologie complète (notamment, en terminologie comparée, l'établissement de correspondance d'un terme avec un terme d'une autre langue), leur diffusion intéressera davantage les terminologues pour la conduite de leurs travaux (tant en recherche ponctuelle qu'en recherche thématique) que les utilisateurs finals.

Le dépouillement terminologique constitue un moyen d'exercer la veille terminologique. Nous disposons maintenant d'outils qui assistent les terminologues dans leur travail de dépouillement terminologique⁽²⁾. C'est l'un de ces outils que nous présentons ici : le système d'aide au dépouillement terminologique Adepte-Nomino.

Mais avant d'aborder l'objet premier de cette communication, nous noterons que les outils d'aide au dépouillement ne peuvent être vraiment utiles à la veille terminologique que s'ils sont adaptés non seulement à la problématique du dépouillement lui-même mais également aux caractéristiques de la veille⁽³⁾. Or la veille terminologique nécessite une rapidité d'exécution dont dépend largement son bien-fondé. Par ailleurs, la veille doit avoir une couverture beaucoup plus large que celle des travaux de recherche thématique ; sans prétendre que la veille doive s'exercer dans tous les domaines de l'activité humaine (on délimitera sans doute son aire de déploiement selon certains critères qui différeront d'une cellule de veille à l'autre), l'extension assez large que l'on donne aux domaines faisant l'objet de la veille n'est pas sans conséquence : les terminologues qui participent à la veille ne peuvent toujours disposer d'une connaissance de domaines de spécialité aussi fine que les terminologues qui effectuent des recherches thématiques, et le projet de veille terminologique à laquelle ils peuvent participer peut

avoir une extension assez large pour empêcher une connaissance approfondie de tous les aspects d'un domaine.

Les outils auxquels on recourra dans des travaux de veille devront donc favoriser la rapidité d'exécution du travail ; l'automatisation de certains traitements, l'intégration des outils et la fiabilité des résultats produits par ces outils contribuent à cette rapidité d'exécution. C'est ainsi que les terminologues utilisant un outil destiné à leur fournir assistance ne devraient pas avoir à accomplir des tâches qu'ils n'ont pas à effectuer dans leurs travaux habituels de dépouillement « manuel » (comme des tâches d'encodage de textes, de constitution, de mise à jour et de gestion de dictionnaires qui nécessitent certains logiciels utilisés pour le dépistage de termes) ; ils devraient disposer d'un ensemble intégré d'outils nécessaires à l'accomplissement de leur travail plutôt que de programmes dispersés. C'est sans doute l'automatisation d'une partie de leur travail qui leur fera gagner du temps tout en leur permettant de fournir des résultats quantitativement appréciables. Alors se pose le problème à la fois de la qualité des résultats des traitements automatiques et des limites de l'automatisation : jusqu'à quel point convient-il d'automatiser le travail du dépouillement ? L'objectif n'étant pas d'automatiser le travail au maximum (à outrance, pourrait-on dire), on n'automatisera que les tâches qui ne nécessitent pas l'intervention humaine (c'est-à-dire les tâches qui ne relèvent pas en propre de la compétence des terminologues) mais le plus grand nombre de ces tâches ; par ailleurs, les traitements automatiques dont les résultats nuisent au travail humain (et les erreurs les plus graves ne sont-elles pas celles qui faussent les données traitées ou qui ne les prennent pas en

compte, laissant croire à leur inexistence?) démontrent la méconnaissance du matériau traité et la non-justification de l'automatisation. Pour que l'utilisation d'un système de dépouillement soit efficace, il faut que le résultat des traitements soit fiable, surtout lorsque l'on considère l'impossibilité qu'auront plusieurs terminologues qui contribuent à la veille terminologique de se familiariser avec tous les domaines dans lesquels ils auront à travailler.

1 Description générale des modules d'Adepte-Nomino

Le système d'aide au dépouillement terminologique Adepte-Nomino, conçu pour aider les terminologues dans leur activité de dépouillement terminologique,

(4) Nomino a été développé par Lucie Dumas, André Plante et Pierre Plante (ATO, Département de linguistique, Université du Québec à Montréal). Nomino fonctionne dans l'environnement Macintosh comme composante de l'Atelier FX (voir (Plante, Dumas: 1995)) et dans l'environnement Windows (le Réseau international de néologie et de terminologie – Rint – a participé au financement des travaux d'écriture de l'analyseur dans ce dernier environnement).

(5) Un fichier «point de vue» regroupe l'ensemble des unités d'un type donné dépitées par Nomino (par exemple: les noms, les verbes, les unités complexes nominales ou l'ensemble des formes fléchies).

(6) Acronyme d'«assistance au dépouillement terminologique». Adepte, développé par l'Office de la langue française, est une adaptation, sous Windows, de l'interface développée initialement pour Macintosh par André Plante et Pierre Plante (ATO-UQAM).

peut également constituer un excellent outil pour la veille terminologique. Il intègre trois modules distincts:

- L'analyseur morphosyntaxique du français Nomino⁽⁴⁾;
- Le module de création des fichiers «points de vue»⁽⁵⁾ qui exploite la sortie de Nomino en extrayant du fichier résultant de l'analyse l'ensemble des unités catégorisées, simples et complexes, dépitées par Nomino pour les répartir en différents fichiers «points de vue»;
- L'interface Adepte⁽⁶⁾ qui se présente comme un poste de travail de dépouillement terminologique à partir duquel l'utilisateur peut lancer l'analyse d'un ou de plusieurs textes et dans lequel il peut exploiter le résultat de l'analyse (consultation des différentes unités relevées par Nomino et de chacun de leurs contextes), créer des fiches et gérer les données comprises sur ces fiches.

1.1 L'analyseur morphosyntaxique Nomino

Nomino constitue le moteur linguistique du système de dépouillement. «Il trouve, pour la majorité des lexèmes d'une phrase, leur catégorie syntaxique et leur lemme. De plus, il dépitte des unités complexes nominales qui sont des candidats à être des termes polylexicaux. [...] Nomino applique, à la suite d'une analyse morphologique en profondeur de chacun des lexèmes (LCMF), une analyse syntaxique locale (DCS) qui désambiguïse les catégories syntaxiques et dépitte par la suite des unités complexes nominales (DUCN)» (Dumas, Plante, Plante 1996: 5-6).

Nomino peut être utilisé seul (les résultats de son analyse peuvent alors être lus à l'aide de gestionnaires de bases de données) ou en conjonction avec d'autres applications: l'Atelier

FX⁽⁷⁾, développé par les auteurs de Nomino, constitue l'une de ces applications tout comme Adepte-Nomino que nous présentons ici. L'utilisation de telles applications est nécessaire pour que du résultat de l'analyse de Nomino soient extraites les différentes listes d'unités et pour que les liens établis entre une unité et ses différents contextes puissent être exploités.

Nous nous limiterons, ici, aux fonctionnalités de Nomino qui sont utilisées dans Adepte-Nomino.

La chaîne de traitement

Nomino applique aux textes⁽⁸⁾ qui lui sont soumis quatre types de traitement.

(7) L'Atelier FX «est un environnement de programmation dédié à la conception de systèmes d'analyse linguistique, d'extraction d'informations dans les textes et de mise au point de progiciels à base de connaissances» (Plante 1996: 4). Il est constitué de trois modules principaux: «le langage FX qui permet la programmation en faisceaux (par Pierre Plante); le calcul de saillance FXS qui permet la comparaison des objets construits par un programme FX (par Pierre Plante et Lucie Dumas); le progiciel Nomino qui est un parseur lexico-syntaxique du français (par Lucie Dumas, André Plante et Pierre Plante)» (*id, ibid.*).

(8) Les textes que l'on fait analyser n'exigent ni préparation ni encodage préalable. Le texte doit être en format «texte» (jeu de caractères ANSI qu'utilise Windows). On peut garder trace du numéro de la page du document imprimé en inscrivant un code, dans le texte, à l'endroit correspondant à la fin de page du document imprimé; les numéros de page pourront être ajoutés à la suite de la source des termes et contextes transcrits sur fiches.

Le premier, le **traitement des marques d'édition**, « consiste à identifier les lexèmes et les phrases, à repérer les noms propres, les abréviations et autres marques d'édition » (*ibid.*: 10). L'utilisateur n'a donc pas à encoder ce texte.

Le deuxième est une **analyse morphologique** qui attribue, hors contexte, à chaque lexème « l'ensemble des catégories syntaxiques possibles de ce lexème avec, pour chacune de ces catégories, une proposition de lemme et des caractéristiques morpho-syntaxiques » (*ibid.*: 10). Lorsqu'une même forme peut correspondre à différents mots (par exemple: la forme *écrit* qui peut représenter le nom, l'adjectif ou le verbe), les différentes catégories et les différents lemmes sont attribués au lexème. L'une des principales caractéristiques de ce traitement est que, pour fournir aux mots leurs catégories et leurs lemmes, il applique

(9) On verra, plus loin, que l'ensemble des formes fléchies – toutes catégories confondues, d'ailleurs – sont également disponibles pour qui souhaiterait y avoir accès; toutefois, elles font l'objet d'une liste isolée (le « point de vue » lexicque) qui n'interfère pas avec les listes d'unités lemmatisées.

(10) L'ambiguïté mal résolue d'un mot crée du bruit: le résultat du dépistage des unités complexes comportera autant d'unités linguistiquement non pertinentes qu'il y aura d'erreurs d'analyse attribuant à un mot une catégorie dont, en contexte, il ne relève pas; toute forme de bruit crée des erreurs d'interprétation et ralentit le travail de dépouillement. Elle est également une cause de silence. D'autre part, le mauvais contrôle de l'ambiguïté marque les limites de l'automatisation des traitements; l'absence de fiabilité de l'analyse restreint d'autant l'assistance que l'on souhaite apporter aux terminologues dans leurs activités de veille terminologique particulièrement.

des règles d'interprétation des suffixes des mots plutôt que de recourir à des dictionnaires; seules les exceptions à ces règles sont comprises dans des listes. L'avantage évident de cette stratégie fondée sur l'application de règles plutôt que sur l'exploitation de dictionnaires est le traitement des néologismes ou des unités trop spécialisées pour être répertoriées dans des dictionnaires; d'autre part, comme il n'y a pas de dictionnaire de toutes les formes à catégoriser et à lemmatiser, l'utilisateur n'a pas, avant de faire analyser le texte pour en dépister les termes potentiels, à vérifier si toutes les différentes formes des mots du texte sont inscrites dans un dictionnaire, ni à compléter ces dictionnaires ni à les gérer. Notons que la lemmatisation (après que sera levée l'ambiguïté catégorielle) d'une part résout certains problèmes de silence (la recherche des occurrences de *transversal* rapportera tous les contextes comportant l'ensemble des formes fléchies: *transversal, transversale, transversales, transversaux* ⁽⁹⁾) mais aussi de bruit (la liste des unités repérées dans un texte étant réduite aux formes lemmatisées, cette liste n'est pas exagérément allongée des formes fléchies des unités et n'oblige pas l'utilisateur à gérer la dispersion des unités).

Le troisième traitement consiste en l'**analyse syntaxique** du texte qui a pour principale fonction de désambiguïser les mots ayant plus d'une catégorie; à la fin de ce traitement, chaque mot reçoit une seule catégorie déterminée par cette analyse contextuelle ainsi que le lemme qui lui convient. L'analyse syntaxique, peu pratiquée par les logiciels de dépouillement, est pourtant nécessaire pour assurer aux traitements une plus grande fiabilité et faire en sorte que les listes d'unités dépistées ne soient pas augmentées inutilement d'éléments qui ne relèvent pas de cette liste d'unités ⁽¹⁰⁾; la qualité du traitement, dans ce sens,

constitue pour les terminologues une économie de temps et d'effort et justifie l'automatisation des traitements adoptée.

Le **dépistage des unités complexes nominales** constitue le dernier traitement. Nomino distingue deux types d'unités complexes nominales, les *ucn* et les *ucna*: « La première liste appelée simplement *ucn* contient les unités considérées les plus fiables. Une seconde liste, *ucna* (unités complexes nominales additionnelles) contient d'autres *ucn* dont la structure est moins fiable. On retrouve dans la seconde liste (*ucna*) des expansions prépositionnelles construites sur les prépositions: avec, pour, sans et sur. On y dépose aussi les constructions qui contiennent des spécificateurs nominaux (*traitement de la parole, infarctus du myocarde, carte au laser*) et les expansions infinitives (*machine à coudre*). La somme des deux listes assure donc le maximum de couverture pour qui veut être certain de ne pas manquer d'*ucn* » (*ibid.*: 14).

Découpage et « combinaison interne des unités complexes »

Une unité complexe peut ne comporter qu'une expansion (ex.: *carte à mémoire*) mais peut aussi en comporter plusieurs (ex.: *carte à courbes de niveau*). « La question se pose alors de combiner les unités composantes d'une *ucn*. [...] Trois variables vont paramétrer le comportement de l'analyse combinatoire interne: *mode-ucnc1*, *mode-ucnc2* et *mode-ucnc3*; ces modes sont cumulatifs. On peut en exciter un ou plusieurs à la fois. Les *mode-ucnc2* et *mode-ucnc3* exigent toutefois que le *mode-ucnc1* soit actif » (*ibid.*: 15). Nomino permet donc à l'utilisateur de choisir l'un ou l'autre de ces modes de combinaison (nous verrons plus loin la façon de définir ces options) que nous allons illustrer avec le terme *carte à piste magnétique*.

Si l'on ne choisit aucun des modes *mode-ucnc1*, *mode-ucnc2* ou *mode-ucnc3*, on obtiendra l'*ucn* de base seulement: *carte à piste magnétique*. Le *mode-ucnc1* rapportera: *piste magnétique*. Le *mode-ucnc2* rapportera: *carte à piste*. En activant le *mode-ucnc3*, on obtiendra enfin l'unité suivante: *carte magnétique*.

Traitement de la coordination

Nomino traite des cas de coordination: «Sur le segment: *lecteur et encodeur de carte*, on aura: *lecteur de carte, encodeur de carte*. De manière symétrique pour la coordination à droite, on aura sur un segment tel: *lecteur de carte et de puce: lecteur de carte, lecteur de puce*. Ou encore, pour les adjectifs coordonnés: *système administratif ou financier: système administratif, système financier*» (*ibid.*: 16).

Sorties de Nomino

L'analyse terminée, Nomino produit deux fichiers dont le nom est constitué de la racine du nom du fichier «texte» analysé et portant les extensions. *ino* (fichier sous format tabulaire comportant les données résultant de l'analyse; voir l'exemple de la figure 1) et. *adr* (un fichier permettant de retrouver l'adresse d'une phrase dans le texte). Si l'on a inscrit des codes de numéro de page dans le texte, un troisième fichier (*.dli*) s'ajoutera.

Le fichier. *ino* qui fournit les résultats de l'analyse morphosyntaxique avec dépistage des unités complexes nominales se présente comme suit: (voir figure 1).

Ce fichier peut être exploité par un gestionnaire de base de données ou par des modules que l'on aurait construits sur mesure. Le module Adepté constitue un type de ces modules d'exploitation.

1.2 Création de fichiers «points de vue»

Une fois que Nomino a terminé son traitement, le module de création des fichiers «points de vue» est appelé automatiquement; ce module a pour fonction d'extraire du fichier. *ino* la liste des *ucn*, des *ucna*, des *noms*, des *verbes*, des *adjectifs*, et le *lexique* complet du texte (sauf les mots inscrits dans une liste de mots vides). Chaque liste est réécrite dans un fichier puis est triée; les multiples occurrences d'une même unité sont ramenées à une seule (avec renvoi à chaque phrase distincte où se trouve l'unité).

Chacune de ces listes fait l'objet d'un fichier «point de vue» dont le nom est composé de la même racine que celle du fichier «texte» analysé à laquelle s'ajoute une extension:

Phra	Cs	Lemme	Ucn	Ucna	Lexème
30					la
30	adj	nouveau			nouvelle
30	nc	carte			carte
30	vfle	permettre			permet
30					l'
30	nc	authentification	< authentification_local >	< authentification_local_du_porteur >	authentification
30	adj	local			locale
30					du
30	nc	porteur			porteur
30	ponc	.			.

Fig. 1 – Extrait du contenu d'un fichier. *ino* produit par Nomino

La colonne lexème reprend tous les mots du texte sur lesquels sont fournies différentes données: numéro de phrase (Phra), catégorie syntaxique (Cs: adjectif, nom commun, verbe fléchi, ponctuateur) et l'unité ou les unités complexes dont le mot de la colonne lexème forme la tête (Ucn: unités complexes nominales; Ucna: unités complexes nominales additionnelles)

*.ucn	pour	unités complexes nominales
*.uca	pour	unités complexes nominales additionnelles
*.per	pour	unités complexes permutées ⁽¹¹⁾
*.nom	pour	noms
*.adj	pour	adjectifs
*.ver	pour	verbes
*.lex	pour	lexique ⁽¹²⁾

Un dernier «point de vue» pourra être disponible: *ucn + ucna* qui, étant construit de façon dynamique, ne fait pas l'objet d'un fichier; ce «point de vue» fusionne la liste des *ucn* et des *ucna*.

Toute cette étape de description linguistique (y compris celle de la création des listes d'unités) est entièrement automatisée; l'intervention de l'utilisateur n'est nécessaire que pour désigner les fichiers à traiter, pour fournir les paramètres initiaux et lancer l'analyse. Le schéma suivant illustre les différents traitements appliqués à un texte et leur résultat.

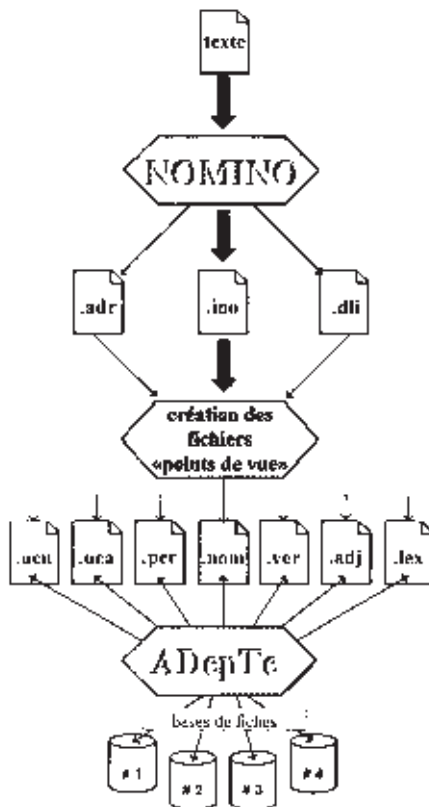


Fig. 2 – Schéma de la chaîne de traitement: de l'analyse d'un texte à la création de bases de fiches terminologiques

1.3 Adepte: un poste de travail de dépouillement terminologique

Dans l'ensemble du système d'aide au dépouillement terminologique Adepte-Nomino, l'application Adepte constitue le poste de travail; il comprend quatre environnements:

- **L'environnement de départ** dans lequel on choisit la base de fiches, on lance l'analyse de textes et l'on choisit un texte à dépouiller;
- **L'environnement de rédaction de fiches** dans lequel on consulte les différentes listes d'unités dépistées par

Nomino (la liste des unités complexes nominales, des noms, des verbes, etc., définie selon un critère fréquentiel), on sélectionne des unités pour consulter leurs contextes, on rédige des fiches;

– **L'environnement de gestion des sources**⁽¹³⁾ dans lequel on crée ou modifie un répertoire de sources et on associe au texte une source qui sera attribuée aux termes et contextes reportés sur fiches;

– **L'environnement de gestion des fiches** dans lequel on peut avoir accès au contenu des fiches pour les consulter ou les modifier et dans lequel on peut effectuer un certain nombre d'opérations dont l'importation ou l'exportation, l'impression, la suppression de fiches, etc.

Chaque environnement dispose d'une fonction d'aide (appelée par le bouton «?») soit sur l'environnement dans lequel on se trouve, soit sur chaque bouton et sur chaque champ de cet environnement.

2 Présentation d'Adepte

La présentation que nous faisons ici d'Adepte est générale; elle vise principalement à donner une vision d'ensemble du logiciel et de son utilisation. Il nous a semblé utile d'inclure dans cette présentation une méthode d'exploitation des résultats de l'analyse qui rendra l'utilisation d'Adepte plus intéressante et surtout plus efficace.

On remarquera que le traitement des textes que nous avons décrit jusqu'à présent était entièrement automatique; on verra que, dans le poste de travail, le terminologue est

(11) Ce fichier est produit par le module de création des fichiers «points de vue» à partir de la liste fusionnée des *ucn* et des *ucna*. Chaque expansion d'une unité complexe est mise en vedette; ex.: l'unité réseau de point de service donnera point: réseau de point de service et service: réseau de point de service.

(12) Le lexique est constitué de l'ensemble des différentes formes flechées comprises dans le texte sans distinction catégorielle.

(13) Une source est une référence bibliographique à un document, généralement codifiée, dont on a extrait des données qui sont inscrites sur fiche terminologique.

seul à décider de la pertinence des termes et de l'intérêt des contextes.

2.1 Environnement de départ

Après avoir lancé l'application Adepté, on se retrouve dans l'environnement de départ (fig. 3):

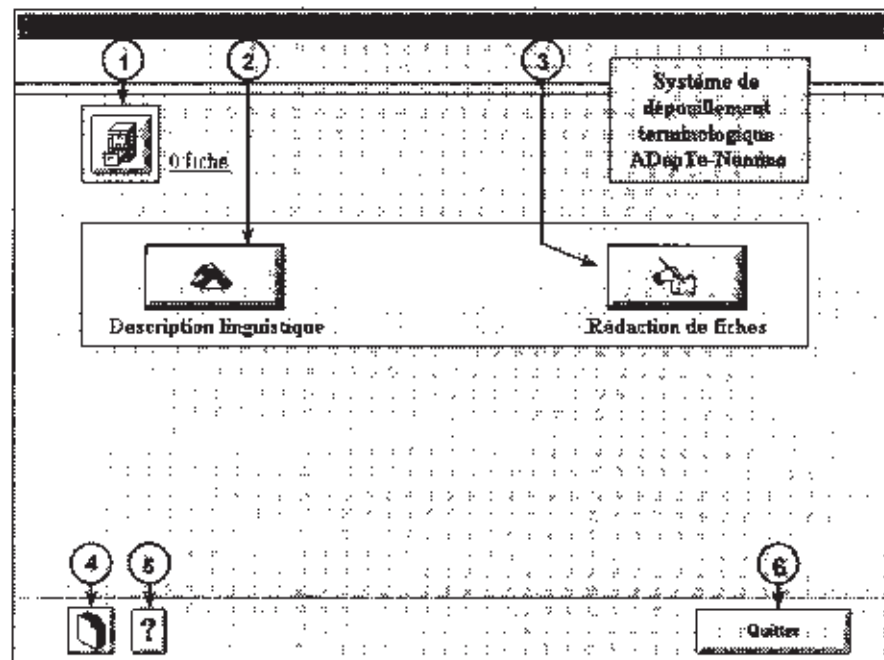


Fig. 3 – Environnement de départ d'Adepté

(bouton 4). Le bouton 5 fait afficher des rubriques d'aide. Le bouton 6 («Quitter») ferme l'application.

Création ou sélection d'une base de fiches

En créant ou sélectionnant une base de fiches (bouton 1), l'utilisateur détermine la base de travail dans laquelle seront versées les fiches qu'il rédigera ou modifiera ou dans laquelle il pourra se reporter directement (par le bouton 4). On peut disposer de plusieurs bases de fiches; toutefois, une seule base est active à la fois. Il est important de bien définir les bases de fiches que l'on crée: les fiches d'une base traitant

Dans l'environnement de départ, l'utilisateur peut créer ou sélectionner une base de fiches (bouton 1), lancer l'analyse d'un ou de plusieurs textes (bouton 2) ou avoir accès à l'environnement de rédaction des fiches (bouton 3) ou à l'environnement de gestion des fiches

de «sécurité informatique» ne sont pas disponibles dans la base «téléinformatique» (à moins qu'on ait importé les fiches de la première base dans la deuxième).

Description linguistique et création des listes d'unités

L'une des premières opérations à effectuer consiste à faire analyser les textes que l'on veut dépouiller. En cliquant sur le bouton 2 de la figure 3 («Description linguistique»), on appelle l'analyseur Nomino (fig. 4).

Nomino est une application autonome; toutefois, pour que le résultat de l'analyse de Nomino puisse être disponible et exploitable

dans Adepté, c'est de cette dernière application qu'il faut lancer Nomino.



Fig. 4 – Écran d'accueil de Nomino

Nomino dispose de sa propre interface. (Voir figure 5)

En cliquant sur le bouton 1 de la figure 5, une boîte de dialogue s'ouvre permettant de sélectionner un ou plusieurs textes à soumettre à l'analyse de Nomino. Le nom des fichiers soumis à l'analyse s'inscrit dans le répertoire de fichiers à analyser (champ 6). Par le bouton 2, on peut retirer de ce répertoire le nom d'un fichier préalablement sélectionné. Un double clic sur la case du menu «Système» (7) ferme l'application. Le bouton 5 fait afficher l'écran d'accueil qui fournit diverses informations sur Nomino (numéro de version, date et auteurs).

Options

L'analyse que fait Nomino peut être paramétrée, comme on l'a vu précédemment. En cliquant sur le bouton 3 de la figure précédente, une fenêtre s'ouvre permettant à l'utilisateur de choisir les options qui lui conviennent (fig. 6).

Pour activer une option (ou la désactiver), on clique dans la case à cocher correspondante. Nous ne

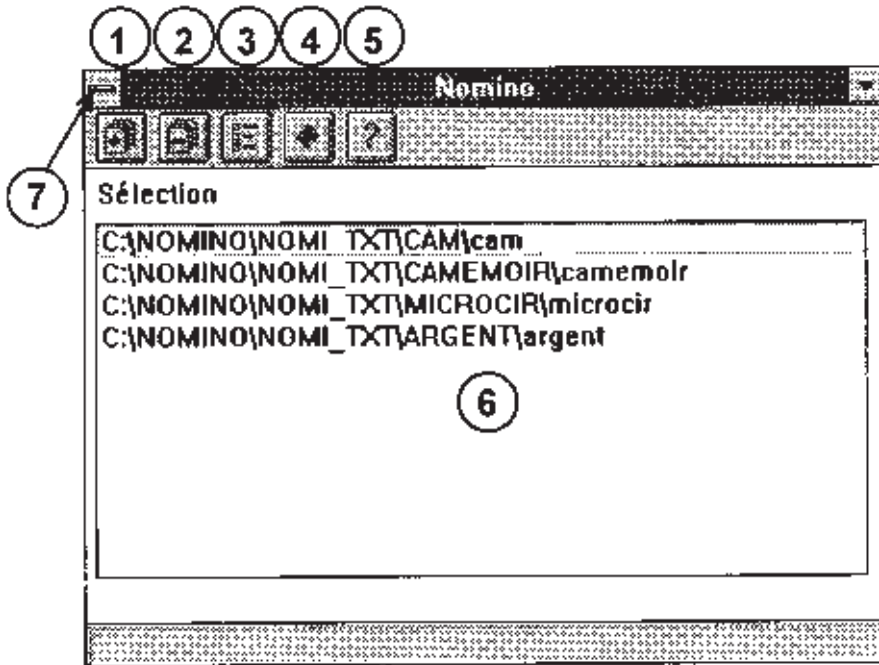


Fig. 5 – Interface de Nomino

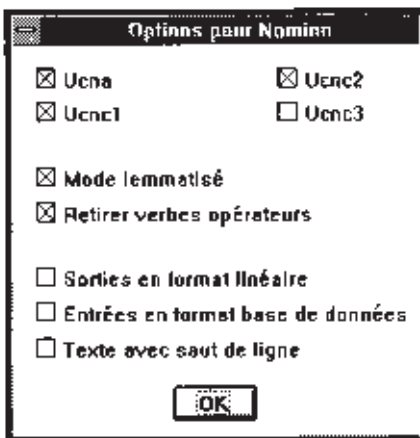


Fig. 6 – Options prédéfinies pour l'exécution de Nomino

retiendrons ici que les options pertinentes dans Adepto-Nomino. Un premier groupe d'options permet à l'utilisateur de déterminer les types d'unités qui seront relevées ou le découpage qui sera effectué sur les unités relevées (cases *ucna*, *ucnc1*, *ucnc2*, *ucnc3*).

L'option «Mode lemmatisé» doit être active pour permettre la sélection de l'ensemble des formes que peut prendre une unité complexe. Avec Adepto-Nomino il faut que les options «Sorties en format linéaire» et «Entrées en format de base de données» soient désactivées; si le texte que l'on soumet à l'analyseur comporte des sauts de ligne, on cochera cette case ⁽¹⁴⁾.

(14) Il est toutefois de loin préférable que le texte ait été enregistré en format «texte» sans saut de ligne. Lorsque l'on choisit l'option «Texte avec saut de ligne», les retours simples sont interprétés comme un espace; seuls les doubles retours sont considérés comme une fin de paragraphe. Notons que, par défaut, le retour constitue une fin de phrase, ce qui a des incidences sur l'analyse syntaxique et le résultat de cette analyse.

Analyse des textes et création des fichiers points de vue

En cliquant sur le bouton 4 de la figure 5, l'utilisateur commande l'analyse des textes inscrits dans le répertoire des fichiers (champ 6). L'analyse est entièrement automatique; il n'y a aucune interaction avec l'utilisateur. La durée relative de cette analyse est indiquée en bas de la fenêtre. Cette durée varie considérablement en fonction de l'appareil dont on dispose.

Une fois l'analyse terminée, Nomino appelle, sans l'intervention de l'utilisateur, le module de création des fichiers «points de vue» et se ferme automatiquement. Le module de création des fichiers «points de vue» extrait du résultat de l'analyse de Nomino des listes d'unités qui seront exploitables dans Adepto. À la fin de cette étape, ce module se ferme automatiquement; l'utilisateur se retrouve dans l'environnement de départ de l'application Adepto.

Sélection d'un fichier à dépouiller

Le nom du ou des fichiers analysés est inscrit automatiquement dans le champ «Rédaction» (même si celui-ci est fermé) à la suite des noms de fichiers qui peuvent s'y trouver déjà. L'utilisateur peut maintenant exploiter le résultat de l'analyse.

En cliquant sur le bouton «Rédaction de fiches» (bouton 1 de la figure 7), l'utilisateur ouvre le champ «Rédaction» (champ 3) dans lequel sont déjà inscrits les noms de fichiers qu'il vient de faire analyser et qui constitue un répertoire de noms de fichiers à dépouiller; le bouton 4 permet d'ajouter à ce répertoire d'autres noms de fichiers. Un double clic sur le nom du fichier à dépouiller ouvre un menu flottant (champ 2) dans lequel l'utilisateur choisira l'article «Sélectionner» pour avoir accès à l'environnement de rédaction; l'article «Supprimer de la liste» permet d'enlever du répertoire le nom

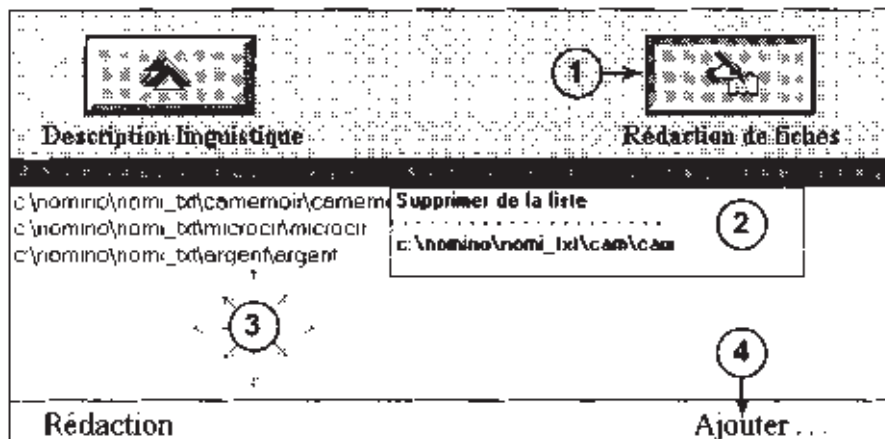


Fig. 7 – Sélection d'un fichier à dépouiller

d'un fichier (sans le détruire sur le disque).

2.1 Environnement de gestion des sources

Si le fichier sélectionné est différent du dernier fichier dépouillé, avant d'avoir accès à l'environnement de rédaction de fiches, l'environnement de gestion des sources s'ouvrira alors, offrant à l'utilisateur la possibilité d'attribuer au texte choisi un code de source ou d'en sélectionner un (fig. 8).

Le répertoire des sources (champ 1) comprend les codes de sources représentant les documents du corpus de dépouillement. En cliquant sur le bouton «Créer» (2), on a accès à une fenêtre de saisie de la source, laquelle pourra être modifiée (bouton 3) ou supprimée (4). Si l'on veut qu'une source soit attribuée aux données que l'on inscrira sur fiches (termes et contextes), on sélectionne une source dans le répertoire et l'on clique sur le bouton 5; cette source apparaîtra dans le champ «Source courante» et sera inscrite sur fiches jusqu'à ce que l'on sélectionne une autre source. Pour avoir accès à l'environnement de rédaction de fiches, on clique sur le bouton 6.

2.2 Environnement de rédaction des fiches

L'environnement de rédaction de fiches est l'environnement principal du poste de travail dans lequel l'utilisateur dépouille le texte qu'il a préalablement sélectionné, crée de nouvelles fiches ou enrichit les fiches de la base. Cet environnement se présente comme suit: (voir figure 9)

Utilisation

La première opération que l'utilisateur effectuera sera de choisir un «point de vue» à faire afficher (bouton 11 de la dernière figure) selon un critère fréquentiel donné (boutons 12 et 13: opérateur et argument); dans l'exemple illustré à la figure 9, la liste d'unités affichée est celle des *ucn* dont la fréquence est supérieure à 4, c'est-à-dire qui ont cinq occurrences et plus.

En sélectionnant une unité de la liste affichée (en faisant un double clic sur cette unité), l'utilisateur peut prendre connaissance de chaque contexte de cette unité. Si, dans la base de fiches, une fiche existe déjà comportant la même unité (selon la même graphie), qu'elle ait été rédigée à partir du texte en cours de dépouillement ou à partir d'un ou de plusieurs autres textes, cette fiche est disposée en arrière-fond de l'écran, prête à recevoir de nouvelles données (comme la figure 9 l'illustre).

Le premier contexte de l'unité sélectionnée est affiché dans la zone «contexte»; pour passer au contexte

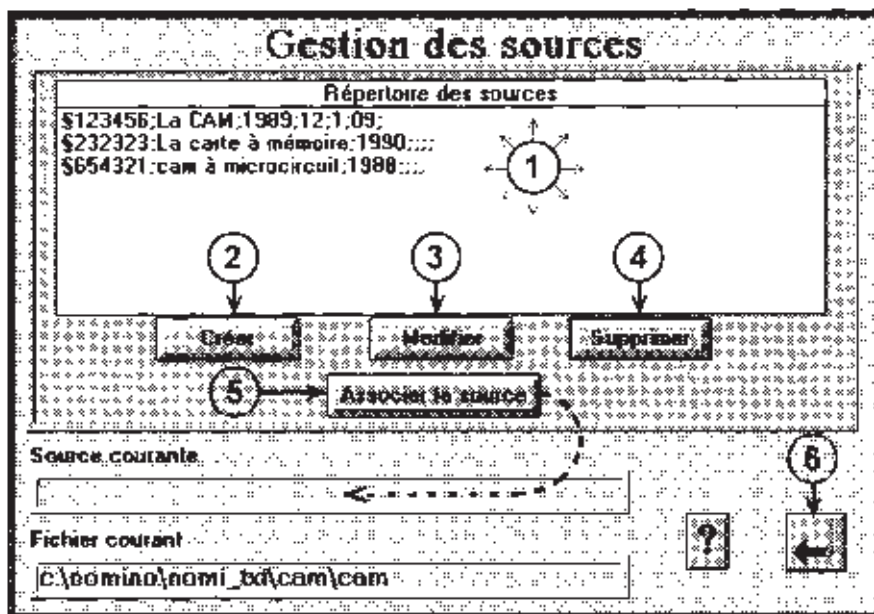


Fig. 8 – Environnement de gestion des sources

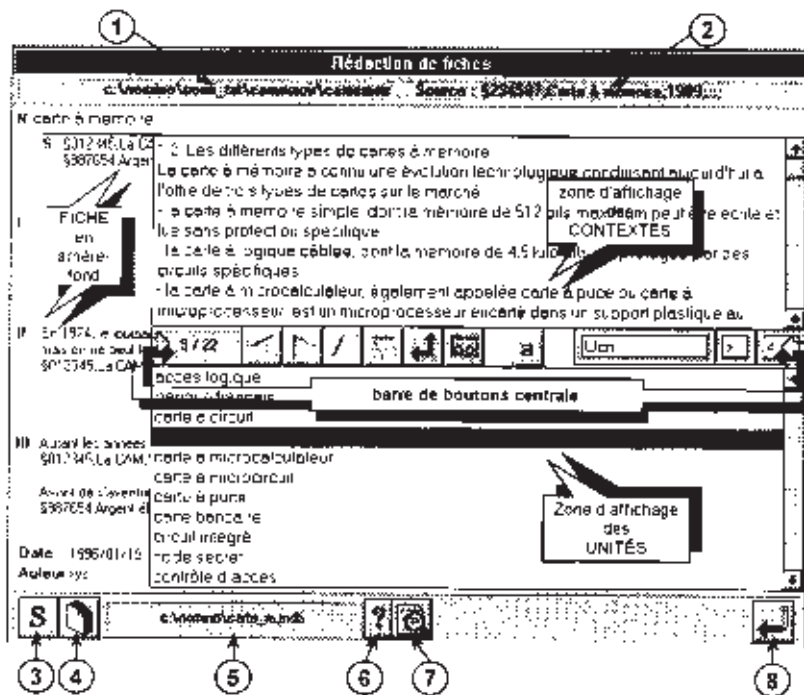


Fig. 9 – Environnement de rédaction de fiches

1 - Nom du fichier en cours de dépouillement. 2 - Source active associée au fichier courant. 3 - Accès à l'environnement de gestion des sources. 4 - Accès à l'environnement de gestion des fiches. 5 - Nom de la base courante. 6 - Aide. 7 - Visualisation complète de la fiche. 8 - Retour à l'environnement de départ.

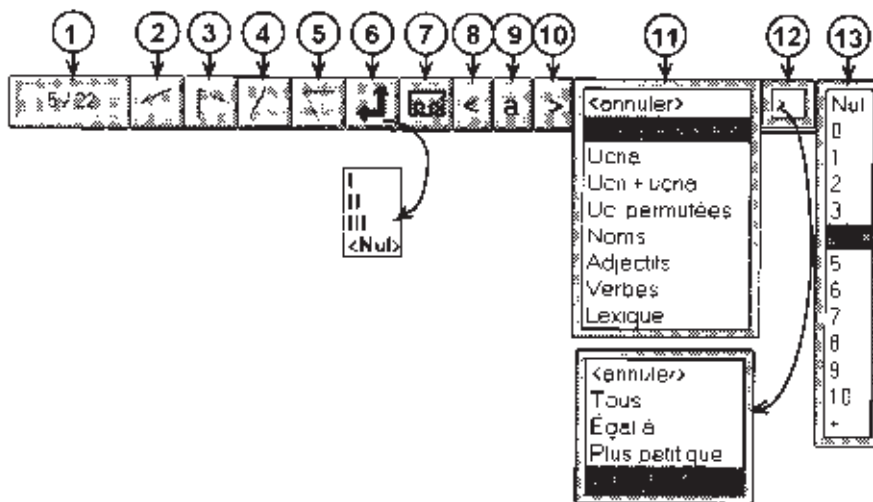


Fig. 10 – Barre de boutons centrale

1 - Numéro du contexte affiché/nombre total de contextes. 2 - Affichage du contexte précédent. 3 - Affichage du contexte suivant. 4 - Allongement du contexte affiché (phrase précédente). 5 - Allongement du contexte affiché (phrase suivante). 6 - Sélection du champ I, II ou III de la fiche dans lequel sera recopié le contexte affiché. 7 - Enregistrement de la fiche. 8 - Affichage du bloc précédent de la liste d'unités. 9 - Recherche d'une unité de la liste. 10 - Affichage du bloc suivant de la liste d'unités. 11 - Sélection d'un « point de vue ». 12 - Sélection d'un opérateur (=, <, >). 13 - Sélection d'un « argument » (chiffre représentant une fréquence).

suivant, on clique sur le bouton 3 de la figure 10; le fait de cliquer sur le bouton 2 permet de revenir au contexte précédent.

Les contextes étant affichés une phrase à la fois, l'utilisateur peut les allonger d'une ou de plusieurs phrases précédentes (bouton 4 de la figure 10) ou suivantes (bouton 5 de cette même figure).

Il peut faire recopier le contexte tel qu'il a pu l'allonger en cliquant sur le bouton 6 et en sélectionnant, dans le menu flottant qui s'est ouvert, le champ I, II ou III de la fiche dans lequel sera transcrit automatiquement le contexte ⁽¹⁵⁾. Si une source a été associée au fichier courant, celle-ci s'inscrit à la suite du contexte (avec l'indication de page si le texte analysé comportait cette indication).

On peut modifier la fiche en cours de rédaction sans avoir à appeler l'environnement de gestion des fiches: un double clic sur le champ à modifier «ouvre» ce champ en fenêtre d'édition, ce qui permet l'affichage de tout son contenu et la saisie des modifications voulues.

Lorsqu'il a extrait du texte toutes les données qu'il jugeait pertinentes à l'unité traitée, l'utilisateur peut enregistrer sa fiche (même dans le cas où il n'a pas recopié de contexte; seul le terme, alors, fera l'objet de la fiche avec, éventuellement, sa source); il lui suffit de cliquer sur le bouton 7. Lors

du premier enregistrement d'une unité sur fiche, il peut modifier la graphie de l'entrée (par exemple la graphie lemmatisée de l'unité complexe *carte à logique câblé* figurant dans la liste des *ucn* n'étant pas la graphie usuelle de ce terme, l'utilisateur pourra préciser la graphie qu'elle portera dans le champ «Entrée» de la fiche: *carte à logique câblée*).

On peut rechercher une unité, dans la liste, en appelant la fonction de recherche (bouton 9) et en inscrivant dans la fenêtre qui s'ouvre les premiers caractères de cette unité (ou tous). Les listes d'unités étant souvent longues, elles ne peuvent être affichées en entier dans la zone d'affichage des unités; elles sont alors découpées en «blocs» d'environ cinquante unités; pour consulter le «bloc» suivant, on utilise le bouton 10; le bouton 8 permet de revenir au bloc précédent.

Les points de vue «unités complexes»

Ces points de vue regroupent, comme nous l'avons vu précédemment, des unités complexes nominales de deux types: les *ucn* et les *ucna*. Ces listes d'unités sont disponibles séparément (points de vue *ucn* et *ucna*) ou fusionnées en une seule liste: points de vue *ucn + ucna* et *uc. permutées*. Dans ce dernier cas, chaque expansion des *ucn* ou des *ucna* est placée au début comme suit:

magnétique: carte à piste magnétique

piste: carte à piste magnétique
toutes les expansions d'une unité complexe (même si elles ne sont pas des *noms*) sont ainsi mises en relief dans cette liste. Il est donc possible de retrouver toutes les occurrences du mot *carte* comme composante d'une unité complexe soit en position de *tête* (par les points de vue *ucn*, *ucna* et *ucn + ucna*) soit en position d'*expansion* (par celui des *uc. permutées*).

Pour les terminologues, ces listes représentent un grand intérêt; c'est leur fournir la liste de la plupart des termes potentiels compris dans le texte, un index des termes du texte, en quelque sorte, avec la possibilité de retrouver les contextes de chaque unité.

Choix de la fréquence

Bien que la fréquence élevée d'une unité complexe ne constitue pas, en elle-même, une preuve de l'importance, dans un texte, de la notion représentée par cette unité, le fait d'entreprendre la «lecture» d'un texte à partir des unités ayant un certain nombre d'occurrences peut aider à y retrouver quelques-unes des principales notions du texte ⁽¹⁶⁾. Aussi, peut-il être utile, lorsque l'on ne connaît d'un texte que son titre, d'en commencer l'exploration par quelques-unes des unités parmi les plus fréquentes, d'évaluer le réseau lexical auquel appartiennent ces unités. Toutefois, on ne saurait se limiter aux unités fréquemment utilisées; il se peut, dans une perspective de veille terminologique, que les unités ayant peu d'occurrences soient celles-là mêmes que l'on veuille recenser. L'utilisation de la fréquence des unités vise l'élimination d'une bonne part du bruit; il existe d'autres moyens de restreindre ce bruit sans que cela n'entraîne une augmentation de silence; en se dotant d'une méthode d'exploitation des données, l'utilisateur pourra à la fois résoudre ces problèmes de bruit et de silence et retrouver les unités les plus importantes du texte. Nous allons exposer très brièvement l'une de ces méthodes.

Contrôle du bruit et du silence

Pour retrouver dans un texte ou un corpus de textes des unités que l'on ne connaît pas au départ, la lecture linéaire des textes peut constituer le moyen le plus long ⁽¹⁷⁾;

(15) Il revient à l'utilisateur de définir ces champs I, II et III et de respecter cette définition pour préserver la cohérence de sa base de fiches.

(16) Ce nombre est d'ailleurs relatif et dépend largement de la longueur du texte, du taux de répétitivité des unités du texte (lequel peut varier d'un type de texte à l'autre), etc.

(17) Il faut, bien entendu, prendre en compte la longueur des textes, la difficulté que représente le fait de ne pas disposer d'une version électronique du texte, etc.

le fait de parcourir un index du texte permet de l'appréhender de façon plus rapide et mieux dirigée.

Toutefois, ces listes d'unités produites automatiquement sont longues, ne comportent pas que de vrais termes et parmi ceux-ci beaucoup ne présentent pas d'intérêt (soit parce qu'ils n'appartiennent pas au domaine de recherche du terminologue, soit parce qu'ils ne concernent pas en propre le sujet traité par l'auteur, soit parce que, dans un projet de veille terminologique, ces termes sont déjà connus). Par ailleurs, le résultat du dépistage de termes, si riche soit-il, ne comporte pas tous les termes du texte (ou ne les présente pas tous selon le découpage le plus adéquat).

Un texte analysé par Nomino constitue une base de données virtuelle en ce sens que chaque unité dépistée a été extraite du texte et renvoie à ses contextes; toutes les unités ne feront pas l'objet d'une fiche, mais toutes sont disponibles. De même que l'on ne lit pas, généralement, un dictionnaire du début à la fin, on ne parcourt pas une base de données en son entier, on l'exploite selon des besoins bien précis et l'on s'en tient généralement aux données qui nous importent. Dès lors que l'on ne connaît pas à l'avance ce que l'on recherche, les «index» du texte, ces listes d'unités extraites de ce texte, constituent le point de départ du dépouillement.

Dans l'enchevêtrement du réseau global que constitue le texte, il faut trouver un fil conducteur qui permette de démêler un ou des sous-réseaux précis. Pour ce faire, relevons, à partir des listes d'unités complexes extraites d'un texte, des réseaux notionnels partiellement représentés par des familles lexicales. Admettons un texte qui traite de la *carte à mémoire*. Cette unité fait partie d'un réseau englobant qui est *carte* et dont chaque unité suivante peut relever: *carte à logique câblé* ([...])

carte à mémoire
carte à mémoire simple
carte à micro-processeur
carte à piste
carte à piste magnétique
carte à puce
carte au laser
carte EEPROM
carte embossé
carte EPROM
carte magnétique
carte multi-services
carte optique au laser

Ce réseau, purement formel peut comprendre plus d'un réseau sémantique: la carte «support d'information», la carte «composante matérielle de l'ordinateur», etc. (sans parler des homonymes de *carte* dans des domaines aussi différents que la cartographie, la restauration, etc.). En ignorant les non-termes, selon que l'on travaille dans un domaine précis, on éliminera les unités non pertinentes de son champ de recherche (et, dans les cas de doute, l'accès immédiat aux contextes permettra souvent de déterminer si une unité appartient ou non au domaine).

C'est ainsi que l'on peut constituer un réseau dont on étudiera

chaque nœud représenté par une unité; par exemple, dans le réseau *carte*, le nœud *carte à mémoire* conduit aux unités *carte à mémoire* et *carte à mémoire simple*. Ce sous-réseau peut être complété d'unités obtenues en consultant le point de vue *Uc. permutées: mémoire: supercarte à mémoire* cette dernière unité qu'on n'avait pu relever à partir de la liste des différentes cartes a pu être relevée grâce à l'expansion commune qu'elle partage avec *carte à mémoire*.

Si la liste des unités peut rapidement informer sur la présence d'un réseau et sur son importance dans le texte, les contextes des unités de ce réseau aideront à confirmer l'existence et la pertinence de ce réseau (qui, jusqu'à présent, n'a été construit qu'à partir d'une famille lexicale) et à le parfaire. Les contextes peuvent permettre de découvrir, d'établir ou de confirmer des relations entre une unité et d'autres unités, relevant ou non de la même famille, ou entre une notion désignée par l'une des unités du réseau lexical et d'autres notions n'appartenant pas à ce même réseau lexical:

– Relations de nature linguistique entre une unité et son abréviation ou sa forme abrégée, entre cette unité et

Contextes

carte à mémoire → CAM

Autant les années 1980 ont été marquées par le micro-ordinateur, autant les années 1990 seront marquées par la *carte à mémoire* (CAM).
(La Salle 1990: 16)

carte à mémoire →
carte à puce

En 1974, le journaliste français Roland Moréno, président d'Innovatron, dépose un brevet sur un objet portable contenant des informations numérisées et un moyen inhibiteur pour les protéger, c'est la *carte à puce* ou à *mémoire*.
(La Salle 1990: 18)

carte à circuit intégré →
integrated circuit card

Messages entre la *carte à circuit intégré* et le dispositif de traitement («messages between the integrated circuit card and the card acceptor device»): concepts et structures, fonctions et messages, types de messages, données communes, éléments de données.
(Guez, Robert, Loret 1988: 176)

– Relations internotions hiérarchiques (de type générique/spécifique ou de type partitif) ou séquentielles ou autres :

Contextes

<p>carte à mémoire → <u>carte à mémoire simple</u> <u>carte à logique câblée</u> <u>carte à microcalculateur</u> <u>carte à puce</u> <u>carte à microprocesseur</u></p>	<p>La carte à mémoire a connu une évolution technologique conduisant aujourd'hui à l'offre de trois types de cartes sur le marché : – la <u>carte à mémoire simple</u> ([...]) – la <u>carte à logique câblée</u> ([...]) – la <u>carte à microcalculateur</u>, également appelée <u>carte à puce</u> ou <u>carte à microprocesseur</u> ([...]). (Ganne, Salomoni 1990: 3)</p>
<p>CAM → <u>dossier portable</u></p>	<p>Autre application de la CAM: le <u>dossier portable</u>. On transporte sur soi, dans une carte, son dossier médical, scolaire ou autre. (La Salle 1990: 21)</p>

une variante, un synonyme, un équivalent anglais, etc.

Plusieurs avantages ressortent de cette méthode : le fait de cerner les termes ou les notions les plus importants, quant au sujet abordé par l'auteur du texte avec une certaine rapidité, mais également de diminuer considérablement l'effet de « bruit » des listes d'unités (particulièrement des listes d'unités complexes) : on ne consulte les contextes que des seules unités utiles, c'est-à-dire les unités faisant partie d'un réseau ou entretenant explicitement des relations. Enfin, les contextes de certaines de ces unités peuvent comprendre des termes qui auraient pu ne pas être dépistés ou qui ne seraient pas découverts convenablement (ce qui permet de compenser le « silence » partiel de la liste des unités).

Attribution à une unité des contextes d'une autre unité

Il est possible d'ajouter aux contextes d'une unité ceux d'une ou de plusieurs autres unités. On voudrait, par exemple, attribuer à carte à logique câblée les contextes de l'unité mal orthographiée carte à logique cablée (comme dans l'exemple de la figure 11). Pour ce faire, après avoir sélectionné l'unité carte à

logique câblée, on peut récupérer les contextes de l'unité mal orthographiée en la sélectionnant et en choisissant, dans le menu flottant, l'article « Union »⁽¹⁸⁾ ; les trois contextes de carte à logique cablée s'ajouteront aux six contextes de l'unité courante, sauf si les deux unités ont des contextes en commun (voir l'exemple illustré de la figure 11). Une « union » peut ainsi être établie entre une unité et un ou plusieurs synonymes, variantes ou abréviations, etc., et même entre des unités relevant de différents « points de vue ».

2.3 L'environnement de gestion des fiches

Les fiches créées ou mises à jour dans le module de rédaction de fiches sont enregistrées dans la base de fiches que l'utilisateur a ouverte dans l'environnement de départ et dont le nom est inscrit dans le champ 1 de la figure suivante. Elles peuvent être consultées et modifiées dans l'environnement de gestion des fiches qui se présente ainsi :

La fiche (ou, plutôt, sa structure) que l'on voit dans cet environnement est la même que celle qui était partiellement affichée en

arrière-fond de l'environnement de rédaction de fiches ; seul le champ « E (ntrée) » n'était pas visible dans ce dernier environnement. Ce champ comprend l'unité telle qu'elle a été modifiée (ou acceptée) par l'utilisateur ; c'est la graphie usuelle, non lemmatisée, qui devrait apparaître dans ce champ, telle qu'on l'inscrirait dans la rubrique d'un lexique ou d'un vocabulaire ou dans un champ « entrée » d'une fiche de terminologie. Cette graphie peut être modifiée à tout moment dans l'environnement de gestion des fiches.

Le champ « N (om) » comporte l'unité lemmatisée telle qu'elle apparaissait dans la liste des unités relevées par Nomino ; ce champ sert à établir un lien entre l'unité qu'il comprend et toute occurrence de la même unité comprise dans l'une ou l'autre liste d'unités dépistées par Nomino dans n'importe quel texte ; si une fiche portant au champ « N (om) » l'unité : carte à logique câblée, à chaque fois que l'on sélectionnera, dans l'environnement de rédaction de fiches, cette même unité dans la liste des ucn de n'importe quel texte, la fiche est ramenée de la base, prête à recevoir de nouvelles données. La comparaison entre les unités d'une liste ne peut être établie avec celles qui sont inscrites au champ « E (ntrée) », ces dernières pouvant différer des unités lemmatisées (ex. : carte à logique câblée).

La fiche comprend plus d'informations qu'il est possible d'en voir : si chaque champ dispose d'un minimum d'espace, toutes leurs données ne sont pas visibles lorsque l'on consulte la fiche. Pour voir la totalité des données d'un champ ou pour modifier ces données, il suffit de faire un double clic sur le champ lui-même : le champ s'« ouvre » en fenêtre

(18) L'article « Union » n'est disponible que si une unité est préalablement sélectionnée.

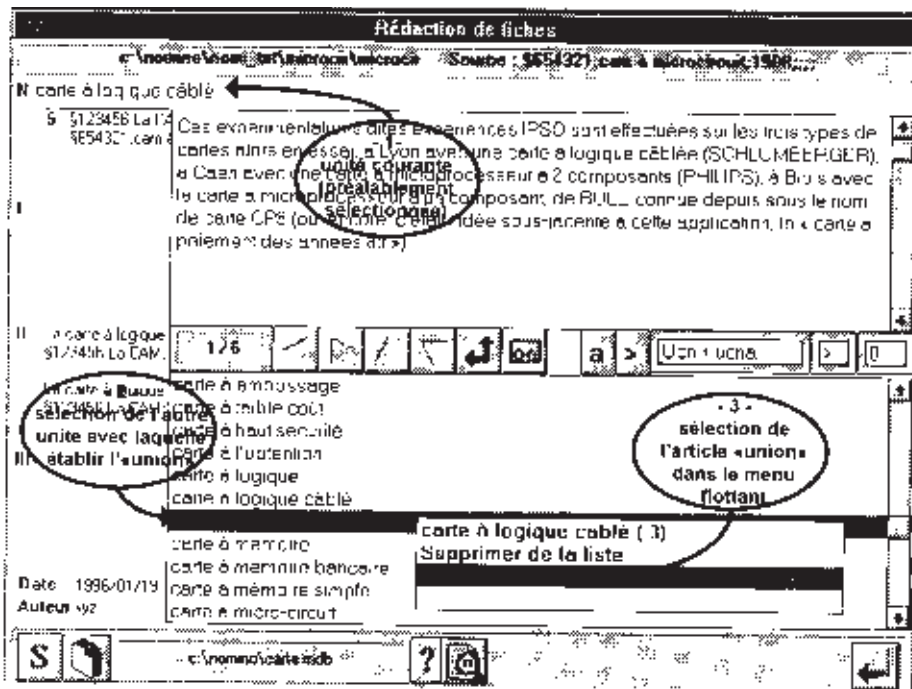


Fig. 11 – Attribution à une unité des contextes d'une autre unité (« union »)

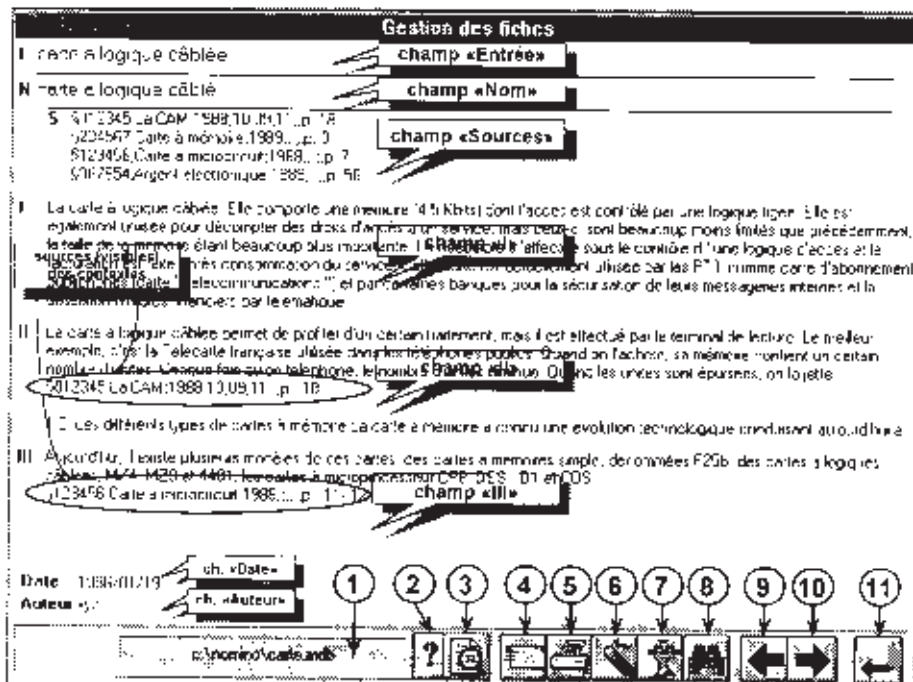


Fig. 12 – Environnement de gestion des fiches

1 - Nom de la base. 2 - Aide. 3 - Visualisation complète de la fiche. 4 - Importation/exportation. 5 - Impression. 6 - Suppression de fiches. 7 - Tri des fiches. 8 - Recherche. 9 - Fiche précédente. 10 - Fiche suivante. 11 - Quitter l'environnement.

d'édition. Le bouton 3 (que l'on retrouvait dans l'environnement précédent) permet de voir toute la fiche, sans possibilité de modifier le contenu de la fiche toutefois.

On peut exporter (bouton 4) les fiches de la base pour les récupérer dans une autre base Adepto ou les exporter vers un autre gestionnaire de bases de données; c'est ainsi que le résultat du dépouillement peut être intégré dans les fichiers d'une banque de terminologie. Le même bouton permet également d'importer des fiches d'une autre base Adepto ou produites à l'aide d'un autre gestionnaire à la condition que le format de ces fiches soit compatible avec celui des fiches Adepto.

Des fonctions générales d'impression (bouton 5), de suppression de fiches (bouton 6) sont disponibles. On peut commander le tri (bouton 7) des fiches par ordre alphabétique du contenu du champ «E (ntrée)».

Le bouton 8 permet la recherche dans les champs «E (ntrée)» et «N (om)»; les flèches 9 et 10 permettent de se déplacer de fiche à fiche. Pour quitter l'environnement de gestion des fiches, on utilise le bouton 11.

3 Conclusion

Le système d'aide au dépouillement terminologique Adepto-Nomino que nous avons présenté en est à sa toute première version; compte tenu de la qualité de ses résultats, de la rapidité de ses traitements, de sa capacité à traiter des textes de fort volume et de sa simplicité d'utilisation, il peut efficacement assister les terminologues dans leurs travaux de dépouillement ou de veille terminologique. Cet outil pourrait être enrichi de nouvelles fonctions qui faciliteraient davantage le travail des terminologues. Ces fonctions les plus immédiatement utiles seraient le développement d'une fonction d'exclusion consistant à éliminer (ou tout au moins à marquer) les unités relevées dans un texte qui seraient comprises dans un corpus d'exclusion⁽¹⁹⁾, la création d'un «point de vue» *réseau* qui fournirait la liste de toutes les têtes (ou bases) des unités complexes et dont le nombre d'occurrences correspondrait à l'ensemble des unités construites à partir de cette tête⁽²⁰⁾. L'intégration d'analyseurs d'autres langues que le français, le traitement des multiples contextes des unités seraient également des développements importants à apporter au système de dépouillement et qui favoriseraient la veille terminologique.

Toutefois, un outil, aussi adapté puisse-t-il être au travail pour lequel il a été conçu et aussi fiables que puissent être ses traitements, ne saurait nous faire ignorer l'importance de certains autres facteurs déterminants pour la bonne conduite d'un projet de veille terminologique. Une méthode appropriée d'utilisation de cet outil, particulièrement lorsque le volume de données à traiter est imposant et que le résultat du travail doit être rapidement disponible, est une condition qui jouera sur l'efficacité du travail.

Un autre facteur du succès de projets de veille terminologique est la détermination des domaines dans lesquels on établira cette veille: ces domaines doivent comporter, avec une certaine abondance, des terminologies nouvelles ou trop récentes pour avoir été répertoriées, sans quoi les travaux de veille ne permettront de recueillir que peu de termes nouveaux; ces domaines doivent occuper des secteurs d'activités humaines jouant un rôle prépondérant ou dont la connaissance et la propagation des terminologies constituent un enjeu important pour les objectifs visés par la veille que l'on veut mener, sans quoi les travaux de veille fourniront des résultats qui seront peu utiles.

La détermination de ces domaines ne saurait se faire sans une veille documentaire préalable, car elle participe à cette nécessaire planification des travaux de veille. La veille documentaire devrait permettre de recenser la production terminologique mondiale, ce à quoi le Rint, par ses inventaires de travaux terminologiques, contribue; elle devrait permettre aussi de dresser un bilan global de l'état de développement des terminologies dans différents domaines afin d'éviter d'entreprendre des travaux réalisés ou en cours de réalisation ou afin d'indiquer les domaines peu couverts. Cette recension devrait prendre en compte la production terminologique anglophone non seulement afin de faire état de ce qui se produit en anglais mais, par comparaison à la production issue de milieux francophones, afin de faire ressortir l'absence de répertoires terminologiques en français ou afin de déceler des terminologies françaises lacunaires (et d'orienter, ainsi, les travaux de veille terminologique). Il est justifié de tenir compte des langues de ces répertoires terminologiques produits par des non francophones: un vocabulaire

(19) Le développement d'une fonction d'inclusion devrait également être possible: seules les unités dépistées figurant dans un corpus d'inclusion seraient retenues ou marquées.

(20) Ce qui permettrait de trouver rapidement les familles lexicales les plus importantes en nombre; à partir de ce «point de vue» *réseau*, lors de la sélection d'une tête, se déploieraient, en cascade, l'ensemble des unités construites sur cette tête.

unilingue anglais, par exemple, *pourrait* comprendre une terminologie non développée en français. Outre les publications terminologiques, il faudrait tenir compte des publications techniques et scientifiques afin de déterminer les domaines ou disciplines qui sont couverts par ces publications ⁽²¹⁾, et, plus spécifiquement, ces domaines ou disciplines qui se démarquent par l'abondance soudaine de la documentation produite; on devra prêter attention à différents facteurs dont celui de la langue dans laquelle cette documentation est rédigée ⁽²²⁾.

Nous ne pouvons, ici, qu'illustrer brièvement l'importance de la veille documentaire et signaler que cette veille ne consiste pas qu'à dresser un inventaire de la production terminologique ou de la production d'information, ou à fournir la documentation utile aux

(21) Nous étendons le sens de «publications» à toute la documentation disponible sous divers formats, sur divers supports. Une sélection de cette documentation s'impose; on s'intéressera particulièrement à cette documentation la plus proche possible des lieux de production de biens et de services, des milieux de recherche et de développement, bref à la documentation la plus recherchée par les acteurs eux-mêmes des milieux commercial, technique et scientifique œuvrant dans un domaine donné.

(22) L'absence de publications françaises dans un domaine donné peut être un indice de l'absence de terminologies françaises et, compte tenu de l'importance du domaine, de l'enjeu que représente le développement et la diffusion de la terminologie française dans ce domaine, pourrait faire ressortir l'intérêt de développer un projet de veille dans ce domaine.

(23) Y compris la documentation électronique, qu'il s'agisse de répertoires terminologiques ou de documents techniques ou scientifiques (publiés ou non, disponibles sur disques optiques ou dans des bases interrogeables, etc.).

terminologiques ⁽²³⁾; l'analyse qui sera faite de cette documentation est essentielle à la définition des projets de veille terminologique.

Jean Perron,
Office de la langue française,
Québec.

Bibliographie

David (Sophie) et Plante (Pierre), 1991: «Le progiciel Terminio: de la nécessité d'une analyse morphosyntaxique pour le dépouillement terminologique des textes» dans *Les Industries de la langue, perspectives des années 1990, Actes du colloque (21-24 nov. 1990)*, Québec, Office de la langue française et Société des traducteurs du Québec, tome I, p. 71-88.

Dumas (Lucie), Plante (André) et Plante (Pierre) 1995: Nomino (document disponible sur le site Web <http://www.ling.uqam.ca/Ato/FX/AtelierFX.html>).

Perron (Jean), 1989: «Termino: un système de dépouillement terminologique» dans *Terminogramme*, n° 54, Québec, Office de la langue française; repris dans la *Revue ICO*, vol. 2, n° 5.

Plante (Pierre), 1996: *L'Atelier FX, version 7 L'environnement de programmation* (document disponible sur le site Web <http://www.ling.uqam.ca/Ato/FX/AtelierFX.html>).

Plante (Pierre) et Dumas (Lucie), 1995: *Nomino Signet Un assistant au dépouillement terminologique* (document disponible sur le site Web <http://www.ling.uqam.ca/Ato/FX/AtelierFX.html>).

Les exemples sont tirés des ouvrages suivants:

Ganne (Roger) et Salomoni (Brigitte), 1990: *La Carte à mémoire*, Paris, Eyrolles.

Guez (Fradji), Robert (Claude) et Lauret (Annette), 1988: *Les Cartes à microcircuit*, Paris, Masson.

La Salle (Michel), 1990: «La carte à mémoire s'utilise partout, par tous» dans *Informatique & bureautique*, Montréal, numéro de mars.

Filtact[©] : un automate d'extraction des termes complexes

Filtact est un prototype en développement d'automate d'extraction de termes complexes. La version présentée à cette table ronde est fonctionnelle, tout en comportant un certain nombre de problèmes en voie de solution pour son optimisation. Le prototype a été conçu dans le cadre du Ciral (Centre international de recherche en aménagement linguistique) et de notre projet CRSH 1994-1997 (Conseil de recherche en sciences humaines) de description lexicale des textes de LSP. La structure générale de l'algorithme d'extraction a été conçue par Alain Auger, Patrick Drouin et Pierre Auger et le programme a été écrit par Patrick Drouin sous Delphi (langage de programmation orienté objet qui dérive de Pascal). C'est à ce dernier également que revient le crédit de l'interfaçage pour Windows 3.1/3.11/95. Le prototype fonctionne sous tout PC évolué (processeur 386 et +, 4 Mo de mémoire vive).

Termes-clés :
Filtact; base de données
textuelles; unité terminologique
complexe; segment.

L'algorithme de Filtact est bâti autour de l'hypothèse que les termes complexes peuvent être repérés en isolant, puis en filtrant les segments complexes (2 à 9 mots) répétitifs d'une base de données textuelles. Pour réaliser cet objectif, les textes de LSP sous format Ascii sont indexés par le programme *makbas.exe* de l'analyseur textuel Tact (versions 1.2/2.1), le résultat de ce premier traitement est la production d'une base de données textuelles indexée (fichier *.tdb).

Dans un deuxième temps, le fichier *.tdb est traité par un programme auxiliaire de Tact, le générateur de collocations *collgen.exe* qui isole tous les segments complexes (2 à 9 mots) répétitifs apparaissant dans la base de données textuelles. Le programme *collgen.exe* produit un fichier Ascii suffixé en *.col où les « collocants » en ordre alphabétique sont accompagnés des segments où ils apparaissent avec l'indice de fréquence pour chaque segment.

Le fichier *.col généré par le programme *collgen.exe* est repris par Filtact qui va filtrer (i.e. inclure ou exclure) les segments isolés lors de l'établissement des collocations. Pour ce faire, Filtact va, dans un premier temps, catégoriser par parties du discours toutes les unités lexicales simples qui commencent et terminent chacun des segments des collocations. Cette catégorisation est effectuée en appliquant les dictionnaires grammaticaux de Sato. Toutes les formes ambiguës (i.e. non reconnues par les dictionnaires) sont marquées de la valeur NIL lors de cette

première étape. Dans un second temps, Filtact va reconnaître toutes les formes complexes selon les valeurs grammaticales sélectionnées préalablement. En réalité, Filtact procède par la négative en écartant les patrons syntagmatiques qui ne sont pas possibles (ou peu probables) pour un TC. Ces patrons sont définis par le terminologie préalablement au filtrage. Par exemple, le logiciel exclut tous les segments qui commencent ou se terminent par un article, une préposition, etc. Ce filtrage s'effectue par exclusion, donc à partir de catégories fermées (prép., conj., etc.) plutôt que par inclusion à partir de catégories ouvertes et ambiguës comme celles du nom, du participe passé, etc. À titre d'exemples, seront retenus les segments :

- Commençant par un NOM et se terminant par un NOM pour repérer les syntagmes du type N + j0 + N; N + jp + N et N + jp + préd + N et N + jp + V + N;
- Commençant par un NOM et se terminant par un ADJECTIF (et inversement) pour repérer les syntagmes des types ADJ + NOM et NOM + ADJ;
- Commençant par un NOM et se terminant par un verbe à l'infinitif pour repérer les syntagmes de forme N + jp + V.

La sélection de ces modèles est effectuée en choisissant comme paramètres de filtrage les catégories : NOM, ADJ, V_INF (+ NIL). L'ajout de la valeur NIL pour retracer les segments ambiguës peut causer toutefois du bruit dans la liste des segments retenus. Les unités complexes filtrées sont enfin enregistrées dans un fichier suffixé en

*.flt, les unités exclues le sont dans un fichier suffixé en *.xlt.

Filtact permet en fait de sélectionner parmi une liste de 39 catégories (valeurs grammaticales) :

abréviation	nombre
adj	nomc
adj_inter	nomp
adv	p_dém
art	p_indef
artdef	p_pers
artgén	p_poss
artind	p_relatif
artpart	punctuation
aux	ppassé
conjonction	pprésent
délim	prép
détdéf	pronom réfléchi
détdém	résidu
détindéf	singulier
détnum	transitif
détposs	v_conj
dentificateur	v_inf
interj	virgule
nil	

Cette flexibilité constitue un atout intéressant pour Filtact, car elle permet de repérer en sus des syntagmes terminologiques complexes tous les modèles de phraséologie qu'on est susceptible de rencontrer dans les textes de LSP. Ainsi, par exemple, les modèles : (NOM + ADJ), (NOM + VERBE), (VERBE + NOM), (VERBE + ADV), (ADV + ADJ) (et (NOM + PRÉP + NOM)). Cette caractéristique s'avérait très intéressante pour notre projet d'atelier de description lexicale complète des textes de LSP.

Pour parler maintenant de l'efficacité de Filtact à reconnaître les unités terminologiques complexes dans un texte de LSP, nous avons pris comme base d'analyse un volume de la collection *Que-sais-je?* intitulé : *Hypertexte et hypermédia* (P.U.F., 199?, 13? p.). Voici quelques données quantitatives concernant ce texte et son traitement :

– Le fichier de départ Ascii a une taille de 237 ko ;

– Le fichier des collocations généré par *collgen.exe* fait 335 ko ;
 – Le filtrage par Filtact du fichier des collocations avec les filtres « nom commun, adj., verbe infinitif, participe passé et nil » a été effectué en 3 minutes environ (machine avec processeur 486 et vitesse d'horloge de 25 Mhz) : 13 466 segments ont été traités, 534 ont été retenus (12 932 exclus) et de ce nombre 203 ont été sélectionnés manuellement comme candidats de termes complexes. Le rendement global tourne donc aux environs de 40% pour l'instant.

Il faut ici signaler deux problèmes qui apparaissent inhérents à l'algorithme que nous avons défini. Le premier problème touche l'exhaustivité des prélèvements effectués sur la base du repérage des seuls segments répétitifs. Les segments uniques étant exclus, le problème augmente avec le traitement de corpus de petite taille et inversement décroît en traitant des corpus de grande taille pour un domaine spécifique. Le second problème est celui du bruit causé par l'absence de désambiguïsation en cours de traitement, par l'attribution de la catégorie NIL à un nombre trop important de lexèmes constitutifs de segments, phénomène normal en raison du caractère d'étrangeté des termes en regard des dictionnaires utilisés qui eux sont plutôt généraux. Ce dernier problème pourra être solutionné en enrichissant constamment les dictionnaires sur le plan terminologique.

Filtact en mode « gestion des résultats » permet de nettoyer commodément la liste des termes filtrés directement à l'écran, en déplaçant d'une fenêtre à l'autre les termes non confirmés. Le choix final des TC peut être confirmé sous Tact en les soumettant à l'épreuve du score d'association (Z score). On utilisera par la suite la liste « nettoyées » des segments filtrés pour interroger la

base de données textuelles sous Tact et repérer en contexte les termes complexes confirmés – concordances/ contextes accompagnées des références topologiques – ce qui va constituer après exportation des données dans un SGBD comme FileMaker Pro, une base de fiches terminologiques. Les termes simples, quant à eux, sont sélectionnés directement sous Tact en utilisant conjointement les indices de fréquence absolue et des antidiCTIONNAIRES pour isoler certaines strates lexicales.

Nous l'avons rappelé en introduction à notre propos, Filtact est encore à l'état de prototype et devra subir encore plusieurs retouches pour améliorer l'efficacité des filtres et réduire le bruit. Nous prévoyons également, dans un avenir rapproché, travailler en étroite collaboration avec l'équipe de Jacques Ladouceur, également rattachée au Ciral, pour la mise au point d'un outil de repérage commun alliant les approches linguistique et statistique.

*Pierre Auger,
 Centre international de recherche en
 aménagement linguistique,
 Université Laval,
 Québec.
 Patrick Drouin,
 Université de Montréal,
 Québec.
 Alain Auger,
 Université Laval,
 Université de Neuchâtel,
 Suisse.*

ANNEXE I

Méthodes d'extraction automatique des TC

I. APPROCHE LINGUISTIQUE

- CATÉGORISATION GRAMMATICALES ET/OU SYNTAXIQUE DES UNITÉS
- APPLICATIONS DE PATRONS DE FOUILLE

II. APPROCHE STATISTIQUE

- FRÉQUENCE D'APPARITION DES SEGMENTS
- CALCUL DE L'ATTRACTION ENTRE LES LEXIES (SCORE D'ASSOCIATION)

III. APPROCHE LINGUISTICO-STATISTIQUE

- CONFORMITÉ STRICTE À DES RÈGLES STRUCTURELLES SYNTAXIQUES
- FRÉQUENCE D'APPARITION DES SEGMENTS
- COEFFICIENT DE VRAISEMBLANCE
- MESURE DE DISTANCE

ANNEXE II

Exemple de fichier de collocations généré par *collgen.exe* (Tact)

humain	2 création d'un --	2 of --
2 esprit --	2 la création d'un --	2 of -- systems
2 l'esprit --	2 dans la création d'un --	2 of -- systems communications
2 de l'esprit --	hyperdocumentation	2 of -- systems communications of
hypercard	4 l'--	2 of -- systems communications of the
2 et --	hyperdocuments	2 of -- systems communications of the
hyperdocument	2 -- accessibles	acm
5 -- électronique	7 d'--	2 eds designing --
2 -- est	4 des --	2 generation of --
2 -- et	2 création d'--	2 generation of -- systems
2 -- 1	2 la création d'--	2 generation of -- systems
10 l'--	hypermedia	communications
2 l'-- est	2 -- systems	2 generation of -- systems
2 l'-- et	2 -- systems communications	communications of
2 même --	2 -- systems communications of	2 generation of -- systems
8 un --	2 -- systems communications of the	communications of the
2 un -- électronique	2 -- systems communications of the	2 generation of -- systems
6 d'un --	acm	communications of the acm
2 d'un -- électronique	2 designing --	2 mandl eds designing --
6 de l'--	2 hypertext --	2 next generation of --

ANNEXE III

Exemple de fichier des TC reconnus généré par Filtact

* Fichier analysé: C:\LOG\LING\FILFACT\HYPER. COL
 * Fichier des segments retenus: C:\LOG\LING\FILFACT\HYPER. FLT
 * Fichier des segments exclus: C:\LOG\LING\FILFACT\HYPER. XCL

* Catégories retenues:
 * adj nomc ppasse v_inf

* Catégories exclues:
 * abréviation adj_inter adv art artdéf artgén artind artpart
 * aux conjonction délim détdéf détdém détindéf détnum
 * détposs identificateur interj nil nombre nomp
 * p_dém p_indéf p_pers p_poss p_relatif ponctuation pprésent
 * prép pronom réfléchi résidu singulier transitif
 * v_conj virgule

* Voir la fin de ce fichier pour les statistiques...

contrainte spécifique
 contraintes techniques
 conversion automatique
 conversion automatique du texte
 conversion automatique du texte en hypertexte
 conversion d'un texte
 conversion d'un texte en hypertexte
 courrier personnalisé
 croiser l'information
 création d'hyperdocuments
 création d'un hyperdocument
 créer des nœuds

* Segments traités: 13466
 * Segments retenus: 534
 * Segments exclus: 12932

Termplus, système d'extraction terminologique

Établir la terminologie propre à un domaine n'est pas chose simple. Pour ce faire, le terminologue doit lire des textes afin de repérer les termes pertinents et trouver pour chacun d'eux les éléments constitutifs de la fiche. Ce travail, plutôt long, est maintenant facilité grâce à Termplus, un système d'extraction terminologique. Il sera donc question, au cours des prochaines pages, de ce logiciel et de son utilité pour le terminologue.

Termes-clés :
repérage de termes ; notions ;
descripteurs ; néologismes.

1 Introduction

Nous avons conçu Termplus dans le cadre de travaux que nous effectuons en gestion informatisée de l'information textuelle et c'est indirectement que nous nous sommes intéressés à l'extraction terminologique. Nos travaux ont d'abord porté sur l'élaboration automatique de listes de descripteurs pour des textes spécialisés et donné naissance au logiciel prototype Notions.

Notre objectif était de mettre au point une méthode d'analyse afin de reconnaître automatiquement les mots qui sont représentatifs du contenu d'un texte, ce que les indexeurs appellent l'analyse conceptuelle. Il fallait ensuite que ces mots soient traduits en un vocabulaire particulier. Notions effectue la première des deux tâches.

Voyons un peu comment Notions procède. À partir d'un texte spécialisé, il extrait une liste de mots simples et de mots complexes qui sont potentiellement représentatifs du contenu du texte. Les mots sont ensuite ordonnés en fonction de la probabilité qu'ils soient représentatifs du contenu. Plus il est probable qu'un mot soit représentatif, plus ce mot est placé haut dans la liste. Finalement, de cette liste seule la tête est conservée, le reste n'étant pas intéressant pour l'indexation documentaire.

Voici les trente premiers mots complexes que nous donne une

analyse effectuée avec Notions sur un texte appartenant au domaine de l'informatique et traitant des multimédias (Aiglin et Scamps) :

cd-rom
cd-photo
séquences vidéo
séquence quicktime
compact disc
micro-ordinateur
vidéo numérique
séquence vidéo
disque dur
format quicktime
présentation multimédia
lecteur de cd
music prose
movie player
macintosh av
conversion quicktime
lecture quicktime
création multimédia
lecteur de cd-rom
temps de calcul
effets de transition
millions de couleurs
effet de transition
techniques de compression
logiciels de présentation
lecteurs de cd
appareils grand public
intelligence artificielle
pcd exchange
screen play

Si pour l'indexation documentaire la plupart des mots retenus en première analyse par Notions ne sont pas intéressants, il en est tout autrement du point de vue de l'extraction terminologique.

2 Termplus

Termplus est une version de Notions adaptée à l'extraction terminologique. Cependant, les

modifications que nous avons apportées à Notions sont tellement importantes qu'un nouveau nom s'est imposé.

Par exemple, en indexation documentaire, la récurrence est un des critères essentiels pour déterminer si un mot est représentatif du contenu. Dans cette perspective, Notions ne pouvait retenir que les mots qui se produisaient au moins deux fois dans un texte. Il devait ainsi exécuter une analyse des segments répétés pour l'aider à isoler les termes complexes.

Pour l'extraction terminologique, on ne peut pas retenir que les mots répétés. Il peut y avoir dans un texte des termes qui ne se produisent qu'une seule fois. C'est d'ailleurs souvent le cas des néologismes. Lorsque nous avons créé Termplus, nous avons donc dû donner un rôle secondaire à l'analyse des segments répétés et ajouter une analyse capable d'isoler les mots simples et les mots complexes qui ne se produisent qu'une seule fois dans un texte.

Puisque Notions n'était destiné qu'à conserver les têtes de listes de mots, il pouvait tolérer, sans trop de conséquence sur la qualité de ses résultats, un taux de bruit assez élevé. D'ailleurs, des tests effectués sur des textes numérisés ont démontré que la correction des erreurs de reconnaissance optique des caractères attribuables à la numérisation n'améliorait pas, de manière significative, la qualité des résultats de Notions. Ce n'est pas du tout la même chose pour l'extraction terminologique. Dans ce cas, le bruit doit être réduit le plus possible sans évidemment provoquer de silence. En travaillant sur Termplus, nous avons donc mis l'accent sur la réduction du bruit.

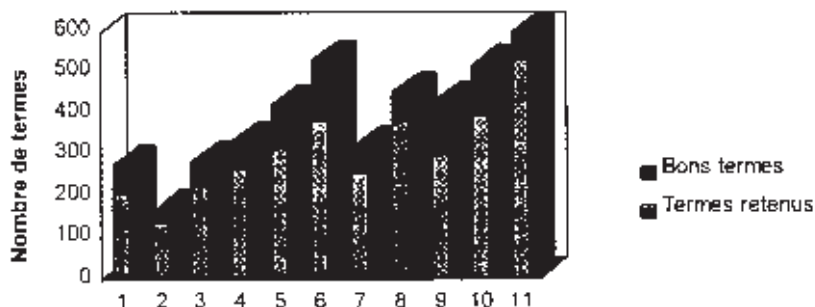
Termplus est un logiciel prototype. Il reste encore plusieurs fonctions à développer. Actuellement, nous disposons de deux versions du logiciel : une version DOS et une

version Windows 95. La version DOS est dotée d'un moteur d'analyse, d'un dictionnaire de verbes et de trois dictionnaires divers qui comprennent environ 2000 mots. Cette version livre ses résultats sous forme de matrice en format Ascii qui peuvent ensuite être récupérés avec un tableur comme Excel. L'utilisateur peut ainsi les réviser facilement et rapidement.

Le moteur d'analyse a été conçu dans une perspective de plurilinguisme. Actuellement, il traite des textes français, mais il peut également traiter d'autres langues sans avoir à subir de modification. Seuls

les dictionnaires doivent être changés. Nous sommes d'ailleurs en train de terminer les dictionnaires anglais et les pré-tests effectués jusqu'à maintenant sont très concluants (voir la section 4 ci-dessous).

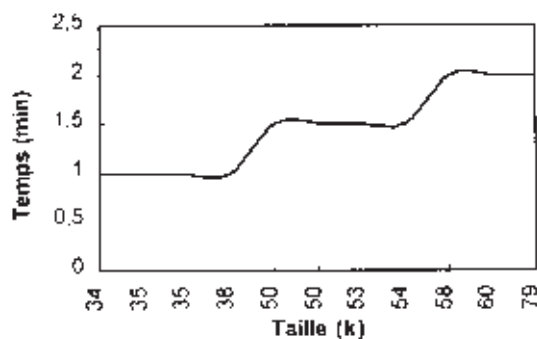
Les performances du moteur d'analyse sont excellentes. Nous l'avons testé sur une série de textes de diverses longueurs et avec des densités terminologiques différentes et dans tous les cas, les résultats sont très bons. Voyons sous forme de graphique, la répartition du nombre de termes que Termplus a relevé pour un corpus appartenant au domaine de l'économie :



Nombre de bons termes par rapport au nombre de termes retenus par Termplus

Il faut dire que les termes potentiels sont ordonnés en fonction de la probabilité qu'ils soient effectivement des termes. La plupart des bons termes tendent donc à se retrouver en tête de liste. L'extraction

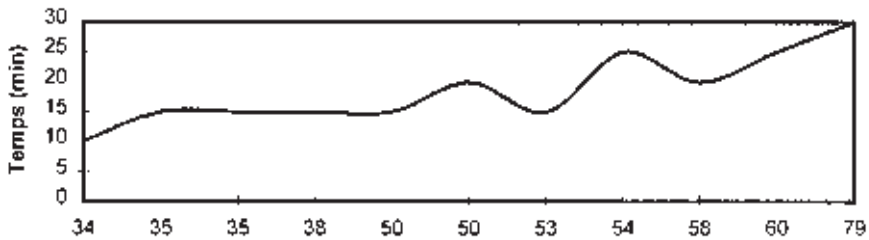
de la majorité des termes d'un texte peut ainsi se faire très rapidement. Toutefois, si on veut éliminer le bruit et ne conserver que les termes d'un texte, il faut alors réviser toute la liste.



Temps de traitement automatique par rapport à la taille des textes

Le temps de traitement automatique augmente de façon régulière avec la taille du texte. L'analyse d'un texte de 100 k, effectuée à l'aide d'un ordinateur 486/66 MHz, prend à peu près deux minutes et demie et l'analyse d'un texte d'un mb. prend environ 25 minutes.

Une fois les textes traités par la machine, l'indexeur humain a épuré les listes de résultats afin de ne conserver que les termes du domaine. Ce travail, effectué à l'aide d'un chiffrier, se fait en une vingtaine de minutes comme le montre la figure ci-dessous.



Temps de traitement manuel par rapport à la taille des textes

Termplus est extrêmement robuste. Jusqu'à maintenant, nous avons traité plus d'une vingtaine de mb. de texte sans aucun problème. Le logiciel produit du bruit mais aucun silence. Nous n'avons intégré aucune technique d'épuration automatique qui occasionne du silence. Nous tolérons l'imperfection lorsqu'elle provoque des résultats erronés, mais pas lorsqu'elle enlève des bons termes. En ce qui concerne le bruit, il est assez bien contrôlé. Le logiciel ordonne les éléments qu'il retient en fonction de la probabilité qu'ils soient des termes. Le bruit tend donc généralement, mais pas d'une façon absolue, à se retrouver en fin de liste et les bons termes en début de liste.

La version Windows 95 de Termplus est presque terminée. Cette version aura le même moteur d'analyse que la version DOS et comprendra en plus une interface de révision. Cette interface permettra à l'utilisateur d'avoir à la fois à l'écran le texte original et la liste de termes potentiels retenus par le logiciel. D'ailleurs, le texte et la liste seront synchronisés. Au fur et à mesure que l'utilisateur naviguera dans la liste de termes, le texte se déroulera afin de

toujours présenter le terme en contexte.

Nous sommes également à parfaire un certain nombre de fonctions d'épurations auxquelles l'utilisateur pourra faire appel selon ses besoins et selon la nature du bruit qu'il y a dans sa liste. Par exemple, nous avons déjà conçu une fonction très simple qui permet d'éliminer tous les termes potentiels qui contiennent des chiffres. Dans certains cas, cette fonction occasionnera du silence en éliminant des termes comme *3-dimension*, mais elle est quand même très utile puisqu'elle évite qu'on ait des choses comme *chapitre 4, figure 1* ou *section 2* dans notre liste. Ce sera donc à l'utilisateur de juger de l'intérêt d'utiliser ou de ne pas utiliser une fonction d'épuration. De toute façon, l'application d'une fonction d'épuration ne modifiera pas la liste de termes potentiels de façon permanente. Ces fonctions agiront plutôt comme des filtres qu'on peut ajouter et enlever à tout moment.

Finalement, nous sommes à mettre au point un éditeur de macro qui permettra à l'utilisateur de combiner et de paramétrer un certain nombre de fonctions d'épuration qu'il

pourra ensuite réutiliser pour l'analyse d'autres textes.

3 Travail du terminologue

Les avantages d'un logiciel comme Termplus sont nombreux quand on regarde de plus près le travail du terminologue. Tout d'abord, le terminologue doit lire des textes dans un domaine précis afin de repérer la terminologie propre à ce domaine, c'est-à-dire les termes simples et les termes complexes. Dans la plupart des cas, il doit ensuite trouver pour chacun de ces termes les principaux éléments constituants de la fiche, c'est-à-dire une définition, un contexte, les variantes orthographiques, les abréviations ainsi que les synonymes. Finalement, il doit repérer les équivalents anglais de ces mêmes notions. Le terminologue doit donc être très attentif lors de sa lecture et il doit très bien connaître le domaine sur lequel il travaille.

La tâche se complique un peu quand le terminologue a comme mandat de repérer uniquement les néologismes d'un domaine. Dans ce cas, il doit faire une sélection afin de retenir certains mots et d'en éliminer d'autres. Deux questions nous viennent à l'esprit quand vient le temps de ces sélections : est-ce que chaque fois que le terminologue trouve quatre ou cinq occurrences d'un même mot, il le laisse tomber et croyant qu'il ne s'agit pas d'un néologisme? Est-ce qu'il en retient une masse considérable juste au cas où il lui en manquerait?

Afin de mieux comprendre ce qui se passe vraiment quand nous lisons un texte à la recherche de néologismes, nous avons fait une expérience. Nous avons dépouillé des textes tirés du domaine des multimédias afin de relever tous les néologismes. Le corpus de textes est composé de sept ouvrages de base,

c'est-à-dire plus de 1 700 pages de textes à lire.

4 Confrontation homme/machine

Nous nous sommes donc demandé s'il fallait vraiment lire tous ces livres. Pour le savoir nous avons confronté l'homme à la machine. L'indexeur humain a donc lu les 85 pages (plusieurs avec des publicités et des photographies) d'un numéro spécial de la revue *Time* qui traitait de la cyberspace. Après cette première tâche qui lui a pris un peu plus d'une heure et demie, l'indexeur n'avait relevé que quelques termes. Nous avons ensuite soumis le même texte à la machine qui a produit une liste de 1 877 mots en 5 minutes.

La liste ci-dessous présente un extrait sommaire des résultats obtenus à partir de la version anglaise.

Comme vous pouvez le constater, cette liste est assez représentative du domaine.

e-mail
meta name
fiber-optic
high-speed
cable-tv
science-fiction
information superhighway
john perry barlow
al gore
city lights
fiber-optic strands
clinton administration
economic growth
hair-thin
science-fiction writer
world wide web
video game
electronic frontier
usenet newsgroups
human system
telephone system

La rapidité du système est stupéfiante. À titre comparatif, la vitesse de lecture d'un lecteur lent est d'environ 20 000 mots à l'heure, celle d'un lecteur moyen est de 28 000,

tandis que celle d'un lecteur rapide est d'environ 60 000. Quant à la machine, elle a fait tout le traitement d'un texte de 111 513 mots (561 pages) en moins de 40 minutes.

Après ce test, remporté par la machine, nous nous sommes demandé si lire les textes valait vraiment la peine à l'étape du dépouillement. Nous avons refait un deuxième test en inversant le processus. Cette fois-ci nous avons inséré chacun des textes dans la machine avant même de les lire et nous avons travaillé sur la liste de termes retenus par le logiciel. Par exemple, pour un texte de 272 pages (Viau et Jolivald) la machine a produit une liste de 4 845 mots complexes en moins de 15 minutes, ce qui aurait été impossible pour l'humain.

5 Fiabilité du système

Pour certains, vite et bien ne vont pas de pair. Dans le cas de Termplus, ce sont deux conditions *sine qua non*. C'est d'ailleurs pour cette raison que nous avons vérifié si la machine n'oublie pas des termes en cours de traitement. Nous avons demandé à un spécialiste de se mesurer au système. Pour ce faire, la personne a lu des extraits de textes et a relevé tous les termes. Ensuite, elle a comparé sa liste à celle produite par Termplus. Tous les termes repérés par l'indexeur faisaient partie de la liste produite par le système. Le système a même marqué un point : l'indexeur avait oublié certains termes que le système avait relevés.

6 Résultats

Lorsqu'on regarde la liste de résultats produite par Termplus pour un seul texte, nous retrouvons souvent plus de 4 000 éléments. Cette liste exhaustive comprend non

seulement des termes, mais aussi des noms propres, des thèmes et du bruit. Étant donné que Termplus n'est doté d'aucun système d'analyse sémantique des phrases, il ne peut déterminer si un mot composé doit être considéré comme un terme ou comme un thème. Ainsi, des choses comme *nouvelles technologies*, *documents existants*, *domaine éducatif*, *ensemble des textes* seront retenues par le système même s'il s'agit plutôt de thèmes.

7 Repérage des néologismes

Nous avons utilisé les listes de termes fournies par le logiciel pour tenter de déterminer à partir de quelle fréquence un terme n'est plus considéré comme un néologisme par le terminologue, c'est-à-dire que le terme fait déjà l'objet d'une ou plusieurs fiches dans la banque de terminologie de l'Office de la langue française.

Prenons, à titre d'exemple, un « Que sais-je ? » (*Textes, hypertextes et hypermédias*). Le logiciel a repéré 3 113 notions (termes, thèmes, noms propres et bruits) répartis comme suit :

Nombre d'occurrences du mot complexe	Fréquence absolue
1	24
1	23
1	22
1	19
2	11
2	9
3	7
8	6
10	5
23	4
62	3
208	2
2791	1

Par contre, l'expérience montre qu'il est plutôt difficile de mesurer la probabilité qu'un terme soit

néologique en ne consultant que la liste des fréquences. Il arrive qu'un terme de fréquence 1 fasse déjà l'objet d'une fiche alors qu'un terme de fréquence 5 n'apparaît nulle part dans les banques. Notre expérience a démontré que la récurrence d'un texte à l'autre est plus importante que la fréquence à l'intérieur d'un seul texte. Ainsi, il ne faut pas croire que tous les termes de fréquence 1 sont des néologismes, puisque, comme nous l'avons mentionné auparavant, certains termes sont des noms propres, d'autres représentent des notions et d'autres encore appartiennent à un autre domaine que celui sur lequel nous travaillons.

8 Utilité

Termplus ne vise aucunement à remplacer le terminologue. Il s'agit plutôt d'un aide qui permet d'accélérer le travail en l'informatisant en partie. Il est ainsi possible pour la personne qui connaît très bien le domaine sur lequel elle travaille de consulter uniquement la liste produite par le logiciel. Comme la liste de termes est synchronisée avec le texte, il est possible en un coup d'œil de repérer l'information nécessaire pour l'élaboration de la fiche. Les avantages d'un traitement semi-automatique sont nombreux :

- Les risques d'erreurs imputables à la fatigue ou au manque de concentration sont nuls;
- Tous les termes (qu'ils soient de fréquence 30 ou de fréquence 1) sont repérés par le logiciel;
- Le traitement est rapide (plus de 200 000 mots à l'heure);
- Le système est convivial puisqu'il tourne sous Windows et qu'il permet à l'utilisateur de changer les paramètres de repérage et d'enrichir les anti-dictionnaires.

9 Possibilité à long terme

Nous venons de terminer la mise au point de la version anglaise de Termplus. Jusqu'à maintenant, les résultats obtenus dépassent nos attentes. Nous avons également eu des demandes pour une version espagnole et une version portugaise du logiciel. Il va sans dire que, dans une optique de terminologie multilingue, les possibilités d'un logiciel comme Termplus sont pratiquement illimitées.

*Jacques Ladouceur,
Guylaine Cochrane,
Centre international de recherche sur
l'aménagement linguistique,
Université Laval,
Québec.*

Bibliographie

- Aiglin (Jacques) et Scamps (Olivier), 1993: Macintosh multimédia, mais c'est très simple, Paris, Dunod.
- Time, 1995: Special Issue : Welcome to Cyberspace, vol. 145, n° 12, Spring.
- Viau (Emmanuel) et Jolivald (Bernard), 1992: Multimédia et CD-ROM, Paris, Sybex.

Le poste de travail du terminologue arabe

L'avènement de la micro-informatique a permis aux chercheurs qui travaillent dans le domaine de la terminologie d'accéder à peu de frais à des banques de terminologie mieux adaptées à leurs besoins. Pour se mettre au diapason, des institutions arabes d'aménagement linguistique ont investi dans la constitution de banques de terminologie. Trois banques qui ont adapté, arabisé ou développé des systèmes de gestion de bases de données pouvant traiter la langue arabe sont présentées ici, soit la Banque de terminologie de l'Académie de Jordanie, la Banque de terminologie de l'Arabie saoudite et la banque Lexar de l'IERA à Rabat (Institut d'études et de recherches pour l'arabisation).

Termes-clés : banque de terminologie ; langue arabe ; terminologie arabe ; harmonisation ; diffusion.

(1) Tout comme les sciences de l'information, la terminologie vise à rationaliser le processus de communication. Par ailleurs, le travail du terminologue, qui consiste à identifier les notions d'un domaine du savoir, à attribuer un terme dans une ou plusieurs langues à ces notions en donnant une définition rigoureuse de chaque notion, rappelle celui du spécialiste de la documentation qui élabore un thésaurus. Cf. River (A), 1990: «Construction des langages d'indexation: aspects théoriques», dans *Le Documentaliste*, vol. 27, n° 6, p. 263-273.

(2) Rapport gouvernemental publié aux États-Unis, en 1963, diffusé par l'ALPAC (*Automatic Language Processing Advisory Committee*). Ce rapport a remis en cause les recherches en TA et les a orientées vers la linguistique computationnelle, la recherche documentaire et la constitution de dictionnaires automatiques.

La terminologie est l'une des disciplines qui ont profité de l'évolution vertigineuse de la technologie informatique. En tant que discipline carrefour elle entretient des rapports étroits avec les sciences de l'information, et ce lien privilégié se retrouve à la fois sur le plan de la visée et sur le plan du fonctionnement ⁽¹⁾. Aussi a-t-elle su exploiter très tôt les SGBD (systèmes de gestion de bases de données) qui étaient conçus au départ pour les besoins de représentation, d'organisation, de stockage et d'accès aux unités de l'information textuelle. En effet, dès les années 65-68, le concept d'une banque de mots sur ordinateur conçue pour fournir aux traducteurs le vocabulaire spécialisé existant dans un domaine de spécialité était né à la suite des recommandations du rapport de l'ALPAC (*Automatic Language Processing Advisory Committee*) ⁽²⁾. Les grandes institutions de traduction ont alors commencé à investir dans la constitution des premières banques de terminologie ⁽³⁾. Les premiers SGBD ne fonctionnaient, cependant, que sur des *main frames*, ce qui rendait l'accès à ces banques ainsi que leur mise à jour des opérations coûteuses. L'avènement de la micro-informatique a changé les données et les SGBD sont devenus disponibles sur micro-ordinateur offrant aux chercheurs, qui travaillent dans le

(3) Dans un recensement établi par Infoterm en 1976, quelques quinzaines de banques de terminologie ont été inventoriées.

domaine de la terminologie, la possibilité de créer et de dégager des banques de terminologie mieux adaptées à leurs besoins et pouvant être mises à jour à relativement peu de frais.

État de la question dans le monde arabe

Pour se mettre au diapason, des institutions d'aménagement linguistique et de mise au point terminologique ont commencé, dès les années 80, à investir dans la constitution de banques de terminologie. Ainsi l'IERA a commencé dès 1979 à stocker ses données à l'ESRIN (Institut de recherche de l'Agence spatiale européenne) par télétraitement en attendant la création de son propre centre informatique en 1984. La Cité Abdul Aziz des Sciences et des Technologies en Arabie saoudite a constitué BASM, sa banque terminologique, dès 1984 grâce à l'acquisition d'une partie des données de la banque de terminologie TEAM. L'Académie de langue arabe de Jordanie a constitué sa banque de termes en 1988. Ces banques de termes ont toutes été constituées sur des gros ou mini-ordinateurs. Le problème qui se pose cependant, quand on veut faire l'état de l'art en matière de banques de terminologie dans le monde arabe, c'est d'une part le manque d'informations et d'autre part le caractère caduque de telles informations. Il n'existe aucun inventaire des banques de

(4) Les propriétés particulières à la langue arabe ont un grand impact sur la conception de toute application en TALN. Ainsi les propriétés morphologiques particulières de l'arabe sont importantes non seulement pour la mise au point de systèmes informatisés d'analyse et de génération, mais aussi pour la manipulation de dictionnaires et notamment leur prise en compte offre une meilleure compression des données. Par ailleurs, pour les questions de codage, cf. Dchiech (M.), 1993: «Réforme du caractère arabe et sa normalisation: réalisation et perspectives», dans *Actes du colloque. Problèmes de l'utilisation de la langue arabe au Maroc*, organisé par l'Académie du Royaume du Maroc, 1993.

(5) L'informations relative à ces banques est tirée des rapports techniques diffusés par Faris Attawil pour la Banque de Jordanie et par Faqari pour la banque BASM lors du colloque *La Technologie informatique au service de la Constitution des Banques de Terminologie et du Dictionnaire Spécialisé* tenu à Tanger en 1995.

(6) F. Attawil expliquait dans une présentation de la banque publiée dans la revue *Al-Lissan Al-Arabi*, 39, 1995 que «L'unité terminologique est classée du point de vue thématique en utilisant parallèlement deux systèmes de classification: le système de codification universel, le CDU, et un système de codification à 10 caractères (chiffres et lettres) conçu à l'institution qui attribue à chaque unité à stocker un code. Ces codes sont considérés comme autant de descripteurs et sont traités comme tels. Ce système a permis la réalisation d'un système de gestion des concepts qui est lui même fondé sur le système ROOT de la BSI (*British Standard Institut*).»

terminologie à l'échelle du monde arabe et il n'y a pas d'observatoire qui puisse assurer la veille technologique et initier de tels rapports. Les informations disponibles sont glanées au hasard des colloques et séminaires panarabes qui se tiennent sur le sujet. Ceci est d'autant plus paradoxal qu'aujourd'hui, avec l'avènement des réseaux et des facilités d'échange des données électroniques, la mise en commun des ressources en matière de terminologie et la coopération entre différentes banques de terminologie s'imposent. (Je reviendrai sur cet aspect de la question.)

Présentation des trois banques de terminologie

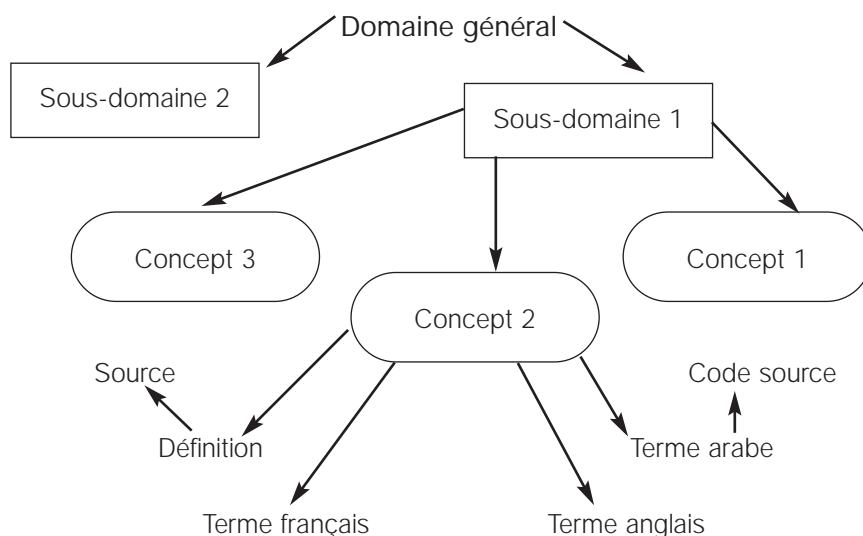
Avant de présenter ces banques, il convient de rappeler que pour adapter, arabiser ou développer des systèmes de gestion de bases de données pouvant traiter la langue arabe, un travail énorme de codification et d'unification a été réalisé à l'échelle du monde arabe afin de résoudre des difficultés de codage informatique dues aux propriétés

inhérentes à la langue arabe telles que son alphabet, le sens droite/gauche de l'écriture, le problème de la vocalisation, la concaténation, etc. (4)

La Banque de terminologie de l'Académie de Jordanie

Constituée en 1988, la banque contient 430 000 termes (5). Elle est supportée par un HP3000/MICROXE et un HP 3000/947, et utilise le Minisis comme système d'interrogation. Du point de vue de sa structuration interne, elle comporte trois sous-systèmes intégrés qui correspondent en fait à trois bases de données complémentaires: un système qui gère les données terminologiques, un système qui gère les données bibliographiques et enfin un système qui gère les données factuelles. Dans la conception de la banque, l'unité terminologique subit le même traitement qu'un document dans une bibliothèque et est classée selon deux modes de classification, ce qui permet de lier cette unité à un thésaurus et de prendre en compte la dimension conceptuelle des données terminologiques (6).

Organisation de la banque du point de vue de sa structure :



La banque terminologique adopte dans sa structuration une approche conceptuelle de système thésaurus et fournit par là même, une représentation cohérente et systématique de la structuration du domaine du savoir traité. Ces relations qui lient les termes aux concepts sont prises en compte au niveau même de la fiche terminologique :

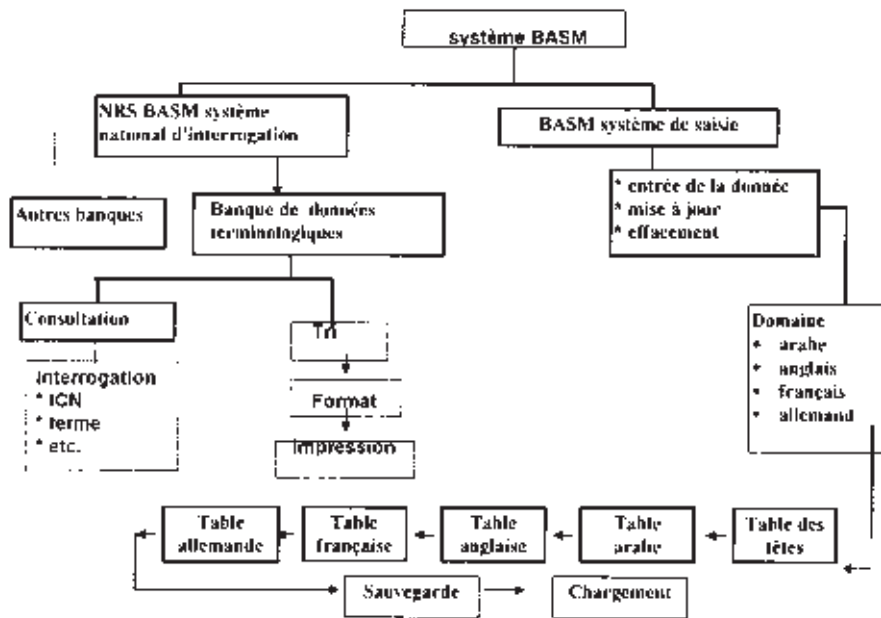
Codes	Données
1 - 01	Terme vedette en arabe, anglais, français et allemand (dans l'ordre)
2 - 02	abréviation
3 - 03	variant orthographique
4 - 04	forme complète
5 - 05	synonyme
6 - 06	degré de synonymie
7 - 07	autonyme
8 - 08	structure syntaxique
9 - 09	déclinaison
10 - 10	code de classement
11 - 11	structure classification des termes domaines
12 - 12	code de pondération
13 - 13	source
14 - 14	code du projet
15 - 15	indicateur grammatical

La Banque de terminologie de l'Arabie saoudite (BASM)

La constitution de la banque de terminologie BASM a commencé en 1983. Une collaboration avec la banque de données TEAM de Siemens (7) a permis à BASM d'acquérir des données terminologiques multilingues qui étaient déjà stockées et prêtes à être

(7) Vers une méthodologie assistée par ordinateur pour le traitement et la diffusion de la terminologie arabe, *Actes du colloque, Casablanca 1993. Langues arabes et technologies informatiques avancées*, Fondation du Roi Abdul-Aziz Al Saoud.

(8) Pour un aperçu historique de l'évolution du Lexar, cf. Ait Taleb (S.) et Benjelloun (F.), 1991: «Présentation de la base de données Lexar», dans *Terminologies nouvelles*, n° 5. Cf. aussi, Richert (N.), 1987: *Arabisation et Technologie*, IERA.



interrogées. BASM a fourni en échange, l'équivalent arabe de ces données. La banque de terminologie BASM a beaucoup évolué et contient d'après les statistiques de 1993 quelque 300 000 données. La banque tourne sur un gros ordinateur IBM ES-9000 et utilise le système VM-SP qui se présente comme un système d'extraction de textes (*Text-Retrieval System*) dont la caractéristique essentielle est son orientation textuelle. L'enregistrement de base est la fiche terminologique identifiée par un ICN (*Internal Control Number*) qui est quadrilingue (arabe, français, anglais et allemand).

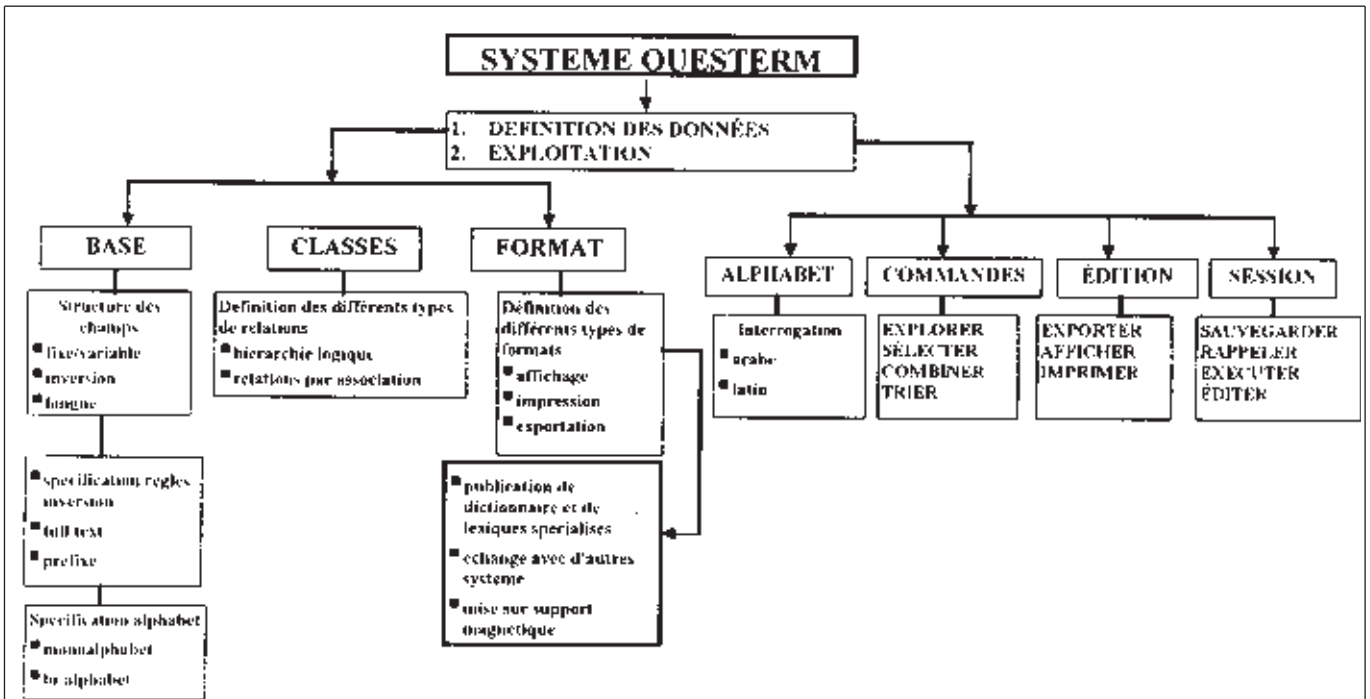
Étant une banque dont les données sont bien structurées, elle permet l'extraction de plusieurs minibases. La mise sur CD-ROM de 200 000 données informatiques a été réalisée et des copies ont été distribuées à des institutions arabes de terminologie. Ce CD est d'autant plus intéressant qu'il fournit deux modes d'accès à l'information, soit une recherche directe en utilisant les fonctions par défaut données sous formes de menu ou un protocole de

recherche plus personnalisé (*Formula Research Mode*) qui améliore l'interactivité pour l'utilisateur.

La banque Lexar de l'IERA

Pionnier en matière de constitution d'une banque de termes bi-alphabétiques, le projet Lexar, démarré en 1978 et ayant fait preuve de sa dimension heuristique par la création même des outils encore inédits, a connu plusieurs virages techniques et méthodologiques. Il n'est pas dans mon propos de faire l'histoire de cette banque (8), mais plutôt de présenter le nouveau système de gestion réalisé pour gérer à la fois la banque telle qu'elle existe et permettre de développer une banque terminologique répondant à toutes les exigences théoriques et méthodologiques en matière de constitution de banques de terminologie.

Ce système fonctionne dans un environnement réseau en architecture client/serveur et est portable puisqu'il peut tourner sur n'importe quelle



Architecture globale du nouveau système

plate-forme informatique. Programmé en langage C ++ pour Windows, il est souple et performant, permettant de créer des bases de données qui répondent à des applications spécifiques sans qu'une hypothèse ne soit faite ni sur le type de données à traiter ni sur leur signification.

Conçue dans l'objectif de renforcer les performances de la recherche, l'interface de consultation permet des clefs de recherche sophistiquées. Les équations de recherche peuvent être simples, en utilisant les termes (terme exact, terme tronqué, terme masqué, etc.), ou complexes en combinant deux ou plusieurs éléments à l'aide d'opérateurs.

En matière d'application à la terminologie, ce système innove dans le sens où il permet de rendre compte de la dimension notionnelle importante en terminologie telle qu'elle a été signalée par J.C. Sager: «A theory of concepts for terminology in essence has only to provide an adequate explanation for cognitive motivations in

Expand (1)

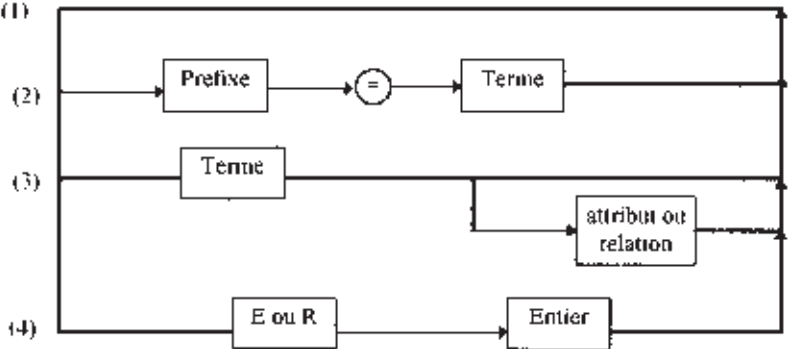


Diagramme syntaxique pour la commande Expand (exploration)

term formation and to provide the basis for structuring vocabularies in more effective way than is offered by alphabetical ordering⁽⁹⁾. » Les préalables théoriques sur lesquels la conception de ce système s'est basée sont :

(9) Sager (J.C.), 1990: A Practical Course in Terminology Processing, J. Benjamins Publishing Company, p. 21.

- Ce qui importe en terminologie ce sont les relations qui lient les termes aux concepts à l'intérieur d'un domaine spécifique, il n'est pas nécessaire d'établir des systèmes conceptuels dans l'absolu et prétendre à l'exhaustivité;
- Les relations entre les référents, en tant qu'objets du monde réel, sont multiples et variées et la structuration

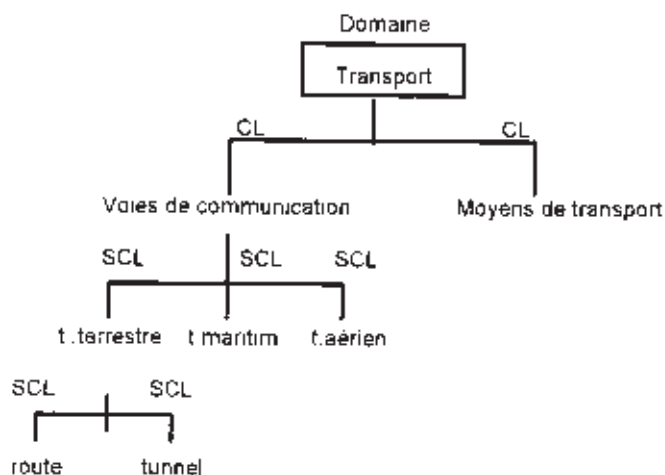
doit s'établir en fonction de l'éclairage qu'elle peut apporter au terminologue, à l'utilisateur ou à l'intégration des données terminologiques dans une application de traitement automatique de la langue (TALN);

– S'il est établi que les relations génériques et partitives sont importantes en terminologie, il n'en reste pas moins que d'autres relations complexes de types cause/effet, fonction, processus, propriété, etc., doivent également être prises en compte;

– Ces relations perçues comme étant pertinentes pour tel ou tel arbre du domaine et sélectionnées à des fins de structuration de ce domaine ne représentent, néanmoins, qu'un sous-ensemble des relations potentielles.

Pour rendre compte au niveau de la consultation et de la mise à jour des données de cette structuration notionnelle, le système utilise le «principe de l'héritage» où les sous-classes héritent l'information des classes; par exemple, si lors de la constitution d'une taxinomie on traite des félidés en tant qu'une sous-classe, celle-ci héritera des propriétés décrites dans la classe des mammifères. Du point de vue technique, l'indexation permet de traduire, selon un langage d'indexation précontrôlé, le contenu notionnel des documents: liant un terme au concept qu'il désigne, précisant les relations conceptuelles qui lient un ensemble de termes à un concept et les relations entre différents concepts. Cette technique est basée sur des règles ⁽¹⁰⁾, ces règles ne sont pas figées et peuvent être définies selon les besoins de structuration du domaine sur lequel travaille le terminologue. Les exemples suivants illustrent cette approche:

(10) Une règle représente les relations potentielles entre deux ou plusieurs concepts.



CL: classe — SCL: sous-classe

Dans les schémas qui suivent: (voir page 62 et suivante)

N.O.T. signifie nombre de termes dans la base;

N.O.D. signifie nombre de documents dans la base contenant le terme;

N.O.R. signifie nombre de relations liées au terme dans la base.

Ces trois banques qui nous concernent ici présentent beaucoup de points communs quant à leurs visées et à leur constitution: – Elles ont toutes les trois pour objectif la collecte, le stockage, l'organisation et l'interrogation de données terminologiques multilingues. Elles se veulent un moyen performant grâce au recours à la technologie informatique pour la diffusion, l'implantation et la mise à jour de la terminologie arabe scientifique et technique;

– Les systèmes utilisés pour la constitution et la gestion de ces banques sont de type documentaire. Orientées vers la consultation et l'exploitation de données terminologiques, elles utilisent la technique de l'inversion avec toutes les possibilités d'indexage, d'organisation, d'accès et d'extraction que permet cette technique;

– Les données sont étroitement structurées selon des champs bien définis, ce qui permet plusieurs points d'accès à l'information contenue dans la fiche terminologique: (l'entrée, une partie de l'entrée, un terme de la définition, une information grammaticale, le champ domaine, etc.) Grâce à cette grande structuration, plusieurs types de sous-ensembles, contenant un ou plusieurs types d'informations selon les besoins ponctuels de la recherche, peuvent être extraits de ces banques. Le système utilisé à l'IERA ainsi que le Minisis, utilisé pour la Banque de Jordanie, permettent en plus l'accès au concept; le premier au moyen des relations conceptuelles structurées et le deuxième par le biais d'un code d'identification qui pointe vers un thésaurus.

Ces banques se caractérisent aussi par un dédoublement des efforts, les corpus stockés dans l'une sont restockés dans les autres, sachant cependant que la saisie des données et leur introduction dans la machine est l'investissement le plus coûteux dans la constitution d'une banque de termes. Elles ont été conçues sans souci de coopération, car il n'y a ni

SYSTEME QUESTERM -Exploitation-

Open Language Query Edit Session Quit Help

EXPAND

No	Classe	N.O.T.	N.O.D.	N.O.R.
E 1	Transport	71	71	0
E 2	>Voies de communication	48	48	0
E 3	>>Transport terrestre	24	24	0
E 4	>>>Route	13	13	0
E 5	>>>Tunnel	11	11	0
E 6	>>Transport maritime	13	13	0
E 7	>>Transport aérien	11	11	0
E 8	>Moyens de transport	23	23	0

Commande >>

LEXAR LATIN EXPAND Transport/Cl

SYSTEME QUESTERM -Exploitation-

Open Language Query Edit Session Quit Help

EXPAND

No	Descripteur	N.O.T.	N.O.D.	N.O.R.
E 1	riverain	56	49	3
E 2	rocade	37	37	6
E 3	roulé	19	13	3
E 4	rouleau	74	63	3
E 5	roulement	45	37	3
E 6	route -----	250	198	80
E 7	route à chaussée unique	25	21	10
E 8	route à grande circulation	23	22	9
E 9	route à sens unique	37	35	11
E10	route collectrice	12	12	10
E11	route de desserte	17	17	10
E12	route de transit	29	28	11
E13	rue	47	34	5
E14	rugosité	23	17	3

Commande >>

LEXAR LATIN EXPAND route

SYSTEME QUESTERM-Exploitation

Open Language Query Edit Session Quit Help

EXPAND				
No	Descripteur	N.O.T.	N.O.D.	N.O.R.
E 1	partie de (partitive)	17	0	0
E 2	type de (générique/spécifique)	42	0	0
E 3	composante de (partitive)	14	0	0
E 4	acton/résultat (associative)	7	0	0

Commande >><<

LEXAR LATIN EXPAND **code/R**

SYSTEME QUESTERM-Exploitation

Open Language Query Edit Session Quit Help

EXPAND				
No	Descripteur	N.O.T.	N.O.D.	N.O.R.
E 1	rocade	37	37	6
E 2	route à chaussée unique	25	21	10
E 3	route à grande circulation	23	22	9
E 4	route à sens unique	37	35	11
E 5	route collectrice	12	12	10
E 6	route de desserte	17	17	10
E 7	route de transit	29	28	11
E 8	rue	47	34	5

Commande >><<

LEXAR LATIN EXPAND **F2**

consensus méthodologique ni structure commune, ce qui représente une limite et un frein à l'évolution de ces banques dans le sens d'une plus grande coopération.

Perspectives et développements futurs

L'arabe étant un bien commun de plusieurs pays, voire de plusieurs centaines de milliers de personnes dispersées aux quatre coins du monde, aucune institution panarabe n'a de prise directe dans le domaine de l'harmonisation et de l'unification de la terminologie ⁽¹¹⁾. Certains spécialistes du domaine craignent même que la babélisation qui

caractérise la terminologie arabe contemporaine ne conduise, à long terme, à la constitution de plusieurs terminologies arabes scientifiques.

La terminologie étant le moyen de véhiculer l'information spécialisée dont elle constitue le vecteur, l'usage d'une terminologie cohérente est le seul garant pour que les notions scientifiques soient comprises de manière identique par les utilisateurs d'une même langue, quels que soient leurs pays ou leurs institutions. Le développement des banques de terminologie semble à l'heure actuelle, l'alternative pour une meilleure diffusion et une plus grande harmonisation: « *The impact of large terminological database is likely to be significant for the improvement of precise, economic and appropriate communication. This development imposes responsibilities on terminologists who should construct and store their information in such a way that it could be widely used* ⁽¹²⁾. »

Banques de terminologie et besoins d'harmonisation de la terminologie

Les grandes institutions arabes de mise au point de la terminologie, comme nous l'avons vu plus haut, n'ont pas hésité à investir très tôt dans la construction de grandes banques de termes, une réflexion sérieuse et une concertation doivent être menées pour faire évoluer ces banques et en faire des systèmes intelligents au service de l'harmonisation et de la diffusion de la terminologie. Le colloque de Tanger, *La technologie informatique au service de la terminologie scientifique et du dictionnaire spécialisé*, tenu en avril 1995 a amorcé le débat. Cependant si on analyse les différentes recommandations, on s'aperçoit que des questions qui sont au cœur de la réflexion des spécialistes de la terminologie dans le monde

d'aujourd'hui sont absentes, à savoir la mise en commun des ressources terminologiques par le développement des échanges électroniques des données, la constitution de systèmes intelligents qui répondent aux nouvelles exigences en matière de terminologie assistée par ordinateur et enfin la prise en compte par la machine de la dimension discursive de la terminologie ⁽¹³⁾.

Banques de terminologie et formats d'échange

Pour mettre en commun leurs ressources terminologiques, se développer rapidement et mieux remplir leur mission de diffusion et d'implantation d'une terminologie arabe scientifique cohérente, et répondre ainsi aux besoins sans cesse croissants en termes nouveaux, ces banques devraient se concerter et adopter l'un des formats intermédiaires standard d'échange électronique conçus et élaborés pour les données terminologiques, le TEI TERM par exemple ⁽¹⁴⁾. En effet, le TEI TERM (*Text Encoding Initiative Term*) est un format qui permet l'échange électronique des données terminologiques entre différents types d'ordinateurs et de logiciels de gestion de banques terminologiques. Les fiches converties dans ce format deviennent accessibles à tous sans qu'il soit nécessaire de connaître les détails du format de départ. Basé sur le SGML (*Standard Generalized Markup Language*), norme internationale (ISO 8879) pour la définition de méthodes de représentation de documents sous forme électronique, le TEI TERM est une norme en vigueur en matière de terminologie dans les plus grands centres, son adoption peut faciliter les échanges des ressources terminologiques informatisées entre nos institutions.

(11) Le Bureau de coordination de l'arabisation, institution panarabe mandatée par l'ALECSO (Organisation arabe pour l'éducation, la culture et la science) pour assurer la coordination entre les différentes institutions de mise au point de la terminologie dans le monde arabe et veiller à la diffusion de cette terminologie parmi les utilisateurs, se débat dans des problèmes institutionnels et ne dispose pas de moyens lui permettant de jouer son rôle de coordination.

(12) *Op. cit.* Note (10) p. 21.

(13) Cf. Condamine (A), 1994: «Bases de connaissances», dans *En Bons Termes*, numéro spécial.

(14) Résultat des efforts conjugués du TC/37 et d'Infoterm pour adapter le TEI pour la terminologie, ce format semble être le standard en la matière. Il existe cependant d'autres formats d'échange, Macro-Mater pour ne citer que celui-là.

(15) En IA les réseaux sémantiques sont des hiérarchies qui représentent la connaissance sous forme de graphes interconnectés où les nœuds décrivent les concepts et les liens (chemins), leurs relations.

(16) Cf. Parent (R.), 1989: « Recherche d'une synergie entre développement linguistique informatisé et systèmes experts: importance de la terminologie », dans *Méta*, 34.

(17) Des expériences concluantes ont été réalisées au Canada par l'équipe de Code et en Europe par l'équipe Aramhis. Code (*Conceptually Oriented Design Environment*) est un système générique d'acquisition des connaissances développé au Laboratoire d'intelligence artificielle de l'Université d'Ottawa. Aramhis (Action Recherche Application Matra/IRIT en interface homme/machine) CNRS, Toulouse. L'approche adaptée dans la réalisation de la base de connaissances est modulaire. Un module prend en charge l'information linguistique associée au terme, un autre module traite les définitions logiques qui dégagent les rapports terme/concept, puis un autre intervient là où les relations qu'entretiennent les concepts entre eux sont structurés.

Banques de terminologie et modélisation des connaissances

Les progrès réalisés en matière de traitement automatique de langues naturelles ont relancé la réflexion sur les fondements de la terminologie, notamment sur la dimension cognitive de cette dernière mettant les formes linguistiques en rapport avec leur contenu conceptuel et soulignant les interrelations tissées dans un même réseau conceptuel appartenant à un domaine du savoir. Par ailleurs les développements de l'intelligence artificielle ont montré qu'il est possible de construire des systèmes conceptuels suffisamment souples et dynamiques ⁽¹⁵⁾, particulièrement quand il s'agit de systèmes construits pour des objectifs et des applications bien définis qui nécessitent une représentation et un transfert des connaissances: « Les réalisations en terminologie et en systèmes experts s'articulent autour de structures notionnelles pour établir des distinctions entre concepts et classes de concepts au moyen: 1) d'une

définition logique basée sur les traits essentiels, 2) d'une description basée sur les traits fréquents 3) des schémas de relations opératoires entre notions ⁽¹⁶⁾. » Nos institutions devront non seulement développer les banques telles qu'elles existent, mais surtout en concevoir d'autres suffisamment intelligentes pour prendre en compte ces dimensions. Les ressources qu'elles accumuleraient pourront être réutilisées dans toutes les applications automatiques qui requièrent des connaissances sur le monde réel. Elles devront rendre compte des termes à la fois dans leur fonctionnement linguistique, à savoir des signes linguistiques qui répondent de manière quasi systématique aux principes de bonne formation qui régissent la langue générale, et en tant que composantes cognitives et « indice d'une connaissance experte ⁽¹⁷⁾ ».

*Saadia Ait Taleb,
Institut d'études et de recherches
pour l'arabisation du Maroc,
Maroc.*

Le système SAMI-BTQ: un système intégré de production et de mise à jour des données de la Banque de terminologie du Québec

SAMI-BTQ est un système de traitement et de gestion de données terminologiques développé à l'Office de la langue française pour alimenter et tenir à jour les fichiers de la Banque de terminologie du Québec. Les principales fonctions de recherche de ce système sont présentées de même que la procédure utilisée pour la mise à jour des données. L'aspect convivial et la facilité d'utilisation de ce dernier permet d'en étendre l'utilisation à l'extérieur de l'Office dans le cadre d'ententes de partenariat.

Termes-clés: banque de terminologie; terminotique; terminologie assistée par ordinateur; système de traitement de données terminologiques.

Introduction

Le présent exposé porte sur le système SAMI-BTQ, système qui permet, de façon intégrée, de produire et de mettre à jour les données de la BTQ (Banque de terminologie du Québec) en vue d'en assurer la consultation et la diffusion sous diverses formes. Toutefois, avant de passer à la présentation proprement dite, il nous apparaît utile de situer brièvement l'Office et le contexte dans lequel y prévaut la production terminologique.

L'OLF (Office de la langue française) est une organisation gouvernementale qui doit veiller à l'application de la Charte de la langue française, loi adoptée en 1977, par laquelle lui sont conférés les mandats de soutenir la francisation des milieux de travail, d'y réussir l'implantation du français et d'assurer le développement de la langue française au Québec en harmonie avec la francophonie.

Dès sa création, l'Office a fait le choix, en matière d'aménagement linguistique, de miser sur le développement de la terminologie comme moyen d'exprimer l'évolution des réalités industrielles et économiques des sociétés modernes ainsi que sur la diffusion des terminologies pour faire du français la langue de travail au Québec. C'est ainsi que la BTQ, dont la création remonte au début des années 70, est devenue un moyen privilégié par l'Office pour réaliser la mission qui lui a été confiée en tant

qu'instrument permettant de recenser les terminologies existantes et de suivre leur évolution.

L'évolution de la production terminologique à l'Office

Au fil des ans, le statut et la fonction de la BTQ ont évolué. Cette évolution est tributaire d'un virage important amorcé, à la fin des années 80, dans le processus de production terminologique à l'Office. Principalement considérée et utilisée au départ comme un système d'archivage et de diffusion en direct de données terminologiques, la BTQ est en effet peu à peu devenue un système dynamique de production de terminologies dont les moyens de diffusion se sont diversifiés (publications, disque optique, etc.). Il était donc devenu nécessaire, pour assurer une bonne exploitation de la masse considérable d'informations qu'avait fini par constituer la Banque (quelque 850 000 fiches terminologiques dans différents domaines scientifiques et techniques avaient en effet été emmagasinées et centralisées dans la BTQ), de l'intégrer en tant que source d'information dans le cadre de toute activité terminologique entreprise à l'Office. Ainsi, dans le souci d'exploiter au mieux le fonds terminologique existant et de fournir aux utilisateurs une information qui soit la plus complète et la plus actuelle possible, les producteurs (actuellement une soixantaine de personnes, qui sont principalement

des employés de l'Office de la langue française travaillant à Québec et à Montréal ainsi que dans les huit bureaux régionaux de l'OLF répartis à travers le Québec) intègrent la BTQ à toutes les étapes de leurs activités, depuis la consultation des fichiers lors des phases préparatoires de leurs travaux jusqu'à la consignation des résultats de ces derniers lorsqu'ils sont terminés. Cette façon de considérer les fichiers de la Banque non seulement comme lieu de destination ultime du résultat des travaux, mais aussi comme source première d'information en assure, par la nécessaire prise en charge du contenu au fil des travaux entrepris, une exploitation efficace ainsi qu'une mise à jour constante.

C'est de ce dépôt dynamique dont sont extraites, périodiquement, les données à jour qui vont faire partie de chaque nouvelle version du disque optique compact de l'Office (mieux connu sous le nom de DOC-BTQ), de même que divers sous-ensembles de fiches destinés à être publiés ou diffusés sous une forme ou une autre (publications sous format papier, disquette, etc.).

L'informatisation de la production terminologique

Pour en arriver à ce résultat, bien sûr, il a fallu définir et développer des outils informatiques qui faciliteraient une mise à jour plus rapide et plus dynamique des données. En effet, les systèmes mis au point à l'époque de la constitution de la Banque, conçus pour le traitement et la gestion de quantités massives de données, ne convenaient plus aux nouvelles caractéristiques du travail ni aux exigences des producteurs. Ces derniers, s'étant entre-temps graduellement convertis à la micro-informatique, exigeaient en effet des outils plus conviviaux, mieux adaptés

à leurs activités, notamment en ce qui a trait à la suppression des opérations répétitives et fastidieuses, à la réduction du nombre d'intermédiaires dans la chaîne de production et à l'intégration des différents aspects de leur travail. De plus, comme dans bien des endroits, des impératifs de rendement de plus en plus exigeants militaient en faveur du développement de systèmes plus performants qui permettraient de réduire les délais de traitement et de production.

Après une évaluation des produits existant sur le marché, il a été décidé, au début des années 90, de développer un logiciel « maison » qui conviendrait à la fois à la gestion d'une base de données de la taille de la BTQ et à la spécificité de l'activité de production terminologique telle qu'elle s'entend maintenant à l'Office: c'est dans ce contexte qu'a été mise au point la première version de SAMI.

À partir de l'automne 1995, les activités de gestion et de diffusion de la BTQ, jusque-là assurées par une société externe, ont été rapatriées à l'Office. Depuis que les données de la BTQ sont logées et gérées à l'Office même, il est devenu beaucoup plus facile et plus rapide, pour les producteurs, d'intervenir sur les données. Cette amélioration résulte en grande partie de l'arrimage entre le dépôt de données et SAMI (qui est alors devenu SAMI-BTQ) et de la poursuite, à l'interne, du développement de fonctions et d'outils logiciels spécialement conçus pour le traitement et pour la gestion de données de nature terminologique.

SAMI-BTQ occupe une position névralgique dans la chaîne de production terminologique, car c'est par là que transitent toutes les données destinées à alimenter les fichiers de la BTQ ou à en ressortir sous forme d'extrants. SAMI-BTQ se caractérise également par ses possibilités d'intégration des différentes composantes du système

de production terminologique situées en amont et en aval du traitement des données proprement dit. Comme la description complète de l'ensemble des outils logiciels utilisés dans la chaîne de production terminologique de l'Office dépasserait largement l'objectif de notre présentation, nous ne mentionnerons ici que quelques-unes des activités que ces programmes permettent d'accomplir: assistance au dépouillement terminologique, comparaison automatique de données aux fins d'alimentation des fichiers, conversion de fichiers, importation et exportation de données, production de sorties sous format de publication, etc.

La présentation de SAMI-BTQ qui va maintenant suivre ne couvrira pas l'ensemble des possibilités du système. Nous avons choisi d'en illustrer certaines fonctions de recherche, de même que la procédure de mise à jour des données, aspects qui confèrent à SAMI-BTQ un intérêt particulier en tant que système de production terminologique.

Présentation de SAMI-BTQ

SAMI-BTQ est conçu sous la forme d'un poste de travail à partir duquel il est possible de consulter directement les fichiers de la BTQ, d'y sélectionner des fiches à traiter, d'en verser une copie dans un fichier de travail, de les traiter et de remettre la version modifiée dans la BTQ. C'est également dans ce même environnement de travail qu'il est possible de créer des fiches nouvelles destinées à alimenter les fichiers de la BTQ.

Le producteur entre dans SAMI-BTQ par un fichier de travail qui lui est propre et dont l'accès lui est réservé par un mot de passe personnel. Dès l'ouverture du fichier de travail, comme le montre la figure 1, le système invite le

producteur à choisir une des trois possibilités qui lui sont offertes par l'intermédiaire de boutons placés dans une barre d'outils: passer directement à la BTQ pour en consulter les fichiers en vue d'une éventuelle extraction de fiches à traiter (bouton BTQ) ou demeurer dans l'environnement de saisie/mise à jour, c'est-à-dire dans le fichier de travail, soit pour y créer de nouvelles fiches (bouton C), soit pour intervenir sur des fiches qui y ont déjà été versées (bouton M).

Nous allons d'abord présenter la démarche qui permet d'extraire des fiches de la BTQ pour les verser dans un fichier de travail, pour terminer avec la présentation de l'environnement de saisie/mise à jour de SAMI-BTQ.

Modalités de recherche de l'information

Le langage d'interrogation

Avant de passer aux modalités de recherche de l'information, il est nécessaire d'apporter quelques précisions sur le langage d'interrogation utilisé dans SAMI-BTQ. Ce langage, qui a été développé à l'Office, permet de poser des questions simples ou complexes et offre les fonctionnalités suivantes: recherche par terme (recherche de l'occurrence exacte du terme ou du syntagme), recherche par mot (recherche du terme seul ou en tant que composante d'un autre terme), inscription des syntagmes en langage naturel, troncature, recherche à l'aide d'opérateurs logiques ou d'opérateurs de proximité. Par défaut, toute recherche dans la BTQ porte sur l'ensemble des champs interrogeables des fiches (entrées principales et sous-entrées) dans les deux langues (anglais et français). Il est toutefois possible de restreindre la portée d'une recherche à une langue ou à un

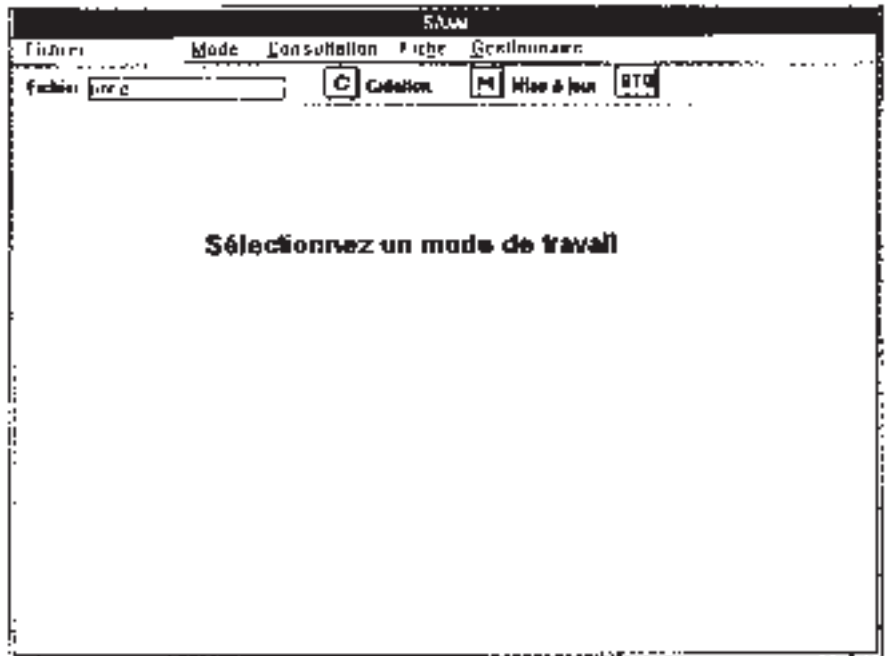


Figure 1

champ en particulier (exemple: recherche du terme xxx dans le champ «synonyme anglais» seulement).

Les modes de recherche

L'activation du bouton BTQ donne directement accès à l'environnement d'où on peut interroger les fichiers de la Banque de terminologie. Pour ce faire, deux modes de recherche sont offerts à l'utilisateur: le mode assisté, dont la page-écran apparaît par défaut (figure 2), et le mode expert. Comme son nom l'indique, le mode de recherche «assisté» a été conçu pour offrir le maximum d'aide à l'utilisateur. Il ne lui est donc pas nécessaire de maîtriser le langage d'interrogation puisque l'écran de recherche a été conçu pour le guider dans la formulation de sa question. Les questions posées dans ce mode peuvent comporter un maximum de quatre éléments.

La recherche en mode expert, par ailleurs, convient davantage à l'utilisateur qui maîtrise le langage

d'interrogation ou qui désire effectuer des recherches plus complexes (un maximum de dix éléments est permis). Dans ce mode, l'interface d'interrogation (figure 3) ne présente qu'une case où l'utilisateur inscrit les différents éléments de sa question en respectant le formalisme du langage d'interrogation. La tâche de saisie des questions est toutefois allégée par la programmation des divers critères de recherche dans les clés de fonctions.

Le repérage de différentes graphies d'un même terme

Certains termes, comme *multiuser* ou *micro-ordinateur*, peuvent s'écrire de différentes façons, c'est-à-dire avec ou sans trait d'union, ou encore en un seul mot. Lorsqu'on recherche de tels termes dans une base de données, ce qui est une situation assez courante en terminologie, on se voit habituellement contraint d'interroger successivement les différentes graphies du terme pour s'assurer d'en repérer toutes les occurrences. Pour contrer ce

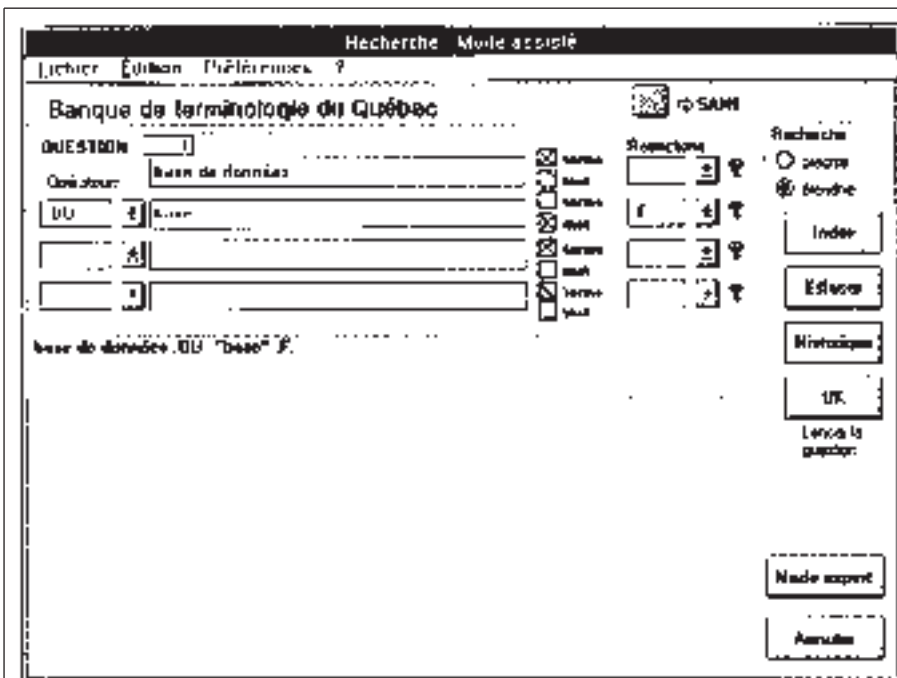


Figure 2

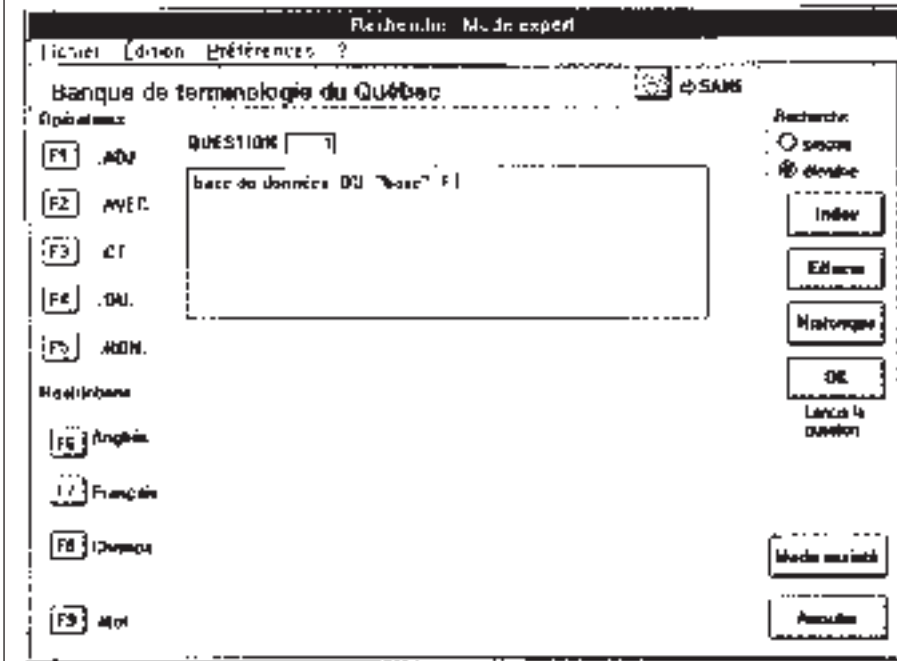


Figure 3

genre de difficulté, le langage de recherche de SAMI-BTQ offre l'option «recherche étendue» qui permet le repérage des différentes graphies de tels termes, et ce, sans égard à la façon dont le terme est inscrit dans la question. Ce repérage unique est rendu possible grâce à certaines interventions qui sont appliquées aux éléments d'information lors de leur indexation, et qui a pour effet de supprimer certains caractères, notamment les parenthèses, les guillemets, les crochets, l'espace, le trait d'union et les signes de ponctuation. Une compression des éléments, par la suite, fait en sorte que les formes *microordinateur*, *micro-ordinateur* et *micro ordinateur*, par exemple, se trouvent en fait toutes indexées sous la forme *microordinateur*. Elles sont donc considérées comme identiques par le système, et, en tant que telles, toutes données en réponse à la question. Lorsque l'utilisateur choisit l'option «recherche étendue», l'existence de graphies différentes lui est signalée dans une fenêtre, (comme celle illustrée à la figure 4) dans laquelle il a le loisir de ne retenir, parmi les formes proposées, que celles qui l'intéressent.

Il est à noter que les différentes graphies d'un terme qui sont fournies en réponse à une question sont des formes qui existent dans la base de données; il faut comprendre que le système ne propose pas à l'utilisateur l'ensemble des graphies qui pourraient éventuellement exister pour le terme recherché.

Visualisation des index

La consultation des index est une autre fonction offerte dans l'environnement de recherche de SAMI-BTQ. L'activation de cette fonction (bouton Index) fait apparaître successivement, dans une fenêtre située dans la portion inférieure de l'écran de recherche, les index anglais et français de tous les

termes compris dans les fiches de la BTQ (figure 5).

Étant donné que le système SAMI-BTQ fonctionne en télécommunication, le contenu de l'index est transféré dans le poste de travail du producteur par « bloc » de 150 termes à la fois, dont une douzaine sont visibles dans la fenêtre. Dans la case située au bas de chaque index apparaît le dernier terme du bloc d'index affiché, ce qui sert de repère à l'utilisateur. Cette case permet également de se positionner directement à un endroit précis dans l'ensemble de l'index, en y inscrivant le nombre voulu de caractères du terme recherché. Les boutons fléchés situés de part et d'autre de cette case permettent de se déplacer d'un bloc à un autre. Tout terme de l'index peut être transféré d'un simple clic dans l'écran de recherche sans avoir à y être réécrit.

On remarquera que les interventions appliquées aux données lors de leur indexation, dont il a été question précédemment, ne nuisent aucunement à la présentation des différentes graphies des termes dans les index visualisables. Dans la portion d'index français présenté à la figure 5, par exemple, on voit en effet que les différentes graphies de *cybercafé* et de *cyber espace* se présentent tout à fait correctement même si ces deux formes sont considérées comme identiques du point de vue du langage de recherche.

L'affichage de la réponse à une question

Le résultat obtenu, une fois la question posée, apparaît dans la partie inférieure de l'écran de recherche. Dans l'exemple de la figure 6, on voit que les fichiers de la BTQ renferment 43 fiches qui répondent à la question posée, réponse qui doit être interprétée comme suit: il y a 43 fiches qui comportent le terme exact *site* (puisque l'option sélectionnée est la recherche par

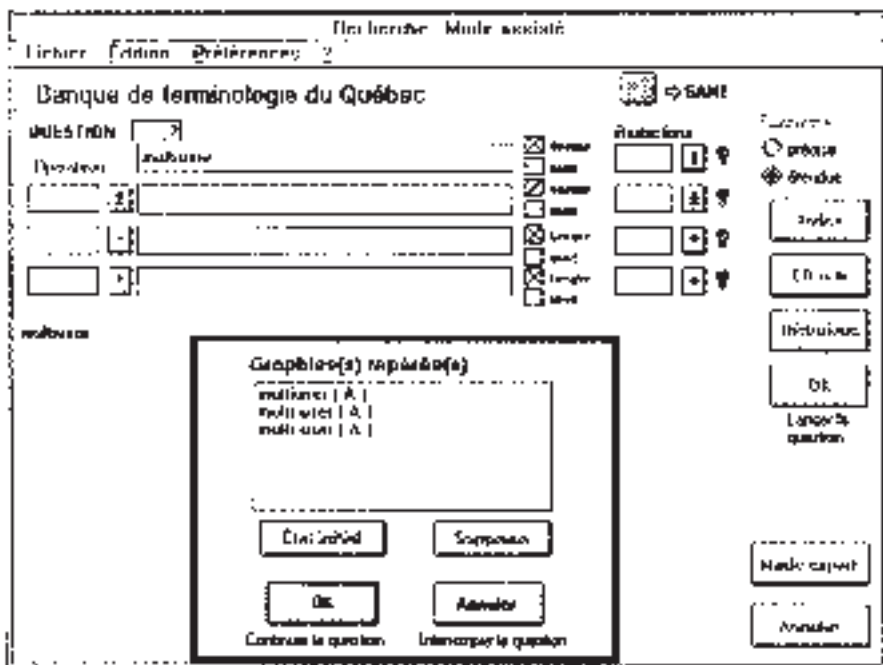


Figure 4

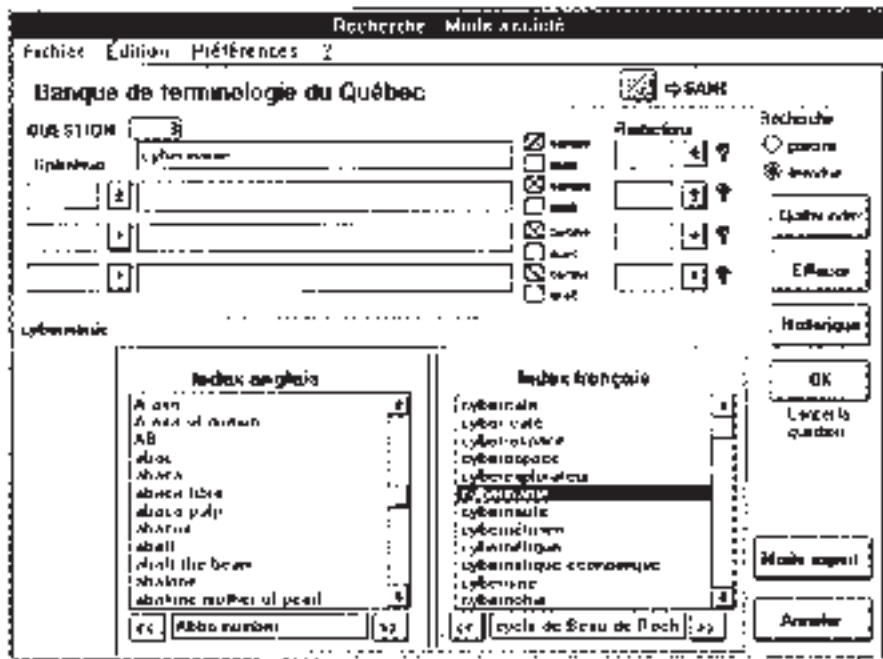


Figure 5

terme) repéré tant en anglais qu'en français (puisque aucune restriction de langue n'a été appliquée à la question), dans l'un ou l'autre des champs interrogeables des fiches (puisque on n'a pas restreint l'interrogation à un champ en particulier).

Le système offre trois modes d'affichage de la réponse à une question donnée. Le choix «fiche par fiche» permet de visionner, l'une après l'autre, en format complet, l'ensemble des fiches obtenues en réponse à une question. À titre d'exemple, la figure 7 présente la 32^e des 43 fiches obtenues en réponse à la question posée.

Il est utile de savoir que dans ce mode d'affichage, le contenu codé des champs 025 (code de document) et 295 (sources) peut être décodé en cliquant sur le contenu du champ en cause. Le résultat de ce décodage s'affiche dans une fenêtre superposée, comme le montre l'exemple de décodage d'un numéro de source présenté à la figure 8.

La deuxième façon de consulter le résultat d'une question est l'affichage des fiches sous format sommaire, illustré à la figure 9. Comme son nom l'indique, ce format permet d'obtenir un aperçu global d'un certain nombre de fiches dont un nombre restreint de champs sont affichés: certaines données de gestion, le domaine d'emploi ainsi que les entrées principales en anglais et en français. De plus, en regard de chaque fiche, à droite, des crochets signalent la présence d'éléments comme les définitions, contextes et notes, donnant ainsi un indice de la richesse terminologique de chaque fiche. Il suffit de cliquer deux fois sur une fiche du sommaire pour la visualiser en format complet. De la même façon que dans le cas des index, les informations du sommaire sont transférées dans le poste de travail par télécommunication; le sommaire s'affiche donc par bloc de 40 fiches à

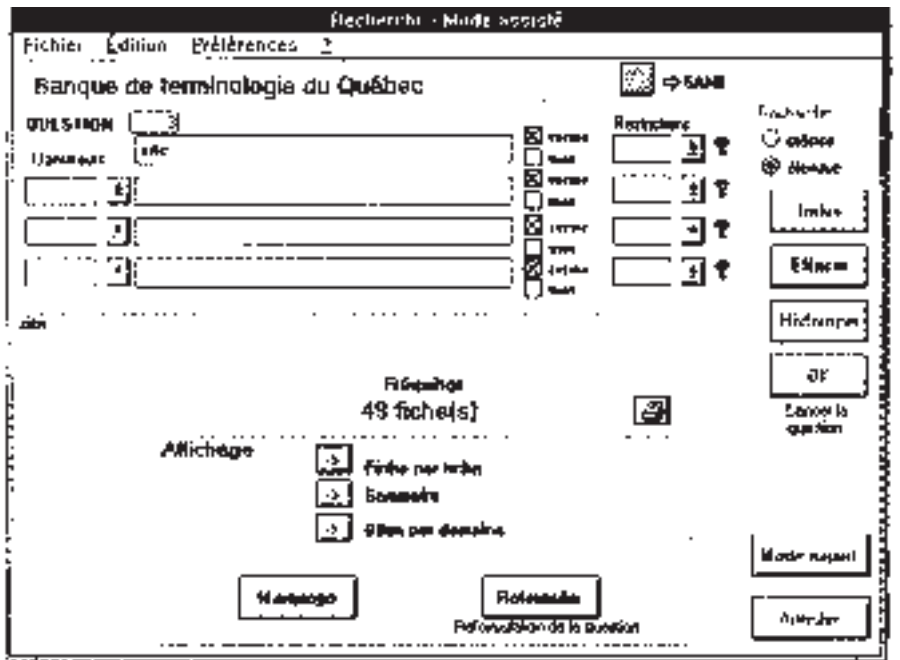


Figure 6

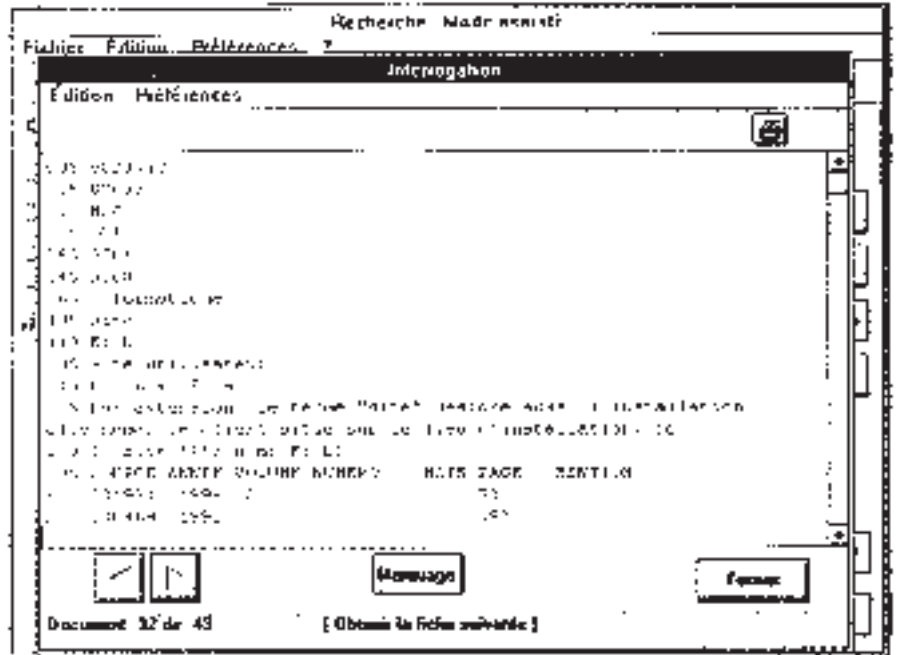


Figure 7

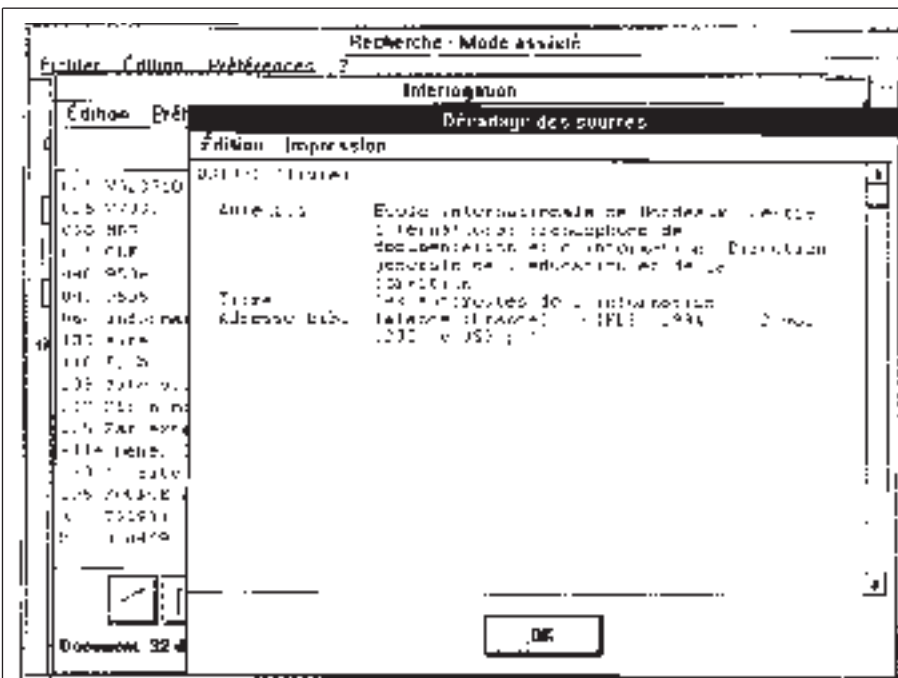


Figure 8

la fois, à raison de huit par page-écran. Des boutons fléchés au bas de la page-écran permettent de passer d'un bloc de sommaire à un autre.

Enfin, le résultat d'une question peut être affiché sous forme de bilan par domaines. Cette option fait apparaître la liste, en ordre alphabétique, des domaines génériques auxquels appartiennent les différentes fiches obtenues en réponse à une question. Comme l'illustre la figure 10, on voit que les 43 fiches de notre exemple sont réparties en 27 domaines, et que deux fiches relèvent du domaines de l'informatique.

Cette forme d'affichage permet donc de choisir, parmi les regroupement de fiches obtenus, ceux qui sont susceptibles de fournir les meilleures réponses, ou encore de faciliter le traitement de certains sous-ensembles de fiches. Ce mode d'affichage permet en effet de restreindre la sélection de fiches à celles relevant d'un domaine particulier en cliquant sur le domaine en question.

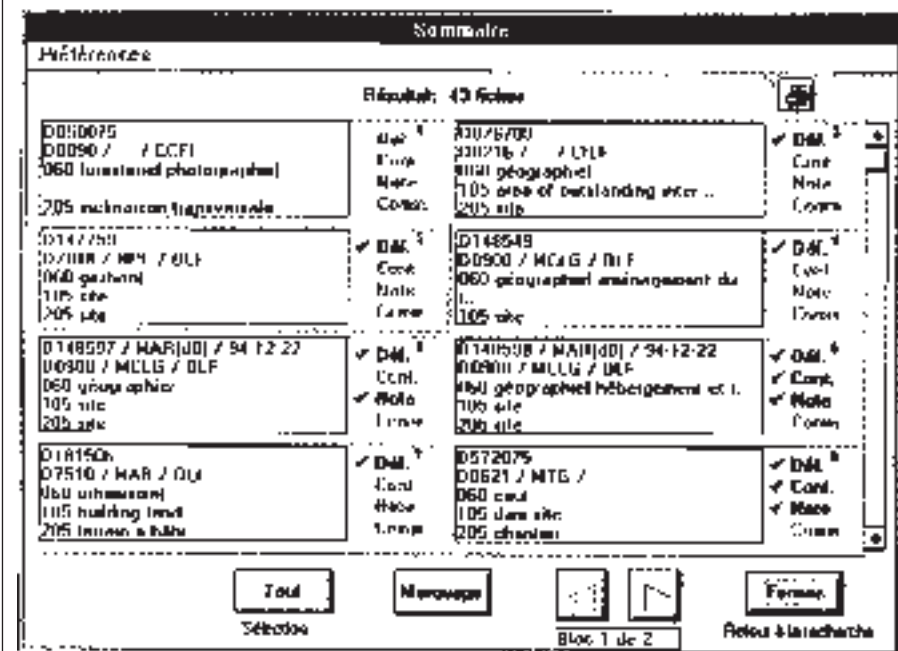


Figure 9

L'historique des questions posées

Le module de recherche de SAMI-BTQ dispose également d'une fonction d'historique qui permet de faire afficher la liste des questions déjà posées dans la séance de travail en cours et, au besoin, de rappeler une question dans l'écran de recherche sans avoir à la réécrire.

La sélection de fiches à traiter et leur transfert dans un fichier de travail

Lorsqu'un producteur décide d'intervenir sur une ou plusieurs fiches de la BTQ, il lui faut d'abord les sélectionner puis commander le versement d'une copie de ces fiches dans son fichier de travail. Au moment où une copie d'une fiche est versée dans un fichier de travail, la version originale de cette fiche, qui demeure entre-temps interrogable dans la BTQ, est simultanément

«marquée» d'un message de mise à jour qui permet d'identifier soit la personne qui veut intervenir sur la fiche, soit le projet dans le cadre duquel elle travaille. C'est la raison pour laquelle cette opération se nomme *marquage*. Elle est déclenchée par l'activation du bouton du même nom, que l'on trouve au bas de la page-écran où s'affiche le résultat d'une question, de même que sur les pages-écrans de chacun des modes d'affichage de la réponse. Dans le cas du mode «fiche par fiche», seule la fiche affichée à l'écran sera transférée dans le fichier de travail, alors que dans les deux autres modes on doit auparavant avoir effectué une sélection parmi les fiches ou les domaines affichés. Dans ces deux derniers cas, il est possible d'agir sur une seule fiche ou sur plusieurs fiches à la fois. La figure 11 présente une sélection de deux fiches affichées en format sommaire. L'activation du bouton «marquage» situé au bas de cette page-écran activera donc le versement d'une copie des fiches sélectionnées dans le fichier de travail qui a été ouvert au début de la présentation, soit le fichier rint.p.

Il faut savoir également que le fait qu'une fiche soit marquée empêche toute autre personne d'intervenir sur son contenu jusqu'à ce qu'elle soit «libérée» par le producteur qui l'avait sélectionnée et transférée dans son fichier.

L'accès aux fiches versées dans un fichier de travail

C'est dans l'environnement de saisie/mise à jour de SAMI-BTQ que l'on peut créer de nouvelles fiches (mode création) et intervenir sur des fiches qui ont été versées dans le fichier de travail (mode mise à jour). Le choix de ce dernier mode donne accès au module de recherche qui a permis d'interroger la BTQ, mais dont la cible, dans cet environnement, est le fichier de travail que l'on interroge afin d'y

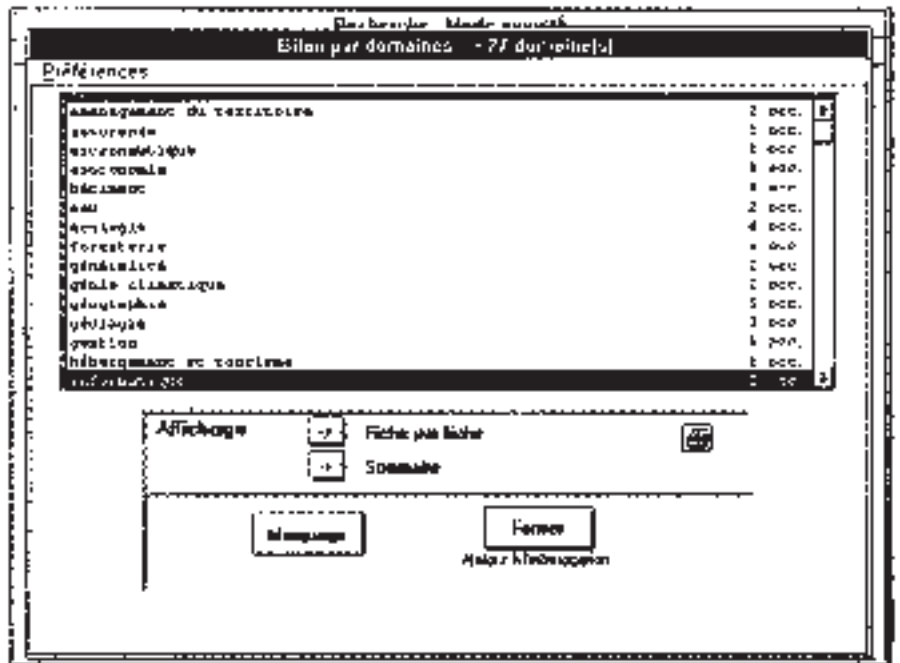


Figure 10

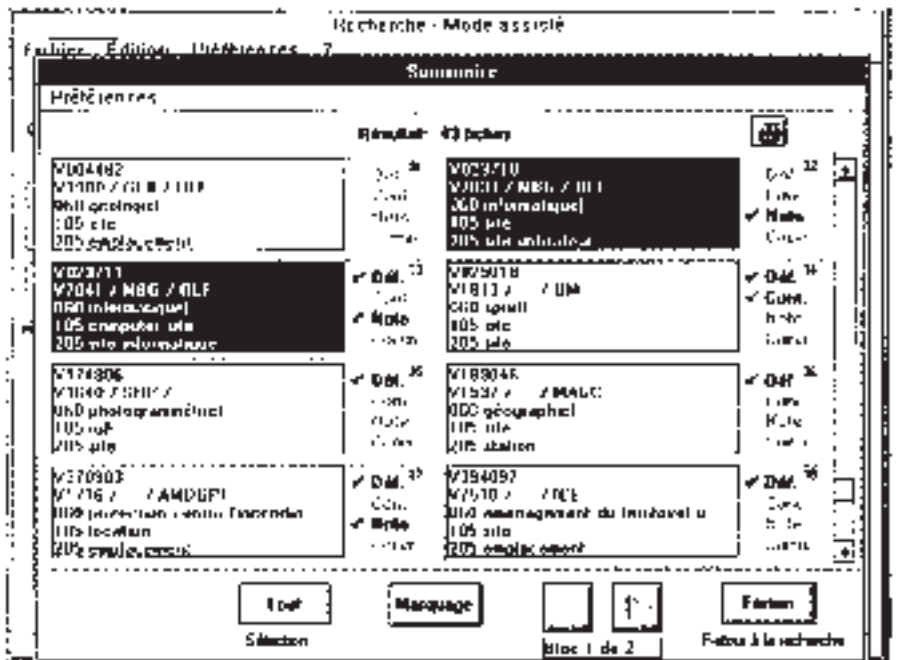


Figure 11

recupérer, parmi toutes les fiches qui y ont été versées ou créées, celles à traiter (figure 12). Selon une procédure analogue à celle du marquage, l'activation du bouton «traitement» permet cette fois-ci de transférer les fiches dans le module de traitement proprement dit, où les interventions peuvent effectivement être faites sur les fiches.

Dans l'environnement de traitement, les fiches sont en quelque sorte fragmentées; leur contenu est réparti dans trois types de pages-écrans, dont deux sont de longueur fixe et correspondent respectivement à la page des champs de gestion et à la page des sources de la fiche. Quant à la partie terminologique proprement dite de la fiche, elle est de longueur variable; dans cette partie, l'interface est dynamique, c'est-à-dire qu'elle s'adapte à chaque fiche de façon à n'en présenter que les champs ayant réellement un contenu. Dans cet environnement, tous les champs sont affichés sous forme de zones d'édition séparées. Les différentes pages-écrans de la fiche font l'objet des figures 13, 14 et 15 (voir page 75 et 76).

L'organisation de l'environnement de saisie/mise à jour de SAMI-BTQ offre de nombreux avantages, en ce qui a trait, en particulier, à l'aide à la rédaction des fiches: inscription automatique de certaines valeurs par défaut, prise en charge de règles d'écriture et de séparation de champs, structuration automatique de la fiche en cours de rédaction, validations automatiques et rétroaction de l'interface, affichage de listes de valeurs permises, fonctions intégrées de décodage de données codifiées, fonctionnalités courantes de traitement de texte, disponibilité des fonctions couper/coller et copier/coller disponibles à l'intérieur d'une même fiche et d'une fiche à l'autre.

Une fois les interventions terminées sur une fiche, le producteur n'a qu'à en commander le retour vers

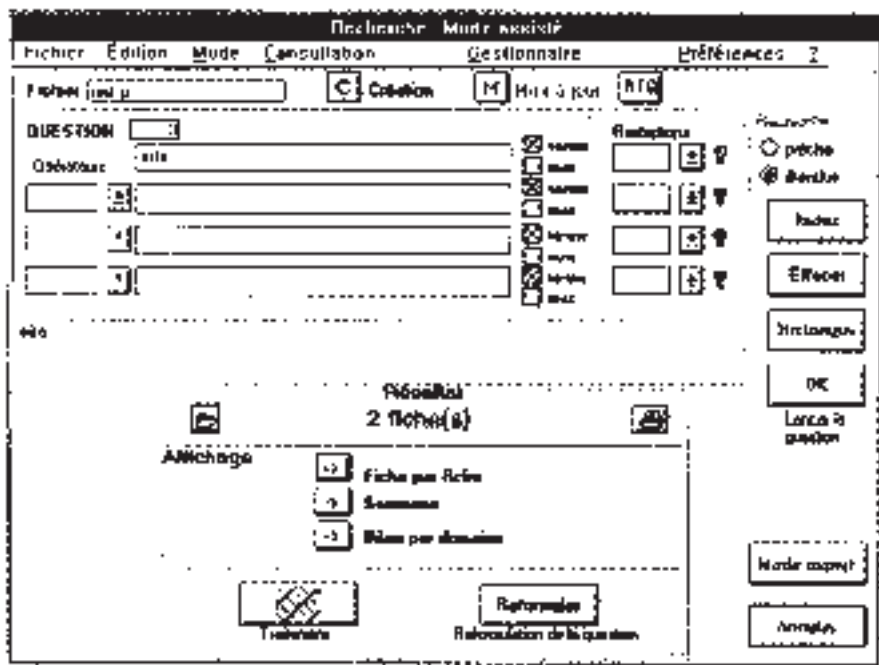


Figure 12

les fichiers interrogeables de la BTQ (option «transmettre» du menu «fiche»). Si toutes les règles de validation sont respectées, la fiche sera extraite du fichier du producteur et rechargée dans la BTQ le soir même; si elle n'est pas valide, le producteur en est informé par un message de l'interface; il peut alors décider de corriger la fiche immédiatement ou ultérieurement. SAMI-BTQ a donc grandement amélioré la souplesse et la rapidité de l'activité de mise à jour des données: en effet, moins de vingt-quatre heures peuvent s'écouler entre le moment où une fiche est repérée, traitée et rendue interrogeable sous sa nouvelle version, alors que la même séquence d'opérations pouvait exiger plusieurs jours, voire des semaines, il y a encore peu de temps. Nul n'est besoin d'ajouter que cette rapidité représente un atout encore plus intéressant dans le contexte de la mise en diffusion de la BTQ dans Internet.

Conclusion

La Banque de terminologie du Québec se situe actuellement à un moment important de son histoire. En effet, en plus d'amorcer un virage technologique marqué préparatoire à son entrée sur l'autoroute de l'information, l'Office s'est récemment engagé dans la voie d'une collaboration plus étroite avec l'extérieur. Aux producteurs internes de l'Office se sont ainsi ajoutés, au cours de la dernière année, quelques partenaires externes avec lesquels des projets de collaboration ont été mis sur pied. L'aspect convivial et la facilité d'apprentissage et d'utilisation de SAMI-BTQ permettent en effet de conclure des ententes avec des partenaires producteurs en vue d'échanger ou de partager des données ou des services. Il est à noter, toutefois, que l'accès aux fichiers de la BTQ est limité, dans le cas des

Champs de gestion

Fichier: Mode: Consultation Fiche: Mise à jour Outils: ?

005 010 015 025 96.02.29

030 035 040 045 050

Domaines/généralisés

060

Domaines/spécifiques

065

Table variable de la fiche

Fichier: Mode: Consultation Fiche: Mise à jour Outils: ? Page: 1

005 0020011 x 002 96.02.29

010 015 020 025 030 035 040 045 050 055 060 065 070 075 080 085 090 095 100 105 110 115 120 125 130 135 140 145 150 155 160 165 170 175 180 185 190 195 200 205 210 215 220 225 230 235 240 245 250 255 260 265 270 275 280 285 290 295 300 305 310 315 320 325 330 335 340 345 350 355 360 365 370 375 380 385 390 395 400 405 410 415 420 425 430 435 440 445 450 455 460 465 470 475 480 485 490 495 500 505 510 515 520 525 530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995 1000 1005 1010 1015 1020 1025 1030 1035 1040 1045 1050 1055 1060 1065 1070 1075 1080 1085 1090 1095 1100 1105 1110 1115 1120 1125 1130 1135 1140 1145 1150 1155 1160 1165 1170 1175 1180 1185 1190 1195 1200 1205 1210 1215 1220 1225 1230 1235 1240 1245 1250 1255 1260 1265 1270 1275 1280 1285 1290 1295 1300 1305 1310 1315 1320 1325 1330 1335 1340 1345 1350 1355 1360 1365 1370 1375 1380 1385 1390 1395 1400 1405 1410 1415 1420 1425 1430 1435 1440 1445 1450 1455 1460 1465 1470 1475 1480 1485 1490 1495 1500 1505 1510 1515 1520 1525 1530 1535 1540 1545 1550 1555 1560 1565 1570 1575 1580 1585 1590 1595 1600 1605 1610 1615 1620 1625 1630 1635 1640 1645 1650 1655 1660 1665 1670 1675 1680 1685 1690 1695 1700 1705 1710 1715 1720 1725 1730 1735 1740 1745 1750 1755 1760 1765 1770 1775 1780 1785 1790 1795 1800 1805 1810 1815 1820 1825 1830 1835 1840 1845 1850 1855 1860 1865 1870 1875 1880 1885 1890 1895 1900 1905 1910 1915 1920 1925 1930 1935 1940 1945 1950 1955 1960 1965 1970 1975 1980 1985 1990 1995 2000 2005 2010 2015 2020 2025 2030 2035 2040 2045 2050 2055 2060 2065 2070 2075 2080 2085 2090 2095 2100 2105 2110 2115 2120 2125 2130 2135 2140 2145 2150 2155 2160 2165 2170 2175 2180 2185 2190 2195 2200 2205 2210 2215 2220 2225 2230 2235 2240 2245 2250 2255 2260 2265 2270 2275 2280 2285 2290 2295 2300 2305 2310 2315 2320 2325 2330 2335 2340 2345 2350 2355 2360 2365 2370 2375 2380 2385 2390 2395 2400 2405 2410 2415 2420 2425 2430 2435 2440 2445 2450 2455 2460 2465 2470 2475 2480 2485 2490 2495 2500 2505 2510 2515 2520 2525 2530 2535 2540 2545 2550 2555 2560 2565 2570 2575 2580 2585 2590 2595 2600 2605 2610 2615 2620 2625 2630 2635 2640 2645 2650 2655 2660 2665 2670 2675 2680 2685 2690 2695 2700 2705 2710 2715 2720 2725 2730 2735 2740 2745 2750 2755 2760 2765 2770 2775 2780 2785 2790 2795 2800 2805 2810 2815 2820 2825 2830 2835 2840 2845 2850 2855

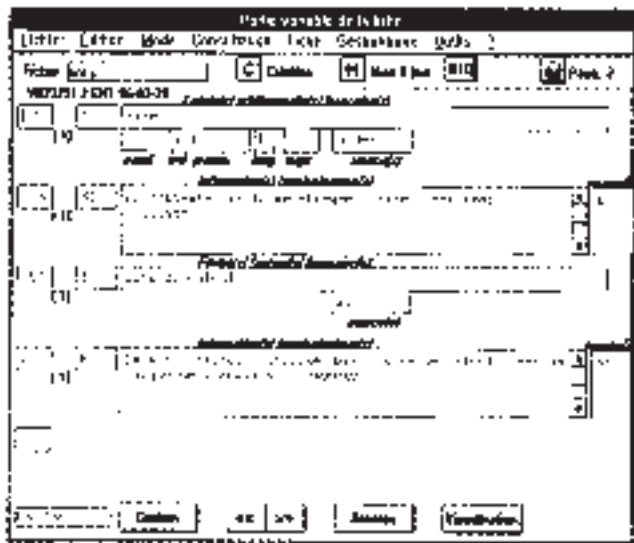


Figure 14

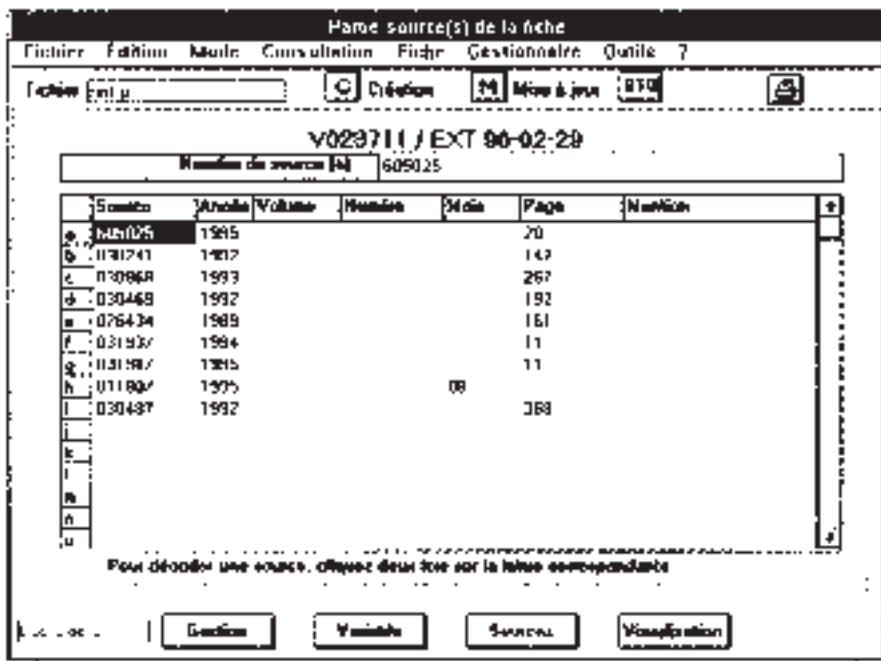


Figure 15

ententes avec des partenaires externes, aux modes consultation et copie, c'est-à-dire qu'il ne leur est pas possible d'intervenir sur le contenu des fiches. Les ententes peuvent avoir des objectifs différents, comme de participer à l'alimentation de la BTQ, comme c'est le cas d'un contrat de veille néologique conclu avec l'Université Laval, ou encore de répondre à des besoins de production propres à un organisme particulier. C'est ainsi que le système SAMI-BTQ a été mis à la disposition du Laboratoire d'informatique et de terminologie de la réadaptation et de l'intégration sociale du Centre François-Charon, dans le cadre d'un projet de collaboration qui a permis la production du premier volet du *Dictionnaire de la réadaptation*. L'expérience s'est avérée fort concluante puisque les auteurs en sont actuellement à la préparation des deuxième et troisième volets de cette publication, et ce, toujours à l'aide de SAMI-BTQ.

Par son ouverture sur de nouvelles formes de partenariat et grâce à son évolution technologique qui en facilite le développement et la diffusion, la BTQ s'assure donc de conserver sa position d'avant-garde sur la scène de la terminologie.

*Sylvie Pelletier,
Office de la langue française,
Québec.*

L'ATelier du TERminologue (Latter[©])

L'ATelier du TERminologue (Latter[©]) ou poste Latter[©] du Gouvernement du Canada regroupe plusieurs applications, dont le programme Latter[©] – application DOS, pouvant assister le terminologue dans son travail. Conçu principalement pour éliminer les doublons de tâches et pour accélérer l'emmagasinement des fiches dans Termium[®] et la production de lexiques et de vocabulaires, le programme Latter[©], à la fois base de données et gestionnaire de fiches terminologiques, fournit aux terminologues un moyen simple et efficace de rassembler, de stocker, d'analyser et de synthétiser l'information terminologique.

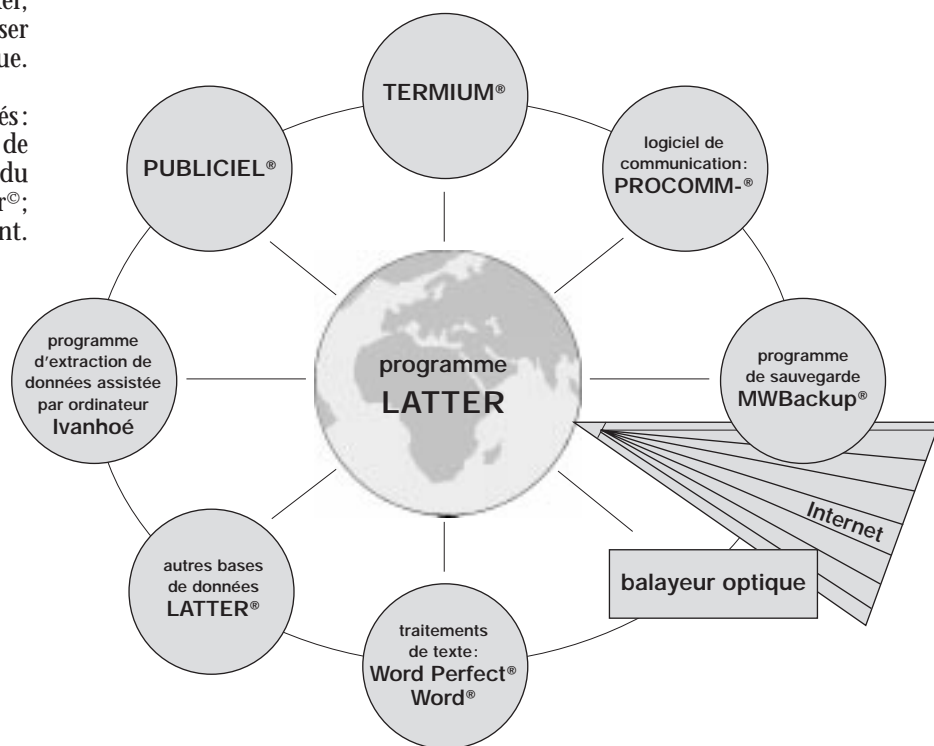
Termes-clés :
base de données; gestionnaire de fiches; Latter[©]; L'ATelier du TERminologue; poste Latter[©]; Gouvernement du Canada; Rint.

Latter[©] désigne deux choses : L'ATelier du TERminologue, poste de travail du terminologue du Bureau de la traduction de Travaux publics et Services gouvernementaux Canada, et le programme Latter[©], base de données et gestionnaire de fiches terminologiques.

L'ATelier du TERminologue englobe tous les outils nécessaires au terminologue dans l'accomplissement de ses tâches :

- Un ou plusieurs traitements de texte comme par exemple WordPerfect[®] pour Windows[®] ou DOS, ou encore Word[®] pour Windows[®];
- Un programme de sauvegarde MwBackup[®] pour la sauvegarde des données;

L'ATelier du TERminologue, Latter[©]



Le programme Latter est au centre de l'ATelier.
Il interagit avec toutes les autres applications.

– Le logiciel d'édition Publiciel® du Bureau de la traduction. Ce logiciel sert à la mise en page bilingue ou trilingue (anglais, français et espagnol) des publications tels les lexiques (*Petits lexiques*) et les vocabulaires (*Bulletin de terminologie* — BT);

– Un logiciel de communication PROCOMM + ® assure l'entrée en communication directe avec Termium®, la base de données linguistiques du Gouvernement du Canada;

– Le programme d'assistance au dépouillement unilingue, bilingue ou multilingue Ivanhoé. Ce programme extrait, d'un texte électronique, les termes marqués par le terminologue: (<< terme retenu, avec justification >> ou [[terme retenu, sans justification]]), les sources correspondantes et, selon le cas, les justifications et les transfert dans un fichier de fiches qu'il crée en un format d'importation Latter®. De là, les fiches seront importées dans la base de données Latter® où elles feront l'objet de recherches complémentaires en vue de leur éventuelle exportation vers Termium® ou Publiciel®, ou vers les deux;

– Les autres bases de données Latter® – le poste de travail peut contenir plus d'une base de données Latter®. Chaque base pourra être réservée à des projets spécifiques, par exemple, un projet de dépouillement en «environnement» et un autre, dédié aux «fractales».

Les postes en commun comprennent:

– Le lecteur optique pour la conversion électronique des textes sur support papier. Ces textes sont dépouillés par le terminologue, et mis en format d'importation Latter® (Voir rubrique Ivanhoé, ci-haut mentionnée).

– Internet qui donne accès à l'inforoute.

Les applications, DOS et Windows®, sont à la fois interreliées

et indépendantes, le noyau central étant le programme Latter®.

Le Bureau de la traduction est en train de créer une version Windows de Latter® améliorée. Cette version sera enrichie d'ensembles permanents qui faciliteront la gestion des fiches. La fonction «Recherche/Remplacement» sera aussi améliorée et des formats d'impression de vocabulaires ou lexiques multilingues s'ajouteront aux options existantes.

Le programme Latter®

Le programme Latter® fournit aux terminologues un moyen simple et efficace de rassembler, de stocker, d'analyser et de synthétiser l'information terminologique. Il fut conçu principalement pour éliminer les dédoublements ⁽¹⁾ de tâches et pour accélérer ⁽²⁾ l'emmagasinement des fiches dans Termium® ainsi que la production de lexiques et de vocabulaires.

Latter® est un programme DOS qui, pour plus de convivialité, utilise un environnement Windows®. Le terminologue dispose donc de certaines fonctions Windows® comme les touches [Alt + Tab] pour basculer d'une application à l'autre et de la fonction «Couper/Coller» pour importer des contextes de Termium®, de WordPerfect®, d'une autre base de

(1) Latter® élimine la triple saisie soit celle effectuée par le terminologue, par la préposée à l'emmagasinement ^{TT} Termium® et par la préposée à l'édition (Publiciel®).

(2) Les fiches rédigées dans Latter® sont exportées vers Termium® et emmagasinées. Les préposées n'ont pas à dactylographier les fiches Latter®, donc il n'y a plus d'engorgement à la saisie.

données Latter®, etc., de les coller et inversement.

L'écran de rédaction Latter®, la fiche Latter®

Inspiré du modèle de conception de la fiche Termium®, l'écran de rédaction Latter® affiche un bloc linguistique imbriqué à l'intérieur de blocs communs d'encadrement. Ces blocs communs peuvent contenir autant de blocs linguistiques qu'il y a de langues autorisées dans Termium® IV. Le tout constitue la fiche Latter® (Voir la fiche Latter®, en annexe). Sur la gauche de l'écran figurent les noms des champs qui sont indélébiles et qui s'affichent automatiquement lors de l'activation d'une fiche. L'utilisateur saisit les données sur la droite de l'écran, dans des fenêtres ponctuelles. La fiche se modifie en cours de saisie. Il n'est pas obligatoire de remplir tous les champs. Pour sauvegarder une fiche, seul le champ «Langue» doit absolument être rempli. Certains champs (à valeur fixe) comportent une validation automatique. Une validation s'effectue sur les fiches exportées vers Termium®. Les blocs principaux (bloc en-tête, bloc linguistique, bloc sources, bloc technique) sont identifiés par un losange et les blocs secondaires, par exemple bloc vedette, justification, réseau notionnel, etc., par une ligne pointillée. Le terminologue peut aussi personnaliser l'affichage de la fiche. Il peut créer un ou plusieurs formats d'affichage ne comportant que les champs sélectionnés, par exemple le code de domaine, le code de langue, le bloc vedette/synonyme, le bloc justification, le bloc source. À chaque projet correspond le format adapté.

Une barre menu figure directement sous la barre de titre de l'écran de rédaction Latter® et de toutes autres fenêtres (Ouverture, Bilan, etc.). La barre menu liste les

divers menus donnant accès à des commandes, à des options ou à d'autres menus. Une ligne d'aide, si activée, spécifique à la fenêtre active, s'affiche au bas de l'écran. Elle liste les différentes touches de déplacement et certaines fonctions principales de Latter[©].

Latter[©] fonctionne sur deux plans. Le premier plan ou avant-plan, visible pour l'utilisateur, correspond à l'écran, c'est l'ensemble «actif». Le second plan ou arrière-plan constitue la base de données proprement dite où les fiches sont conservées.

Base de données

La base de données (arrière-plan) est conçue pour emmagasiner et gérer des fiches. L'utilisateur de Latter[©] peut donc créer, modifier, supprimer, imprimer, exporter et importer des fiches unilingues, bilingues ou multilingues. Latter[©] les classe par ordre de création ou d'importation. Chaque fiche porte un numéro unique suivant l'ordre d'entrée dans la base. Pour accéder aux fiches et les amener à l'écran, il faut passer par l'option «Ouverture» du menu «Ensemble» et sélectionner un champ indexé. Seuls ces champs indexés permettent d'extraire des fiches de la base de données, par champs combinés ou non.

Le terminologie peut donc :

Créer, Modifier ou Supprimer	Une ou plusieurs fiches ou fiches modèles. Par fiches modèles, on entend une fiche sur laquelle on consigne les éléments communs à un lot de fiches, par exemple : le code d'auteur de la fiche, le domaine, les paramètres (correct, masculin, etc.), les observations, les sources, etc.															
Valider	Une ou plusieurs fiches. Latter [©] vérifie la présence obligatoire d'éléments d'information exigés en vue d'une exportation vers Termium [®] .															
Imprimer	Une ou plusieurs fiches en format continu, parallèle, lexique ou vocabulaire. (Voir exemples en annexe).															
Exporter	<table border="0"> <tr> <td>Vers Termium[®]</td> <td>Importer</td> <td>de Termium[®]</td> </tr> <tr> <td>Vers Publiciel[©]</td> <td></td> <td>de Publiciel[©]</td> </tr> <tr> <td>Vers un traitement de texte</td> <td></td> <td>d'un traitement de texte</td> </tr> <tr> <td>Vers un autre Latter[©]</td> <td></td> <td>(Ivanhoé, autres lexiques ou vocabulaires, etc.)</td> </tr> <tr> <td></td> <td></td> <td>d'un autre Latter[©]</td> </tr> </table>	Vers Termium [®]	Importer	de Termium [®]	Vers Publiciel [©]		de Publiciel [©]	Vers un traitement de texte		d'un traitement de texte	Vers un autre Latter [©]		(Ivanhoé, autres lexiques ou vocabulaires, etc.)			d'un autre Latter [©]
Vers Termium [®]	Importer	de Termium [®]														
Vers Publiciel [©]		de Publiciel [©]														
Vers un traitement de texte		d'un traitement de texte														
Vers un autre Latter [©]		(Ivanhoé, autres lexiques ou vocabulaires, etc.)														
		d'un autre Latter [©]														
Interroger Annuler Modifier	Termium [©] fiches Termium [®] Accès non direct à Termium [®] fiches Termium [®]															

Choisissez une valeur

Entrée
Projet L
Projet T
Domaine L
Domaine T
N° de fiche
N° matri T
Renvoi
Clé regr.
Source
Année de source
ID cr. L
Date cr. L
Date soum. rév.
Date L à T
Date MAJ T

Fig. Option *Ouverture*, menu *Ensemble*: liste des champs indexés.

Ensemble

L'ensemble se compose des fiches actives à l'avant-plan, c'est-à-dire des fiches apparaissant à l'écran. Ces fiches peuvent être remplies ou non, porter ou non un même code de projet Latter[©], aborder ou non un même thème. Un ensemble peut donc contenir des fiches de même code de codes différents (Normes, ISO, Métrologie). Seules les fiches de l'avant-plan constituent l'ensemble.

La base de données regroupe toutes les fiches. L'ensemble renferme les fiches extraites de la base de données, créées ou importées à l'écran. La barre titre de la fiche à l'écran donne le nombre total de fiches contenues dans l'ensemble. L'option «Bilan» [F2], si activée, affiche la liste des vedettes principales de chacune des fiches contenues dans l'ensemble.

Latter[©] est plus qu'un outil de saisie, c'est un outil de gestion de fiches. Le terminologie peut rassembler dans un ensemble des

fiches portant sur une même notion, sans égard à la provenance (fiches de dépouillement, fiches Termium®, etc). S'il étudie, par exemple les types de déchets, il peut les étudier, les comparer, les analyser, en faire la synthèse et en créer de nouvelles ou en modifier selon le cas. Celles-ci peuvent être exportées immédiatement ou gardées pour une exportation ultérieure vers Termium®, vers un système de traitement de texte ou vers Publiciel®. Pour faciliter la gestion et la création de fiches, Latter® offre la fonction «Copier/Coller» qui permet à l'utilisateur de copier, par exemple, un contexte et de le coller le nombre de fois désiré sur la même fiche, sur une autre ou sur plusieurs. De plus, la fonction «Recherche» permet de rechercher, d'ajouter ou de remplacer un élément dans un champ donné ou dans toute la fiche de l'ensemble actif.

Prochaine étape : Latter® – Windows

Le Bureau de la traduction est en train de créer une version Windows de Latter® améliorée. Cette version sera enrichie d'ensembles permanents qui faciliteront la gestion des fiches. La fonction «Recherche/Remplacement» sera aussi améliorée et des formats d'impression de vocabulaires ou lexiques multilingues s'ajouteront aux options existantes.

*Louise Claude,
Terminologie et Documentation,
Bureau de la traduction,
Travaux publics et Services
gouvernementaux Canada,
Canada.*

Références – Marques de commerce

PROCOMM Plus®:
Datastorm Technologies, Inc.,
Columbia, MO

Publiciel®:
Direction de la terminologie et de la
documentation, Gouvernement du
Canada

Termium®:
Direction de la terminologie et de la
documentation, Gouvernement du
Canada

Windows®:
Microsoft Corporation, États-Unis

WordPerfect®:
WordPerfect Corporation, Orem, UT

Bibliographie

Claude (Louise) avec la collaboration de
Cloutier (Yvan), 1995: Le manuel de
référence Latter®, Travaux publics et
Services gouvernementaux Canada,
Bureau de la traduction, Direction de la
terminologie et de la documentation,
version provisoire.

Leonhardt (Christine), 1994:
«Termium® and Latter®: An Update»
dans L'Actualité terminologique, vol. 27,
1, Ottawa, Approvisionnement et Services
(1994: 15-18).

Leonhardt (Christine) et Valiquette
(Michèle), janvier 1996: «Termium® et
Latter®: développement et diffusion»,
dans Terminogramme, Numéro 78,
(1996: 12-14).

La fiche LATTER

LATTER 2.1 Édition de fiche n° 1458 1/1

Fiche Édition Ensemble Utilitaires A propos... Sortie

- Projet L : Code de projet LATTER, indispensable pour la gestion de LATTER. Ex. : PRO-SVP MAI 1995
 Domaine L : Domaine LATTER, disponible pour la gestion de LATTER.
 Domaine T : Code de domaine TERMIUM IV Ex. : SHC CAB SCH
 Dom. sec. : Domaine non codé, exporté vers TERMIUM IV.
 Fonds : Indicatif de fonds TERMIUM IV Ex. : TER, APN ou TRA
 Fichier : Indicatif de fichier TERMIUM IV Ex. : A ou B
 Exploitant : Exploitant TERMIUM IV - champ non utilisé, ne rien inscrire dans ce champ.

BLOC LINGUISTIQUE

- Langue : Indicatif de langue du bloc linguistique. Ex. : EN, FR, DE, ES, IT, PT, NL, NO, SV, FI, RO, DA, IU
 Vedette : Terme principal. Appuyer sur [Ins] pour créer le bloc synonyme.
 Source : Appel de source de la vedette.
 Param. : Paramètre de la vedette.
 Abrév. : Abréviation de la vedette. Appuyer sur [Ins] pour créer un autre bloc abréviation.
 Source : Appel de source de l'abréviation.
 Param. : Paramètre de l'abréviation.

BLOC JUSTIFICATION

- Type jus. : Indicatif du type de justification. Ex. : DEF, DÉF, CONT, EX, PHR ou OBS
 Jus. : Justification.
 Source : Appel de source de la justification.
 Terme clé : Terme clé TERMIUM IV.

BLOC RÉSEAU NOTIONNEL

- Qualifié : Terme repère d'un ensemble notionnel.
 Opér. : Opérateur. Expression de la relation unissant deux notions d'un réseau notionnel.
 Lien : Expression qualifiant certains opérateurs.
 Qualifiant : Terme secondaire de l'ensemble notionnel.
 Clé regr. : Clé de regroupement. Champ disponible pour la gestion de LATTER.
 Renvoi : Champ disponible pour la gestion de LATTER.
 Dossier : Indicatif de dossier de synthèse.
 Ens. not. : Ensemble notionnel. Champ disponible pour la gestion de LATTER.
 Aut./lang. : Code d'auteur de première rédaction - bloc linguistique. Champ TERMIUM IV.
 Date/lang. : Date de première rédaction - bloc linguistique. Champ TERMIUM IV.
 Aut. MAJ/1 : Code d'auteur de la dernière mise à jour du bloc linguistique. Champ TERMIUM IV.
 Date MAJ/1 : Date de la dernière mise à jour du bloc linguistique. Champ TERMIUM IV.

BLOC SOURCE

- Appel : Lettre de renvoi à la source.
 Source : Code de source TERMIUM.
 Ann1 : Première année de la source.
 Ann2 : Deuxième année de la source, si nécessaire.
 Volume : Volume de la source.
 Numéro : Numéro de la source.
 Page : Page de la source.
 Non codée : Source en clair, non codable par le Service de codage des sources.

BLOC TECHNIQUE

Projet T	: Code de projet TERMIUM IV.
Formule	: Formule chimique. Champ commun à tous les blocs linguistiques.
Source	: Appel de source de la formule chimique.
Param.	: Paramètre de la formule chimique.
Doc. d'acc	: Indicatif d'un document d'accompagnement (schéma, illustration, tableau ou diagramme, etc.).
Langue dép	: Indicatif de la langue de départ de la fiche LATTER.
ID cr. L	: Code de l'auteur de la fiche LATTER.
Date cr. L	: Date de création de la fiche LATTER.
Date MAJ	: Date de mise à jour de la fiche LATTER.
ID rév. L	: Code de réviseur de la fiche LATTER.
Date soum.	: Date de soumission à la révision de la fiche LATTER.
Date revue	: Date de révision de la fiche LATTER.
Date L à T	: Date d'exportation de la fiche de LATTER à TERMIUM IV.
Synthèse	: Champ disponible pour la gestion LATTER. Terme repère d'un ensemble notionnel.
Bloc-notes	: Champ fourre-tout LATTER, à l'usage de l'utilisateur.

CHAMPS TERMIUM

Nmatri T	: Numéro matricule de la fiche TERMIUM IV.
Date T à L	: Date d'importation de TERMIUM IV à LATTER.
Nom cr. T	: Nom de l'auteur de la fiche TERMIUM IV.
Date cr. T	: Date de création de la fiche TERMIUM IV.
Nom MAJ T	: Nom de l'auteur de la mise à jour de la fiche TERMIUM IV.
Date MAJ T	: Date de la mise à jour de la fiche TERMIUM IV.
Nom rév. T	: Nom du réviseur de la fiche TERMIUM IV.
ID Rév.	: Code du réviseur de la fiche TERMIUM IV.
Rév. MAJ	: Nom du réviseur de la fiche TERMIUM IV.
Nom vérif.	: Nom du vérificateur de la fiche TERMIUM IV.

À une question

de M. Pierre Isabelle,

M. Jean Perron explique qu'aucun test n'a été effectué pour mesurer l'importance du bruit et du silence engendré par le logiciel Adepte-Nomino. Cependant, il signale que c'est par des choix méthodologiques judicieux que les problèmes du silence et du bruit peuvent être contournés, notamment par la constitution de réseaux de termes qui permettent l'élimination du bruit et du silence.

Par ailleurs, M^{me} Caroline de Schaetzen fait remarquer que certains logiciels peuvent être détournés de leur fonction première et pourraient servir non pas à la constitution de réseaux, mais à l'analyse en grappes de proximité de concepts et de dénominations. Ce sont des logiciels qui s'appellent *Cluster Analysis* et qui sont utilisés dans les sciences humaines (analyse du discours), ce qui permettrait de compléter l'utilisation de Adepte-Nomino.

M^{me} Saadia Ait Taleb

signale la difficulté d'accès aux banques de terminologie arabophones en l'absence de réseaux télématiques satisfaisants. Certaines banques sont diffusées sur CD-ROM, alors que d'autres doivent diffuser la terminologie sur disquettes. L'arrivée d'Internet résoudra sans doute le problème.

Le développement et la diffusion des banques de terminologie

Président de séance :
Loïc Depecker
(Délégation générale à la langue française, France)

Thème 3

La banque de données terminologiques TermRom, petite expérience d'une petite banque de terminologie

La présente communication traite de la banque de terminologie BDT, créée par l'Association roumaine de terminologie – TermRom. Elle donne un aperçu du travail accompli par les terminologues et des difficultés rencontrées lors de sa constitution. La structure de la fiche de terminologie y est décrite et des publications, qui sont le résultat de l'exploitation des données, y sont présentées.

Termes-clés:
banque de données;
TermRom; normalisation;
terminologie roumaine.

(1) Micu (Samuil), Şincai (Gheorghe), 1980: *Elementa linguae daco-romanae sive valachicae* [Viena, 1780]-Studiu introductiv, traducerea textelor și note de Mircea Zdrenghea, Cluj-Napoca, Editura Dacia.

En 1992, avec le soutien de l'Union latine, a été constituée l'Association roumaine de terminologie, coordonnée par le professeur Ion Coteanu, membre de l'Académie roumaine à titre de président. Elle devait grouper tous ceux qui avaient des préoccupations concernant la terminologie en Roumanie (tant linguistes que spécialistes dans d'autres domaines d'activité). Son principal objectif était la création d'une banque de terminologie qui soit bientôt accessible aux spécialistes de divers domaines ainsi qu'au public. En ma qualité d'administrateur et de programmeur de cette banque de données, je désire donner un aperçu du travail de terminologie accompli par les terminologues de TermRom et des problèmes que nous avons rencontrés au cours de la constitution de cette banque de terminologie à cause de la situation particulière de la langue roumaine. Le roumain, en tant que langue nationale ayant une diffusion culturelle relativement restreinte, peut être qualifié de «petite» langue par rapport au français et à l'anglais, deux langues ayant le statut de «grandes» langues, de langues internationales. Au cours des trois derniers siècles, en commençant par la première grammaire en caractères latins (au

début, le roumain était écrit à l'aide de l'alphabet cyrillique) écrite en 1780 en latin ⁽¹⁾, le roumain a fait le pas décisif vers son établissement comme langue romane littéraire, reconnue aujourd'hui comme une langue nationale d'État. En ce qui concerne la terminologie, les préoccupations des gens de culture roumaine étaient plus anciennes que la grammaire mentionnée, qui est considérée comme la limite entre l'ancien roumain et la langue roumaine moderne. Étant une petite langue littéraire, fondée sur une culture écrite relativement jeune, le roumain a formé son inventaire lexical terminologique surtout par la voie de l'emprunt notamment aux langues modernes occidentales, voire romanes. La Roumanie ayant une politique linguistique permissive, le processus de normalisation de cette terminologie de spécialité en est encore à ses débuts. Ce qu'on observe en roumain littéraire qui, se trouvant dans la dernière phase de sa modernisation, s'est dirigé vers le latino-roman, surtout vers le français pour l'emprunt lexical, c'est le déplacement de ses sources d'emprunts vers l'anglo-saxon, surtout vers l'anglais et notamment vers sa variante américaine, situation qui pose de grands problèmes de graphie et de prononciation pour de nombreux mots.

On peut exemplifier cette situation par le terme français *ordinateur*, en anglais *computer*, qui est appelé couramment en roumain *calculator*. Dans la terminologie usuelle le terme anglais a supplanté le terme français, même si celui-ci existe dans une forme parfaitement adaptée à la langue roumaine, *ordinator*. Le mot *computer*, écrit comme en anglais, est prononcé dans le langage courant avec *o* [o] et non pas avec [ə] et même avec *-er* au lieu de [(r)] en finale, donc *kompjuter* et non pas [kəm'pju:tə^r].

La prononciation du mot *computer* en concordance avec la graphie roumaine est acceptée tacitement dans les travaux universitaires normatifs, parce que dans le plus important, le *DOOM*⁽²⁾, on ne fait aucune mention de la prononciation, ce qui aurait été nécessaire. Don Matei, vice-président et coordonnateur du groupe TermRom, qui est également responsable de la terminologie roumaine du glossaire d'informatique de TermRom, propose qu'on écrive *compiuter* pour ce terme sélectionné comme «terme problème», c'est-à-dire comme terme qui requiert d'être normalisé sur le plan international.

Les glossaires de TermRom ont, presque chacun, une section de «termes-problème».

Compte tenu des moyens financiers modestes et du personnel restreint qui travaille à la création de notre banque de données, la méthodologie du travail s'est orientée vers le seul objectif qui soit effectivement réalisable. Il s'agit d'un objectif minimal, mais acceptable, à mon avis, dans la première phase de la recherche. Au début, on a désiré

constituer un fonds assez important de termes spécialisés des différents domaines avec leurs équivalents en français et en anglais. On a travaillé à la constitution de ce fonds pendant deux ans, en collectant des termes «en vrac» qui ont été mémorisés dans des fichiers en format Microsoft Word. Le format des fichiers a été conçu d'après le modèle d'une fiche terminologique, chaque champ de la structure de la fiche ayant un identificateur qui permettra la transformation par la suite du fichier en une base de données à proprement parler. Nous avons emmagasiné de cette manière, jusqu'à présent, environ 20 000 termes.

La structure d'une fiche terminologique est assez complexe; elle comprend, en plus du terme vedette et de ses équivalents en sept langues (français, anglais, espagnol, italien, russe, allemand, portugais), les synonymes, les antonymes du terme vedette et de ses équivalents dans d'autres langues (s'il en existe), les termes dérivés, des indications sur la flexion, les variantes lexicales, les termes génériques, le sens large du

terme, le terme scientifique (s'il existe), la représentation symbolique (si elle existe), l'abréviation, la prononciation, la catégorie grammaticale du terme et des équivalents (y compris les syntagmes), le contexte, la zone (le pays), l'étymologie, la bibliographie, la définition en roumain et dans les autres langues, des notes, le statut du terme (terme validé ou pas), le domaine, le nom du terminologue, la date, des renseignements sur des modifications ultérieures et la contribution du terminologue.

Dans cette première phase de l'élaboration de la banque de données, le groupe de TermRom a choisi la terminologie roumaine consignée dans les dictionnaires et glossaires roumains publiés au cours des dix dernières années.

La banque de données offre de la sorte une image sur la relation entre la terminologie roumaine et la terminologie internationale telle qu'elle est reflétée dans les dictionnaires. Il va de soi que cette première étape sera suivie de l'extraction des termes utilisés

(2) Dicționarul ortografic, ortoepic și morfologic al limbii române, 1982-1989, Bucarest, Editura Academiei, (élaboré par l'Institut de linguistique de Bucarest et approuvé par l'Académie Roumaine).

*AM

{0}	contract de cauțiune	{cg}	s.m.sg.
{asc}	cauțiune, contract de	{7}	contrato de caución
{cg}	n. sg.	{asc}	caucin, contrato de
{f}	contracte de cauțiune	{cg}	s.m.sg.
{asc}	cauțiune, contracte de	{dm}	com.
{cg}	n. pl.	{d}	Contract portivit căruia o
{1}	contract of guarantee		persoană numită girant se
{asc}	guarantee, contract of		obligă să îndeplinească
{cg}	s. sg.		obligapiile unui debitor
{f}	contracts of guarantees		către creditorul acestuia, în
{asc}	guarantee, contracts of		cazul în care debitorul nu-și
{cg}	s. pl.		poate chita datoriile.
{2}	cautionnement	{g}	contract
{cg}	s.m.sg.	{s}	p
{f}	cautionnements	{b}	MDE
{cg}	s.m.pl.	{r}	AM
{3}	Bürgschaft	{z}	sep. 95

Fiche terminologique telle qu'elle apparaît en format Microsoft Word

effectivement dans les textes roumains de spécialité; cependant, cette première étape d'information fondée sur les dictionnaires est imposée par les nécessités pratiques d'un objectif d'une importance nationale comme l'est à mon avis la création d'une banque de données terminologiques du roumain.

De nombreuses fiches terminologiques sont encore incomplètes; elles seront groupées, corrigées et validées par des linguistes et par des spécialistes des différents domaines, par les membres et les collaborateurs de TermRom. Pour que cette quantité considérable d'information terminologique soit utile et facilement utilisable, on a dû trouver une solution convenable pour la mémoriser dans une base de données ayant une structure correspondant à la structure de notre fiche terminologique et qui soit exploitée dans un environnement qui reconnaisse les signes diacritiques spécifiques du roumain. Il existe une grande variété de logiciels de terminologie à l'étranger, mais aucun en Roumanie. On pourrait employer certains de ces logiciels s'ils étaient adaptés pour faire le tri alphabétique en tenant compte des signes diacritiques propres au roumain. Pour se procurer et modifier un tel logiciel, il faudrait des ressources financières dont TermRom ne dispose pas encore. Nous avons été obligés, par conséquent, d'employer un logiciel de gestion de bases de données que nous possédions, conçu pour Windows, Foxpro for Windows (nous travaillons avec Windows CE qui a des polices pour le roumain), et de construire nous-mêmes nos programmes de tri. Pour une recherche de l'information plus rapide et plus facile, nous avons créé plusieurs petites bases de données liées aux mots-clés communs.

Un premier résultat de l'exploitation de ces données est représenté par l'impression de petits glossaires de terminologie (*Glossaire*

des termes d'informatique, Glossaire des termes financiers et bancaires, Glossaire des termes de commerce, Glossaire des terminologies, Glossaire des termes d'énergétique) groupant des informations extraites directement de la banque de données à l'aide des programmes que nous avons conçus pour transformer les contenus des champs choisis des bases de données en textes pouvant être formatés et imprimés. Nous travaillons maintenant à l'impression du *Glossaire des termes de la sociologie* et du *Glossaire de terminologie linguistique*.

données regroupe par conséquent des noms tant pour des objets réels abstraits (processus, actions, relations) que pour des objets réels concrets (appareils, pièces, instruments techniques, etc.).

Actuellement nous travaillons à l'élaboration des programmes nécessaires à l'exploitation de la banque de données et à la création d'une interface à menus en roumain qui soient plus accessibles pour les usagers.

En Roumanie les préoccupations liées à la terminologie existent depuis longtemps. Les recherches ont été

achita o factură = *to settle an invoice* (en.) = *régler une facture* (fr.) = *liquidar una factura* (sp.) = *eine Rechnung begleichen* (ger.)

A plăți o factură cu bani lichizi, cec, virament în cont etc.

achiziție s.f. = *purchase* (en.) = *acquisition* (fr.) = *adquisición* (sp.) = *Ankauf* (ger.)

Termen care semnifică o acțiune economică din sfera comerțului privind procurarea, cumpărarea de bunuri, urmată de trecerea în proprietatea celui care face achiziția, a obiectului achiziționat.

[Spre deosebire de cumpărare, achiziția se poate realiza nu numai cu titlu oneros ci și cu titlu gratuit.]

ACORDUL MONETAR EUROPEAN = *European Monetary Agreement* (EMA)

(en.) = *Accord monétaire européen* (AME) (fr.) = *Acuerdo Monetario Europeo* (AME)

(sp.) = *Europäisches Währungsabkommen* (EWA) (ger.)

Acord de plăți încheiat între țările membre ale Uniunii Europene de Plăți (UEP), care conține dispozițiile privind decontările și creditele reciproce.

Extrait du *Glossaire des termes de commerce* (roumain, anglais, français, espagnol)

En examinant l'inventaire roumain de ces glossaires on peut observer que le groupe de TermRom a intégré dans la sphère de la terminologie des composantes lexicales représentant aussi bien des termes techniques et scientifiques que des nomenclatures⁽³⁾. La banque de

poursuivies par des groupes de linguistes ainsi que par des spécialistes dans d'autres domaines, mais les résultats concrets de ces recherches imposent un examen attentif pour pouvoir développer l'étude de la terminologie en tant que discipline scientifique.

Les conclusions d'une enquête effectuée l'année dernière dans les universités, les institutions culturelles et les entreprises de Roumanie ont démontré que même si on ressent toujours davantage la nécessité d'une terminologie rigoureuse dans les différents domaines d'activité qui

(3) Cf. Dubois (Jean), 1974: Dictionnaire de linguistique, Paris, Larousse, s.v.; Vascenco (Victor), 1975: Probleme de terminologie lingvistică, București, Editura Științifică și Enciclopedică, p. 5-9.

sont de plus en plus développés dans notre pays, il n'existe pas une préoccupation constante pour ordonner, systématiser et normaliser ce secteur spécial de la linguistique. Les tentatives pour créer des centres de terminologie dans les universités sont plutôt le résultat de l'initiative de groupes restreints d'enthousiastes (à l'Académie des études économiques, à l'Université polytechnique de Bucarest, à la Faculté des lettres). Dans les domaines de la télécommunication et de l'informatique, dans le domaine financier, etc., on ressent le besoin réel d'une terminologie adéquate. L'informatisation de l'économie roumaine, après 1989, a obligé de larges catégories de gens de différentes professions à apprendre à utiliser l'ordinateur. Certains d'entre eux ne connaissaient pas l'anglais, et comme la terminologie de l'informatique en Roumanie est presque totalement anglaise, comprendre le fonctionnement de l'ordinateur leur a été très difficile; beaucoup ont été obligés d'apprendre de façon mécanique l'usage de l'ordinateur, sans être capables de tirer profit de tous les avantages que les nouveaux programmes ont mis à la disposition des usagers. C'est une situation à laquelle je me suis heurtée à l'occasion des cours de formation que j'ai donnés.

L'existence d'une banque de données en Roumanie qui soit utilisée par le public, même si pour le moment elle ne couvre que quelque domaines, est évidemment essentielle.

Pour les spécialistes des différents domaines d'activité, la banque de données offre une classification des sens et des rapports entre la terminologie roumaine et la terminologie d'autres langues. On pourrait dresser des programmes de vocabulaires pour l'enseignement du roumain ainsi que des programmes d'étude pour les autres disciplines ou pour les termes que les élèves doivent

posséder dans leur bagage général. Pour les linguistes, la banque de terminologie permet d'examiner les relations entre le roumain et les autres langues modernes dans le domaine du vocabulaire de spécialité. L'existence de la banque de données terminologique est utile pour normaliser le roumain littéraire actuel en réalisant un inventaire terminologique roumain standard, facile à consulter et à actualiser. On peut éliminer de la sorte les ambiguïtés de la langue écrite et parlée et des documents officiels et on peut protéger la langue nationale.

Le développement de la banque de données BDT a pour objectif principal, dans un avenir prochain, l'adjonction de nouveaux termes en même temps que la correction des fiches terminologiques existantes. Nous continuerons d'imprimer de petits glossaires contenant des extraits de la terminologie présente dans la banque de terminologie. Nous envisagerons également la possibilité d'accéder au réseau Internet pour pouvoir avoir des échanges avec des usagers externes et créer des liens qui, nous espérons, seront de plus en plus étroits et mutuellement profitables.

*Mihaela Ursu,
TermRom, l'Association roumaine
de terminologie,
Roumanie.*

Étude comparée de dictionnaires électroniques et de banques de termes

Les dictionnaires électroniques sont très diffusés; Elsevier a notamment annoncé la parution sur DON de sa collection de dictionnaires multilingues spécialisés. Les dictionnaires en plusieurs volumes de la langue générale ne sortent plus sur papier. La présentation de ces ouvrages électroniques est-elle similaire? Exploite-t-elle de la même manière le contenu? Comme à toutes les bases de données, les butineurs d'Internet ouvrent un public neuf aux banques de termes. Le contenu de ces banques, dont la présentation diffère des dictionnaires, peut-il rivaliser avec celui des dictionnaires électroniques spécialisés?

Termes-clés:
banques de termes; dictionnaires électroniques spécialisés; ergonomie; analyse de contenu.

1 Introduction

1.1 Étude menée

Un premier volet de cette recherche compare les fonctions de consultation offertes par les banques de terminologie à celles des logiciels de trois dictionnaires électroniques. Nous avons ajouté un gestionnaire de vocabulaires à ces dictionnaires. Ce type de logiciel est destiné à l'encodage, la consultation, les modifications, l'impression et l'échange des vocabulaires individuels élaborés par les traducteurs. Ces petites collections étaient qualifiées de «fichiers boîtes à chaussures» à l'époque où leur support était des fiches en carton. Nous les avons inclus parce que de plus en plus de ces gestionnaires sont fournis avec des dictionnaires prêts à l'emploi.

La seconde partie de cette communication compare le contenu d'un dictionnaire électronique, le Grand Dictionnaire Bilingue d'Informatique Dicosoft de la firme Mysoft, avec les termes de l'informatique que comportent les banques de données terminologiques Eurodicautom et Termium. Les termes uni- et plurilexématiques cherchés dans ces fonds terminologiques sont extraits de livres à caractère didactique, relatifs au matériel et aux logiciels de micro-informatique, ainsi que de revues spécialisées en micro-informatique. Ces publications sont rédigées en

anglais et en français. Les critères d'élaboration du corpus textuel ont été les suivants:

- Registre de vocabulaire limité à la micro-informatique, qui fait l'objet d'une littérature abondante et de traductions nombreuses dans le marché commercial de la traduction;
- Densité élevée de termes récents de l'informatique;
- Diversité des sous-domaines couverts par les textes;
- Diversité des niveaux de technicité de ces termes.

Le lecteur notera que les termes cherchés proviennent de six sources bibliographiques seulement. Nous regrettons en outre de ne pas avoir eu l'occasion de consulter la BTQ du gouvernement québécois. Quant à la version dont nous disposions pour Termium, elle date de mai 1992. Le volume du corpus de termes cherchés donne malgré tout de l'intérêt à cette étude comparée.

2 Présentation des logiciels

Mis sur le marché en 1994, Dicosoft est un dictionnaire électronique anglais/français - français/anglais pour Windows, à installer sur disque rigide. Il compte 35 000 vedettes anglaises et 35 000 vedettes françaises. Aucune information n'est donnée sur la provenance de ce lexique terme à terme d'informatique.

Comme Dicosoft, le dictionnaire unidirectionnel Wordswitch de langue générale, bilingue néerlandais/français et tournant sous Windows, est commercialisé dans les boutiques

d'informatique. Il en existe une version pour disques rigides et une autre pour les disques optiques numériques. Nous avons utilisé la version sur disque rigide pour la présente étude. Contrairement à Dicosoft, Wordswitch permet d'ajouter et de supprimer des mots ajoutés par l'utilisateur dans le dictionnaire (le vocabulaire du dictionnaire proprement dit est verrouillé contre l'effacement). Aucune information n'est donnée sur les auteurs de Wordswitch, commercialisé en Belgique pour le français et l'anglais, l'allemand, l'espagnol, l'italien, l'espagnol, le néerlandais.

Larousse a, quant à lui, édité plusieurs dictionnaires électroniques bilingues et unilingues pour disque rigide ou sur DON, notamment un dictionnaire d'anglais commercial et le dictionnaire Larousse bilingue général bidirectionnel, anglais/français et français/anglais. Nous avons étudié la version sur disque rigide du Larousse bilingue général. L'utilisateur peut y chercher un mot, sauter à un autre mot (synonyme, par exemple) du même dictionnaire ou d'un autre dictionnaire Larousse, si un renvoi a été prévu. Larousse distingue les acceptions des mots polysémiques.

Ni fenêtres, ni souris, ni boutons, ni menus déroulants dans la banque Eurodicautom de la Commission de l'Union européenne. Les codes sont cependant parfaitement mnémoniques et leur frappe peut s'avérer plus rapide à l'usage que certaines sélections successives par des clics! La lettre C, suivie de la pression de la touche retour-chariot, fait par exemple afficher la fiche suivante, lorsque Eurodicautom en comporte plus d'une pour la vedette demandée par l'utilisateur, E clôt la session d'interrogation de la banque, etc.

Termium allie la rigueur à la présentation conviviale des gestionnaires de bases de données

pour les DON tournant sous DOS: affichage permanent des menus, des codes les plus courants et des commandes qui y correspondent, couleurs paramétrables par l'utilisateur, etc. Trois portions d'écrans affichent respectivement la langue du terme cherché, la liste alphabétique de tous les termes alphabétiquement proches de la vedette demandée et contenus dans la banque pour cette langue ainsi que l'intitulé du ou des termes ou syntagmes dont l'utilisateur a demandé l'affichage. Cette présentation permet à l'utilisateur de se raviser avant l'affichage des résultats de la recherche effectuée dans la banque.

Avec le gestionnaire de vocabulaires Multiterm pour Windows, l'utilisateur peut concevoir à son gré le contenu, la taille et la présentation de ses fiches. L'affichage coloré cache un gestionnaire de fiches terminologiques flexible et riche en commandes. Des fonctions fréquemment utilisées s'obtiennent par un clic! du bouton droit de la souris dans certaines zones de la fenêtre d'affichage (menus dits

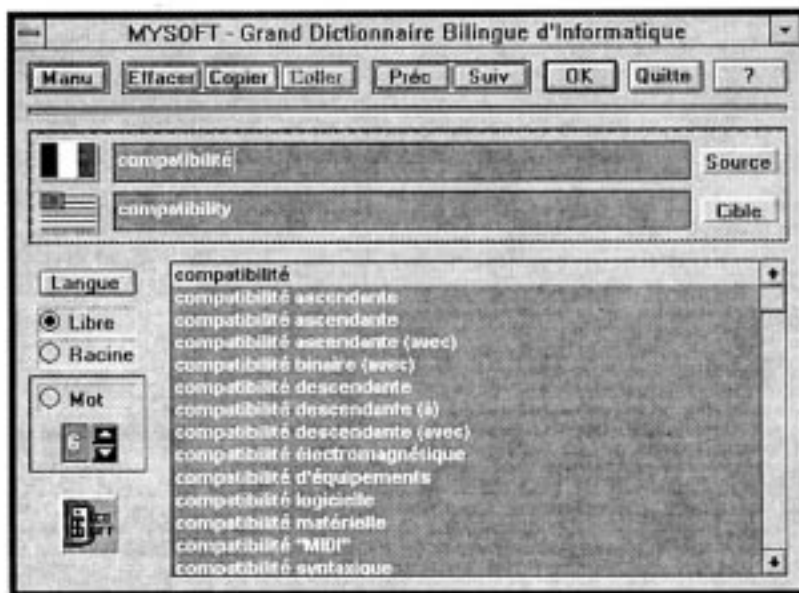
« contextuels », typiques de la dernière génération des logiciels pour Windows).

3 Comparaison des contenus

3.1 Modes de recherche des vedettes

Le logiciel donnant accès au contenu de Dicosoft offre trois filtres aux termes sélectionnés en réponse aux requêtes de l'utilisateur. Ces requêtes sont la liste de termes parmi lesquels l'utilisateur peut choisir celui ou ceux dont il veut voir afficher l'équivalent dans l'autre langue :

– L'option dite « Libre » propose tous les syntagmes contenant le mot ou le syntagme demandé par l'utilisateur, indépendamment de l'ordre de dactylographie des mots lors de la requête. Dans cette liste, l'utilisateur désignera le ou les termes dont il veut voir l'équivalent dans la langue d'arrivée. Si l'utilisateur demande par exemple *base de données*, il verra



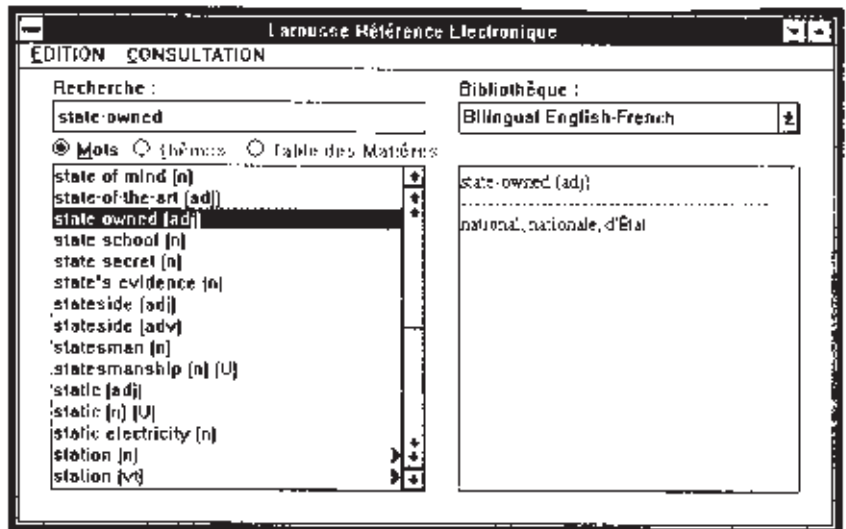
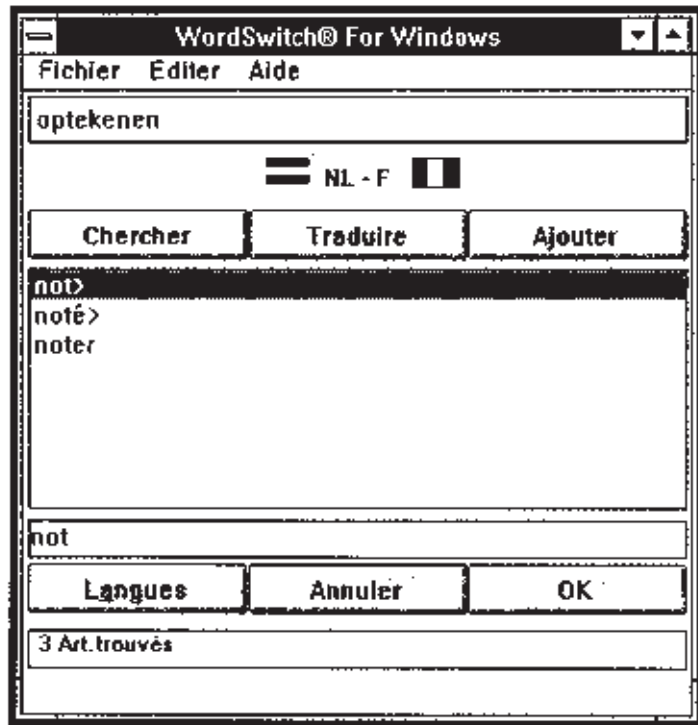
s'afficher une liste comportant *administration de base de données, automatisation des données de base, base de données active, données de base, GDB (gestion de base de données)* etc. Il y pointera le terme dont il veut voir l'équivalent en anglais;

– Lors de la recherche par le bouton «Racine», tous les termes et syntagmes commençant par la chaîne de caractères demandée par l'utilisateur sont tronqués à droite.

Cette chaîne doit comporter au moins deux caractères. Exemple : si l'utilisateur demande *chiffre*, c'est dans la liste des termes *chiffre, chiffre codé primaire, chiffre décimal, chiffrage, chiffrage de jonction, chiffrer, chiffres successifs, chiffreuse*, etc., qu'il pourra désigner au logiciel un terme dont il veut voir l'équivalent en anglais;

– Le mode de recherche «Mot» liste le mot demandé ainsi que les syntagmes contenant ce mot. L'utilisateur indique la longueur maximale (de 1 à 6 mots) des syntagmes éventuels parmi lesquels il souhaite choisir. Exemple : *conversationnel*, si l'utilisateur a limité la taille des syntagmes à deux mots, fera afficher *conversationnel, mode conversationnel, solution en conversationnel, système conversationnel, terminal conversationnel, traitement conversationnel*, etc.

– Le logiciel de Wordswitch se réduit à deux fonctions de recherche principales : troncature à gauche ou recherche littérale de la chaîne de caractères demandée par l'utilisateur. Le nombre de fonctions de consultation est moins réduit dans le logiciel sous-jacent à Larousse que dans le logiciel de Wordswitch. Le même système de liste que dans Dicosoft a été prévu : la validation du mot à chercher par la touche retour-chariot donne toujours accès à la liste alphabétique des mots et termes voisins du mot inconnu ou méconnu. Larousse reconnaît les mots tronqués



d'Eurodicautom: DOC (numéro de la fiche active), PAGE (n° de la page-écran), BE (bureau de terminologie émetteur), TY (code de la collection), NI (numéro d'identification de la fiche), DATE (date de création ou de mise à jour de la fiche), CF (code de degré de fiabilité), CM (domaine), AU (auteur de la fiche), code de la langue (DA, FR, DE...), CR (toutes les langues), VE (vedette), PH (phrase, c'est-à-dire contexte), MC (mot-clé), DF (définition), AB (abréviation), RF (références bibliographiques), SI (signification d'une abréviation), PS (pays de provenance ou d'usage pour l'abréviation), NT (différentes notes explicatives), ainsi que les codes-matières (domaines) et les codes des bureaux émetteurs.

3.3 Nombre de termes trouvés

Ce chapitre donne les résultats chiffrés de la consultation de Dicosoft, Eurodicautom et Termium. La liste alphabétique des termes cherchés pour le français ainsi qu'un extrait de l'échantillon de textes ayant servi de base à notre recherche sont dans les annexes I et II.

Dans ce tableau, la colonne «Exactement» totalise les termes qui, parmi ceux dont nous avons demandé l'équivalent au dictionnaire, y sont effectivement présents et sous le libellé précis que nous avons dactylographié lors de la requête.

D'autre part, les textes français charriant de la terminologie d'informatique sont souvent truffés d'emprunts à l'anglais. Il fallait donc étudier ces emprunts, qui possèdent la plupart du temps leurs équivalents français. Nous avons donc réservé une rubrique aux termes pour lesquels le sens de la consultation a dû être inversé. C'est la colonne «Ang/Fr».

Enfin, la colonne «Variante» reprend les termes pour lesquels nous n'avons trouvé d'équivalents que pour une variante orthographique, typographique ou morphologique du terme en langue source.

Exemples : à la place de l'équivalent de *boîte aux lettres*, nous avons trouvé l'équivalent de *boîte à lettres*, à savoir : *mailbox*; pour *canal DMA* (*DMA* signifiant *Direct Memory Access*), celui de *canal d'accès direct à la mémoire* : *direct memory access channel*; pour *cristaux liquides*, celui de *cristal liquide* : *liquid crystal*; pour *chiffage*, celui de *chiffrement* : *ciphering*; pour *synchronised*, celui de *synchronized* : *synchronisé*.

La rubrique «Explication» annonce un acronyme pour lequel le dictionnaire ne propose aucun équivalent, mais uniquement le syntagme entier désigné par le sigle. Exemples : *VESA* : *Video Electronics Standards Association*; *ISA* : *Interrupt Storage Area*.

«Polysémiques» qualifie des termes polysémiques en langue source mais pas en langue cible. Pour eux, un ou plusieurs équivalents ont été trouvés, mais nous ignorons

l'acception précise du terme de la langue source traduit par cette ou ces dénominations. Exemples : *piloter* : *to control*; *palette* : *vane*; *palette*, *sample board*.

La rubrique «Constituants» comptabilise les syntagmes nominaux dont nous n'avons trouvé que les éléments constitutifs ou certains de ces éléments. Elle inclut aussi des termes proches qui pourraient aider l'utilisateur à trouver par similitude un équivalent pour le terme concerné. Pour *logiciel intégré*, nous avons par exemple trouvé l'équivalent de *logiciel*, à savoir *program package*, *software*, *soft*, etc., ainsi que l'équivalent d'*intégré*, qui est *built-in*, *integrated*; pour *local*, nous avons trouvé l'équivalent d'*opération en ligne ou locale*, à savoir *line or local operation*; pour *piste concentrique*, nous avons trouvé l'équivalent de *piste* : *track* ainsi que celui de *concentrique* : *concentric* et l'équivalent de *sillon concentrique* : *locked groove*, *concentric groove*.

Enfin, la colonne «Néant» reprend les termes absents du dictionnaire ou de la banque, que ce soit pour leur libellé précis, partiel ou approximatif.

Les pourcentages figurant à droite de chaque rubrique sont calculés par rapport au total des termes demandés (377 termes anglais dont nous voulions l'équivalent français, par exemple). Le pourcentage relativement faible de variantes trouvées chez Eurodicautom s'explique par le fait que la banque de données ne donne accès qu'au terme

	Exactement		Variante		Explication		Polysémiques		Constituants		Néant		Total	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Dicosoft	339	57	15	2,5	0	0	4	0,75	206	34,5	31	5,25	595	100
Eurodicautom	296	49,75	12	2	14	2,5	30	5	209	35	34	5,75	595	100
Termium	328	55	26	4,25	3	0,5	23	4	172	29	43	7,25	595	100
Moyenne	321	54	17,65	3	5,65	1	19	3,25	195,65	32,75	36	6	595	100

Total des termes trouvés dans les sens français vers anglais et anglais vers français.

recherché dans son libellé littéral ainsi qu'aux syntagmes nominaux formés à partir de cette chaîne. La troncature dans Eurodicautom oblige à faire afficher immédiatement et successivement l'entière de toutes les fiches dont la vedette commence par les mêmes lettres que celles qu'a précisées l'utilisateur, au contraire de la fonction «Libre» de Dicosoft. Ces fiches apparaissent dans l'ordre chronologique de leur dactylographie au sein de la banque. Termium se déplace quant à lui dans la liste-index alphabétique de ses vedettes au fur et à mesure de la frappe du mot cherché. Cette liste permet donc également de repérer un terme proche et d'en faire afficher la fiche.

Les chiffres obtenus séparément pour la direction français/anglais et anglais/français sont similaires pour Dicosoft. Ce dictionnaire est en effet élaboré «en miroir», c'est-à-dire que tout terme anglais a son équivalent français et vice versa. Dans ce dictionnaire, toutes les abréviations sont expliquées mais elles ont également toujours leur équivalent dans l'autre langue. Ce n'est pas le cas pour Eurodicautom et Termium, qui possèdent toutes deux des fiches unilingues (rares dans Termium), voire plurilingues dans le cas d'Eurodicautom, mais ces fiches ne comportent pas forcément la langue cible recherchée.

Il était prévisible qu'un dictionnaire monodomaine comme Dicosoft ne comporterait qu'un faible pourcentage de termes polysémiques: les termes qui s'y trouvent appartiennent nécessairement au domaine de l'informatique. Le pourcentage plus élevé d'Eurodicautom est dû à la géométrie variable de ses fiches. Un certain nombre d'entre elles comportent uniquement le terme et sa traduction, sans autre indication, donc sans information sur le domaine d'utilisation. Le domaine précisé par

nous, l'automatisation, recouvre en outre une série de sous-domaines.

Quant aux termes non trouvés, les trois dictionnaires donnent à peu près les mêmes chiffres. Le nombre élevé de termes trouvés dans Eurodicautom montre la priorité donnée par la Commission des Communautés européennes aux technologies nouvelles (dans le cadre des projets Esprit-LRE). Dicosoft, mis sur le marché en 1994 et spécialisé en informatique, n'est donc pas la source la plus complète. Mais on se souviendra aussi que notre version de Termium a deux ans, soit une éternité pour l'informatique!

Quant aux équivalents en toutes lettres et aux explications des abréviations, c'est Eurodicautom qui s'est avérée la plus riche, grâce à sa banque d'abréviations. Celle-ci est bien utile pour les sigles et les acronymes, qui souvent ne possèdent pas d'équivalents en informatique ou pas d'équivalents adoptés par les usagers. Nous avons vu que la structure de Dicosoft ne permettait pas d'obtenir ce genre de renseignements, tandis que Termium, privilégiant l'équivalence d'une langue à l'autre, possède peu de fiches unilingues.

3.4 Ventilation par domaines des vocables trouvés

Les 595 termes ont été répartis dans les 4 domaines dont traitait le corpus. Nous avons donné une

	Termes trouvés		Trouvés en partie		Non trouvés		Total	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Dicosoft	106	62,5	53	31,5	10	6	169	100
Eurodicautom	87	51,5	72	42,5	10	6	169	100
Termium	106	62,75	55	32,5	8	4,75	169	100
Moyenne	99,75	59	60	35,5	9,25	5,5	169	100

Domaine des logiciels et applications, total des termes trouvés dans le sens français vers anglais et anglais vers français.

extension assez large à ces domaines en raison de la taille modeste de notre échantillon.

Comme certains termes appartiennent à deux ou plusieurs domaines, les totaux cumulés de chaque tableau dépassent 595. Certains des tableaux ci-dessous alignent les chiffres d'échantillons réduits (logiciels et applications: 35, périphériques: 39, télécommunications et réseaux: 29). Leurs résultats sont donc sans valeur statistique.

Par «Termes trouvés», nous entendons les termes trouvés soit dans leur libellé précis soit sous une variante orthographique, morphologique ou typographique. Les termes «Trouvés en partie» sont une ou plusieurs acceptions de termes polysémiques, un ou plusieurs constituants de syntagmes ou encore la dénomination complète d'abréviations trouvées. Une colonne «Non trouvés» totalise les termes absents du dictionnaire, que ce soit pour tout ou partie de leurs constituants.

La moyenne des deux sens de consultation montre les chiffres inférieurs d'Eurodicautom pour les termes et syntagmes trouvés dans leur formulation exacte, ce que compensent les constituants de syntagmes. Dans Termium, le pourcentage de termes non trouvés est un peu inférieur à celui des deux autres sources, mais le nombre de termes trouvés dans leur libellé exact est équivalent à celui de Dicosoft.

	Termes trouvés		Trouvés en partie		Non trouvés		Total	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Dicosoft	167	61,25	96	35,25	10	3,5	273	100
Eurodicautom	143	52,5	116	42,5	14	5	273	100
Termium	161	59	92	33,5	20	7,5	273	100
Moyenne	157	57,5	101,25	37	14,75	5,5	273	100

Domaine du matériel de micro-informatique, total des termes trouvés dans le sens français vers anglais et anglais vers français

Pour ce tableau, les chiffres moins élevés d'Eurodicautom et de Termium pour le sens anglais/français sont compensés par les meilleurs résultats obtenus pour l'autre sens de consultation. Grâce à eux, les résultats sont relativement similaires pour Termium et Dicosoft. Eurodicautom est en léger recul, qui est comme toujours compensé par des constituants nombreux.

Termium comporte bon nombre des termes relatifs au matériel périphérique (10% de plus que Dicosoft pour les termes trouvés dans leur formulation exacte). Le nombre de termes inconnus est, chez lui, similaire aux deux autres sources, ce qui s'explique par le nombre de constituants trouvés chez Dicosoft et Eurodicautom.

	Termes trouvés		Trouvés en partie		Non trouvés		Total	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Dicosoft	59	57,75	40	39,25	3	3	102	100
Eurodicautom	60	58,75	40	39,25	2	2	102	100
Termium	69	67,5	31	30,5	2	2	102	100
Moyenne	62,75	61,35	37	36,35	2,25	2,35	102	100

Domaine du matériel périphérique, total des termes trouvés dans le sens français vers anglais et anglais vers français

	Termes trouvés		Trouvés en partie		Non trouvés		Total	
	Nombre	%	Nombre	%	Nombre	%	Nombre	%
Dicosoft	60	60	31	31	9	9	100	100
Eurodicautom	52	52	41	41	7	7	100	100
Termium	52	52	34	34	14	14	100	100
Moyenne	54,75	54,75	35,25	35,25	10	10	100	100

Domaine des télécommunications et réseaux, total des termes trouvés dans le sens français vers anglais et anglais vers français

En télécommunications et réseaux, Dicosoft a 8% de plus que ses homologues, quoique le nombre de termes non trouvés soit moins

grand dans Eurodicautom. Termium est un peu moins fourni, sans doute en raison de l'ancienneté de notre version.

4 Interface avec le traitement de texte

Lorsqu'ils tournent sous Windows, un grand avantage des dictionnaires pour disque rigide est leur lien avec le traitement de texte. Ils permettent d'importer, de faire «atterrir» dans un texte, l'entrée cherchée ou n'importe quelle rubrique d'une fiche de ce dictionnaire. Le texte chargé dans le logiciel de traitement de texte en même temps que le dictionnaire sera le plus souvent une traduction, en cours d'élaboration. Cette importation de portions de fiche requiert plus ou moins de manipulations selon les dictionnaires.

4.1 Facilité du va-et-vient entre dictionnaire et texteur

	Nombre de touches ou de clics!	
	Avec macro	Sans macro
Dicosoft	4	4
Eurodicautom	-	-
Termium	-	-
Multiterm	-	3
Wordswitch	4	7
Larousse	8	9

Nombre de manipulations requises pour remplacer un terme, apparaissant dans un document chargé sur traitement de texte, par son équivalent dans l'autre langue.

Le nombre de manipulations est plus ou moins important selon le dictionnaire électronique consulté (au minimum 3, pour Multiterm, au maximum 9, pour Larousse). La macro mentionnée dans le tableau est une macrocommande programmée dans le langage de programmation WordBasic du traitement de texte Word. Cette macro réduit le nombre de manipulations requises pour «couper/coller» une dénomination,

c'est-à-dire aller du traitement de texte au dictionnaire, en copier un fragment, revenir au traitement de texte et insérer dans le document ouvert cette portion de texte copiée. Rappelons qu'une macrocommande est un fichier comportant des commandes de traitement de texte (et éventuellement des instructions de programmation, ce qui n'est pas le cas ici). Ces commandes s'exécutent l'une après l'autre, sans interruption, dès que ce fichier de macros, dit «exécutable», est appelé.

Les deux logiciels les mieux interfacés avec le traitement de texte sont Dicosoft et Multiterm. L'interface de Wordswitch et Larousse avec le traitement de texte est moins

étroite, plus classique : elle recourt au mode «copier/coller» par le presse-papiers (mémoire-tampon partagée par tous les logiciels chargés en même temps sous Windows). Cependant, même si cette interface a mieux été programmée, et est donc plus élaborée dans Wordswitch, elle est encore améliorable par une macrocommande simple.

On le voit, le passage du dictionnaire au traitement de texte requiert un grand nombre de commandes chez Larousse. En outre, la sélection de plusieurs mots dans le dictionnaire, par exemple une définition, afin de les exporter vers le traitement de texte n'est pas aisée (surbrillance tardant à suivre les

mouvements de pointage prolongés de la souris).

L'interface de Multiterm est la plus sophistiquée. Elle affiche dans Word sa propre barre d'icônes-outils.

Cette barre d'outils permet à l'utilisateur de consulter et feuilleter les dictionnaires tournant sous Multiterm sans sortir du traitement de texte. Pour la recherche :

- D'une fiche d'une vedette précise : clic! de souris sur l'icône de la loupe de la barre d'outils Multiterm, dactylographie du terme à chercher, puis clic! sur le bouton «OK»;
- D'une fiche active : clic! de souris sur l'icône ad hoc (représentant un point d'interrogation) de la barre d'outils Multiterm;

INSERTION D'UN ÉQUIVALENT TROUVÉ DANS LES DICTIONNAIRES ÉLECTRONIQUES

Dicosoft, en mode automatique	Multiterm	Wordswitch, avec macrocommande	Larousse, avec macrocommande
<ol style="list-style-type: none"> 1. Sélection, dans le texte en cours de frappe ou de correction, du/des termes à chercher; 2. Clic! sur l'icône «copier» ou pression de la combinaison de touches Ctrl + C (<i>copiant le mot inconnu dans le presse-papiers et lançant la consultation du dictionnaire, pour le terme sélectionné</i>); Éventuellement, sélection d'une autre vedette dans la liste affichée par le logiciel. 3. Alt + Tab (<i>entraînant le retour au traitement de texte</i>); 4. Clic! sur l'icône «coller» ou Ctrl + V (<i>insertion par le logiciel de l'équivalent du terme dans le traitement de texte, à la place du terme en langue source</i>). 	<ol style="list-style-type: none"> 1. Sélection du terme inconnu dans le document, au sein du traitement de texte; 2. Clic! de souris sur l'icône de la loupe de la barre d'outils Multiterm (<i>pour faire afficher au bas de l'écran du texte l'équivalent de ce terme</i>); 3. Clic! dans le traitement de texte sur une icône («smiley») de la barre d'outils Multiterm (<i>pour insérer cet équivalent dans le texte lui-même</i>). 	<ol style="list-style-type: none"> 1. Clic! de souris à droite du terme inconnu, dans le document en cours de frappe ou de correction; 2. Pression de Alt + Y (<i>lançant la macrocommande qui copie le mot dans le «presse-papiers» puis qui consulte le dictionnaire</i>); 3. Clic! dans la boîte réservée au mot à faire chercher par le logiciel du dictionnaire (<i>lançant l'insertion de ce mot dans la zone de frappe du mot à chercher puis l'affichage de son équivalent</i>); Clic! de souris sur le bouton «OK» ou double-clic sur un autre équivalent de la liste affichée par le logiciel, ou encore pression sur la touche retour-chariot (<i>lançant l'insertion de l'équivalent dans le traitement de texte, à la place du terme en langue source</i>). 	<ol style="list-style-type: none"> 1. Clic!, dans le document en cours de frappe ou de correction, à droite du terme inconnu; Alt + Y (<i>lançant la macrocommande qui copie le terme dans le «presse-papiers» puis qui lance la consultation du dictionnaire</i>); 2. Clic! dans le champ «Recherche» si ce champ est vide ou sélection du mot inséré à cet endroit lors de la consultation précédente du dictionnaire; Ctrl + V (<i>lançant l'insertion du terme cherché dans la fenêtre de consultation du dictionnaire</i>); 3. Pression sur la touche retour-chariot (<i>pour confirmer le choix du terme à chercher</i>); 4. - Soit : deuxième retour-chariot si le mot sélectionné automatiquement par le logiciel convient; - Soit : sélection d'un autre mot dans la liste alphabétique des vedettes du dictionnaire et retour-chariot (<i>lançant l'affichage de l'équivalent du terme demandé</i>); 5. Éventuellement, clic! sur une autre acception du mot demandé, dans le cas de termes polysémiques (exemple : conflit : 1. situation tendue, 2. entre États); 6. - Dans le cas d'un équivalent en un mot : sélection de l'équivalent par double-clic de souris; - Dans le cas d'un syntagme : sélection avec la souris, puis clic! sur le menu «Édition» puis le menu «Copie»; Alt + Tab (<i>entraînant le retour au traitement de texte</i>); 7. Clic! sur l'icône «coller» ou Ctrl + V (<i>entraînant l'insertion de l'équivalent dans le traitement de texte, à la place du terme en langue source</i>).

- D'une fiche précédente: clic! de souris sur la flèche vers la gauche de la barre d'outils Multiterm;
- De la fiche suivante: clic! de souris sur la flèche vers la gauche de la barre d'outils Multiterm;
- Du champ précédent de la fiche active: clic! sur la flèche vers le haut de la barre d'outils Multiterm;
- Du champ suivant de la fiche active: clic! sur la flèche vers le bas de la barre d'outils Multiterm.

L'utilisateur peut aussi insérer les vedettes ou une rubrique sélectionnée dans ce document. Multiterm ne cache pas derrière une fenêtre temporaire, même petite, une partie du texte en cours de traduction ou de rédaction. Mais l'utilisateur peut aussi bien consulter le dictionnaire actif en avant-plan, à savoir basculer dans Multiterm lui-même et y consulter le dictionnaire ouvert (dans ce mode, les trois-quarts de son texte resteront visibles). De plus, ce gestionnaire de fiches terminologiques permet d'importer et d'exporter des fichiers entiers ou filtrés selon une combinaison puissante de critères et éventuellement de plusieurs opérateurs booléens. Selon le filtre défini par l'utilisateur, le texte ainsi généré comporte en arrière-plan tout ou partie des fiches du dictionnaire et pour elles, tout ou partie des rubriques. Plusieurs modèles de fiches préprogrammés automatisent la présentation en traitement de texte de ces fichiers importés: grand tableau ou typographie analogue à celle des dictionnaires imprimés sur papier. L'exportation de fiches peut aussi se faire de et vers tout autre logiciel, en passant alors par le format Ascii. L'importation est la lecture par Multiterm d'un document créé en traitement de texte, en Ascii, en format RTF, SGML etc. Elle facilite la conception de dictionnaires assistée par ordinateur (récupération par Multiterm de fiches générées par des concordanciers-analyseurs, etc.).

Eurodicautom et la version de 1994 de Termium n'offrent aucune interface avec le traitement de texte. L'utilisateur doit donc prendre note des termes trouvés ou demander une impression des fiches extraites de ces banques. Cependant, Termium permet d'imprimer ou d'enregistrer sur le disque rigide une ou des fiches sélectionnées sur critère. Le document Ascii ainsi créé peut être lu et utilisé dans le traitement de texte, par exemple pour être diffusé dans un service de traduction. Les interrogations en arrière-plan, en suivant l'ordre d'occurrence des termes problématiques dans un texte à traduire, sont prévues sur Eurodicautom. Notons que la Commission de l'Union européenne donne désormais accès à Eurodicautom par Internet.

5 Ergonomie des consultations directes

La consultation d'une banque ou d'un dictionnaire se fera bien sûr souvent depuis cette source elle-même, non à partir d'un terme sélectionné dans un document écrit en traitement de texte.

5.1 Durée de consultation

Le tableau ci-dessous montre des écarts de vitesse importants dans la consultation selon le support employé.

	Nombre de minutes
Dicosoft	8
Eurodicautom	25
Termium	12
Multiterm	-
Wordswitch	8
Larousse	8

Durée moyenne de consultation pour 20 termes

La durée de consultation pour le même échantillon d'une vingtaine de termes passe du simple (8 minutes pour l'ensemble des dictionnaires électroniques) au triple (25 minutes pour Eurodicautom). N'ayant pas accès à une liste alphabétique des vedettes qui font l'objet de fiches dans la banque, l'utilisateur est obligé de lire toutes les fiches retenues par Eurodicautom en réponse au terme demandé, sans même en connaître le nombre. Il en fera parfois défiler une cinquantaine sans trouver celle qui l'intéressait. Une deuxième raison est l'amplitude du domaine. Le domaine est le seul filtre de recherche disponible pour la banque, mais le nôtre est si vaste que beaucoup de fiches sans lien avec l'informatique sont proposées. La lenteur de consultation s'explique également par les fiches en double, en triple... Les autres chiffres ne sont pas pertinents. Les chiffres moyens de Termium s'expliquent en effet par la lenteur du lecteur de DON utilisé pour notre test (un des plus anciens du marché, à vitesse simple, et qui tournait sur un 386 SX à 25 MHz). Les dictionnaires sur disque rigide étaient sur deux 486 SX à 25 MHz. La différence par rapport aux dictionnaires sur disque rigide est cependant bien moins grande qu'avec d'Eurodicautom.

L'intitulé des boutons de Dicosoft est significatif. Avant d'utiliser le dictionnaire pour la première fois, l'utilisateur se passe

donc aisément du mode d'emploi en ligne et de l'unique feuille d'explications fournie avec le logiciel.

Avec Termium, l'utilisateur n'est jamais bloqué par l'oubli ou l'ignorance d'une commande. Toutes les fonctions sont en effet accessibles soit par le biais d'une touche-fonction (F1, F2... Fn), dont les plus courantes sont affichées en permanence au bas de l'écran, soit par le biais des menus, l'intitulé de ces menus étant parfaitement intelligible. L'utilisateur choisit d'afficher les fiches dans leur totalité ou uniquement les rubriques qui

l'intéressent. Dans Wordswitch, l'ajout et le retrait de termes dans le dictionnaire sont un peu plus difficiles, afin d'éviter les suppressions intempestives de termes. Une petite ombre au tableau : l'absence de traduction de certains menus (l'utilisateur peut choisir la langue d'utilisation du logiciel). C'est ainsi que le menu «Aide» est rédigé uniquement en néerlandais. La confection et la gestion de fiches terminologiques étant plus complexes que la consultation d'un dictionnaire (même assortie de l'ajout de vedettes terme à terme comme le permet Wordswitch), les menus de

Multiterm semblent plus opaques. C'est la (légère) rançon à payer pour la puissance de ce logiciel, notamment la paramétrabilité de l'affichage et de l'exportation des fiches qu'il permet de créer. À cela s'ajoute le fait que les menus contextuels ne sont pas visibles en permanence (l'utilisateur doit cliquer avec le bouton droit de la souris pour les faire apparaître). Cet usage, récent et pratique, du bouton droit de la souris n'est pas encore entré dans «l'inconscient collectif» des utilisateurs des logiciels pour Windows.

5.2 Consultation du dictionnaire depuis son logiciel même

Eurodicautom	Termium	Dicosoft	Multiterm	Wordswitch	Larousse
<ol style="list-style-type: none"> 1. Pression de la touche Q puis de la touche retour-chariot; 2. Frappe du mot à faire chercher par le logiciel; 3. Pression sur la touche retour-chariot (entraînant l'affichage de la première fiche retenue pour l'entrée demandée); 4. Pression sur la touche C puis de la touche retour-chariot (entraînant l'affichage de la fiche suivante, lorsque la banque en comporte au moins deux); 	<ol style="list-style-type: none"> 1. Choix de la langue de départ des termes à chercher (à savoir des vedettes de la banque ou n'importe quelle rubrique de leur fiche), par touche-flèche de déplacement puis retour-chariot; 2. Dactylographie du terme à chercher et/ou sélection d'un terme dans la liste alphabétique des vedettes, au moyen des touches-flèches puis du retour-chariot; 3. Pression de la touche F9 pour faire afficher la ou les fiches. 	<ol style="list-style-type: none"> 1. Frappe de la vedette à chercher; clic! de souris sur bouton OK (entraînant l'affichage de l'équivalent du terme sélectionné); 2. Éventuellement, sélection d'un autre terme dans la liste affichée. 	<ol style="list-style-type: none"> 1. Dactylographie du terme à chercher; 2. Pression de la touche retour-chariot. 	<ol style="list-style-type: none"> 1. Frappe du mot à chercher par le logiciel; 2. Clic! sur le bouton «Traduire»; 3. Éventuellement, sélection d'un autre mot dans la liste affichée par le logiciel. 	<ol style="list-style-type: none"> 1. Dactylographie du mot à faire chercher par le logiciel; 2. Pression sur la touche retour-chariot; 3. - Soit: deuxième retour-chariot si le mot sélectionné automatiquement dans la liste par le logiciel convient (entraînant l'affichage de l'équivalent du mot demandé); - Soit: sélection d'un autre mot dans la liste alphabétique des entrées du dictionnaire et retour-chariot (entraînant l'affichage de l'équivalent du mot demandé); 4. Éventuellement, sélection d'une autre acception du mot demandé, dans le cas de termes polysémiques (entraînant l'affichage de l'équivalent du mot demandé).

5.3 Feuilletage du dictionnaire depuis le logiciel même

Eurodicautom	Termium	Dicosoft	Multiterm	Wordswitch	Larousse
<ul style="list-style-type: none"> - Pour accéder à la fiche suivante: pression sur la touche C puis sur retour-chariot; - Pour avancer ou reculer de 1, 2... ou n fiches par rapport à la fiche affichée à un moment donné: dactylographier + 1, + 2... + n ou -1, -2... - n, puis pression du retour-chariot; - Pour accéder à la fiche comportant un numéro précis: dactylographier = 1, = 2... = n, puis retour-chariot. 	<p>Dans la série de fiches sélectionnées sur base du terme demandé,</p> <ul style="list-style-type: none"> - Pour passer à la fiche suivante: touche de fonction F7; - Pour revenir à la fiche précédente: Shift + F7; - Pour revenir à la première fiche: Alt + 1; - Pour passer à la dernière fiche: Alt + L; - Pour accéder à une fiche précise: Alt + G, puis le numéro de cette fiche; - Dans la liste alphabétique des entrées: touches-flèches et/ou touches Page Up et Page Down. 	<p>Dicosoft permet d'accéder à tous les équivalents des termes de la liste compilée sur base de l'entrée demandée, soit par:</p> <ul style="list-style-type: none"> - Un clic! sur le mot dont on veut connaître la traduction (<i>en utilisant éventuellement l'ascenseur situé à droite de cette liste</i>); - Les touches-flèches; - Un ou plusieurs clics! sur les boutons; «Préc» (terme précédant le terme mis en surbrillance dans la liste) et «Suiv» (terme suivant le terme mis en surbrillance dans la liste). 	<ul style="list-style-type: none"> - Fiche précédente: clic! de souris sur le bouton situé à gauche de la vedette de la fiche active ou F4 ou bien encore CTRL + F4 (si la recherche doit passer par un filtre préalablement défini); - Fiche suivante: clic! de souris sur le bouton situé à droite de la vedette de la fiche active ou F5 ou encore CTRL + F5 ou F4 (si la recherche doit passer par un filtre préalablement défini); - Fiche n° 1, 2... ou n: Ctrl + N puis 1, 2... ou n; - Renvois hypertextuels: l'utilisateur peut donner le statut de bouton hypertextuel à certains termes, dans les définitions ou d'autres rubriques d'une fiche. Multiterm permet donc de créer aisément de tels renvois hypertextuels. Ces termes renvoient directement à la fiche comportant en vedette ces mots-renvois. Pour accéder à cette fiche par saut hypertextuel: cliquer sur le terme (affiché dans une couleur distincte) sous la rubrique qui le comporte. 	<p>Pour obtenir un aperçu du dictionnaire, qui est terme à terme, l'utilisateur ne peut que voyager dans son index:</p> <ol style="list-style-type: none"> 1. Dactylographie de la racine d'un mot, ou de son initiale; 2. Clic! de souris sur le bouton «Recherche» (il lance l'affichage de la liste des mots qui, dans le dictionnaire, commencent par la ou les lettres demandées); 3. Sélection d'un terme dans la liste affichée; 4. Clic! de souris sur le bouton «Traduire» pour faire disparaître cette liste. 	<ul style="list-style-type: none"> - À tout moment, l'utilisateur peut sélectionner un mot dans l'index alphabétique, toujours affiché; - Renvois: le logiciel permet de renvoyer à un autre terme du dictionnaire ouvert ou même d'un autre dictionnaire Larousse disponible sur le même ordinateur. On sélectionne un mot apparaissant dans la liste des entrées, on clique sur les menus «Édition» puis «Renvoi» (pour sauter à un autre mot dans le même dictionnaire). Ou bien, pour sauter à un mot d'un autre dictionnaire, on sélectionne un mot apparaissant dans la liste des vedettes du dictionnaire actif, on clique sur le menu «Édition» puis sur le menu «Renvoi dans un autre dictionnaire» et on sélectionne le dictionnaire dans lequel on veut effectuer la recherche. Le logiciel affiche directement la liste des mots les plus proches de celui qui a fait l'objet du renvoi.

5.4 Inversion des langues source et cible

Eurodicautom	Termium	Dicosoft	Multiterm	Wordswitch	Larousse
<ol style="list-style-type: none"> 1. Pression de la touche L, puis du retour-chariot; 2. Frappe des initiales de la langue source (NL pour néerlandais, par exemple), puis retour-chariot; 3. Frappe des initiales de la ou des langues-cibles ou de A pour faire afficher les équivalents dans toutes les langues d'Eurodicautom, puis retour-chariot 	<p>Positionnement sur la clé de recherche symétrique (par exemple TEN quand on avait choisi TFR), puis retour-chariot.</p>	<p>Clic! sur le bouton «Langue».</p>	<p>Les langues source et cible peuvent être changées indépendamment l'une de l'autre (Multiterm est multilingue), et ce, en fonction des langues présentes dans le dictionnaire actif de l'utilisateur. Pour changer de langue:</p> <ol style="list-style-type: none"> 1. Clic! de souris sur la flèche vers le bas, à côté de l'intitulé de la langue active, à droite de la rubrique «Langue source» ou «Langue d'arrivée»; 2. Sélection d'une autre langue dans le menu déroulant affiché. 	<p>Wordswitch est unidirectionnel. Pour inverser le sens de la consultation, l'utilisateur doit changer de dictionnaire, donc acheter séparément les dictionnaires français-néerlandais (ou anglais, allemand, espagnol) et néerlandais (anglais, allemand, espagnol) - français.</p>	<ol style="list-style-type: none"> 1. Clic! de souris sur la flèche vers le bas située à côté de l'intitulé du dictionnaire en cours d'utilisation (menu «Bibliothèque»); 2. Sélection dans le menu déroulant d'un autre dictionnaire disponible; 3. Clic! sur le bouton OK.

6 Tendances générales

Les conclusions à tirer de cette étude seraient nombreuses mais nous n'en épingleons que quelques-unes.

6.1 Présentation

Certaines sources comportent des fautes de frappe. Elles sont assez nombreuses dans Eurodicautom, beaucoup moins dans Termium (*allowy* au lieu de *alloy*, par exemple). Normales dans les banques, ces erreurs le sont-elles dans les dictionnaires proprement dits, perçus par l'utilisateur comme des outils de référence, à l'instar de leurs homologues imprimés? Quelques exemples de fautes de frappe mais aussi d'autres erreurs commises par Dicosoft: *police*, traduit en anglais par *fonte*; *redundant array of inexpensive disk* (au lieu de *disks*); *facility*, traduit uniquement par *possibilité*; *power*, traduit uniquement par *puissance*, mais tous deux traduits correctement dans des syntagmes nominaux dont ils sont un constituant.

Les fiches en double, en triple, en quadruple... restent un peu trop nombreuses dans Eurodicautom. L'exemple ci-après justifie cependant la répugnance que pourraient avoir des terminographes de la Commission à fusionner des fiches hypothétiquement identiques. Comment traduire *interface enfichable*, sachant que: *organe enfichable = plug-in unit*, *unité enfichable = plugboard unit*, *carte enfichable = circuit card*, *gabarit enfichable = insert pattern*?

Des défauts subsistent dans certains programmes. Dans Dicosoft, l'inversion du sens de la recherche efface d'office le terme d'entrée, un effet gênant dans un domaine truffé d'anglicismes; s'il tente en effet la recherche d'un même terme dans les deux langues, l'utilisateur est obligé de redactylographier ce terme après

avoir changé de langue source. Une faiblesse de Termium: si l'utilisateur est bien guidé dans ses consultations, la sélection de la langue de recherche avant chaque demande de terme ralentit un peu les recherches consécutives de termes.

6.2 Contenu

La principale qualité d'un dictionnaire est son exhaustivité. Dicosoft et les banques la possèdent: seuls 5,75% de termes n'ont pas été trouvés dans Eurodicautom. La formule de la banque, avec ses mises à jour, est préférable aux dictionnaires spécialisés fermés (dont Internet interrompra peut-être la présente diffusion sur DON). Surtout dans un domaine en évolution rapide comme l'informatique, l'analyse comparée du contenu des banques et de Dicosoft le montre. Il faut par contre consulter les trois sources pour avoir toutes les variantes, notamment de syntagmes.

L'équivalence d'une langue à l'autre dépend souvent de l'environnement phrastique et textuel du terme. Grâce à la mention de leurs sources, les banques donnent le contexte ou le moyen d'y accéder.

Les terminographes de Termium inscrivent tous les termes dans un domaine, délimité de manière claire et scientifique. L'attribution des domaines pourrait être plus systématique et cohérente dans Eurodicautom, pour éviter que l'automatisation ne comporte aussi bien des termes désignant des abrasifs pour le travail à froid du verre que des termes de télécommunication. L'extension des domaines gagnerait en outre à être réduite dans cette banque.

Une série de données utiles manquent dans Dicosoft: ce dictionnaire traductif terme à terme n'indique ni auteur, ni sources, ni date... Soulignons par contre son

effort de traduction des emprunts usuels à l'anglais. Une série de fiches d'Eurodicautom et de Termium comportent les synonymes et termes proches, des notes de traduction, des explications... Notons que quelques fiches d'Eurodicautom comportent, en vedette, beaucoup de termes qui ne sont pas tous synonymes.

Quelques commentaires sur l'usage de ces termes proches, à l'image de ceux de Termium, seraient bienvenus. Des fiches relatives à des homonymes ne comportent ni domaine ni définition dans cette banque. Ne devraient-elles être complétées? Termium inclut de nombreux acronymes et abréviations, des commentaires sur l'usage des termes, des fiches réservées aux difficultés de traduction et d'autres données administratives telles que les sources, le nom du terminographe auteur de la fiche, la date de création et éventuellement de mise à jour... Les fiches d'Eurodicautom comportent toutes les variantes orthographiques d'un terme (notamment les variantes nationales comme l'anglais américain et britannique).

Rappelons également des rubriques souhaitées des traducteurs pour les sources terminographiques, signalées dans un rapport Eurotra, un consortium Esprit, deux enquêtes auprès des traducteurs belges:

- Syntaxe: mention de la transitivité ou de l'intransitivité des verbes, des prépositions accompagnant ces verbes (données présentes dans le Larousse), des collocats du terme;
- Données encyclopédiques: mention de la définition, mention du contexte uniquement s'il éclaire le sens ou les cooccurrents du terme, faux amis, distinction entre termes proches, quasi-synonymes et synonymes, accompagnée de notes explicatives;
- Emploi du terme: domaine, sous-domaine, type de texte, statut de la dénomination (usage recommandé, toléré, à éviter, obsolète), avertissement du statut éventuel de

néologisme et dans ce cas, source de ce néologisme, variantes orthographiques;

- Acronymes et abréviations, sous la forme de vedettes à part entière, distinctes de l'expression qu'ils remplacent et désignent;

- Fiabilité des différentes données de la fiche, non de la seule vedette;

- Renvois à l'hyperonyme, méronyme, aux hyponymes, antécédents et conséquents, collatéraux directs limités aux termes très techniques ou aux équivalences partielles;

- Renseignements sur le type d'équivalence d'une langue à l'autre: partielle, totale, nulle, réversible ou asymétrique;

- Références bibliographiques de graphiques, dessins et corpus de textes utiles (plutôt que des extraits textuels encombrants et jamais substitutifs);
- Motivation du terme, c'est-à-dire lien logique entre la notion et sa dénomination (origine de telle métonymie, etc.).

Il est un danger auquel les alimentations décentralisées et les sous-traitants des terminographes doivent rester attentifs. Il est dangereux de compléter des définitions et d'ajouter des néologismes dans un dictionnaire, lors de mises à jour, sans vérifier l'obsolescence d'anciennes vedettes ou définitions. Les trois sources terminologiques étudiées mettent par exemple sur la même fiche *diskette* et *floppy disk*, et parallèlement *disquette*, *disque flexible* et *minidisque*, etc., sans mentionner (sauf Termium) les différences entre les notions désignées. Or *disque flexible* ne dénomme pas une disquette de 3,5 pouces protégée par un petit boîtier plat, mais uniquement une disquette de 5,25 pouces glissée dans une enveloppe. Ces deux supports appartiennent cependant bien à la famille des disquettes. Quant au *minidisque*, c'est une disquette,

également «non flexible» mais de dimensions inférieures à 3,5 pouces.

6.3 Accès au contenu

Sauf Larousse, aucun dictionnaire ne dispose d'un lemmatiseur. Or, ce dernier facilite la recherche, depuis un texte chargé dans un traitement de texte, de l'équivalent des mots situés à gauche du curseur, évitant la reffrappe du mot ou de l'expression canonique.

Les opérateurs booléens de Termium optimisent les résultats des requêtes par les filtres qu'ils permettent. Dans notre version, il aurait fallu pouvoir les utiliser pour des conditions portant à la fois sur la chaîne à trouver et sur le domaine de la vedette. Un index de termes-candidats est également précieux pour les banques, comme celui de Termium. L'indication du nombre de fiches retenues par Termium sur base de l'équation (booléenne ou non) de recherche, permet d'affiner une requête avant de lire les fiches elles-mêmes, si l'utilisateur a «ratissé» trop large.

Dans une banque sans index comme Eurodicautom, un second frein aux recherches est l'absence de différenciation entre substantifs et verbes. L'utilisateur est obligé de passer en revue un nombre imposant de fiches comportant en vedette le substantif *chain* avant de pouvoir lire les quelques fiches comportant en vedette le verbe *to chain*.

Prévu dans Eurodicautom et surtout Dicosoft, l'affichage des syntagmes comportant un mot, indépendamment de l'ordre de leurs constituants, optimise également l'accès au contenu.

Même chose pour les renvois hypertextuels de Larousse et Multiterm et les menus contextuels de Multiterm.

6.4 Interfaces

Le Système 7 étant lentement mais sûrement balayé, malgré l'universalité du PC Power Mac, Windows est devenu l'interface-utilisateur incontournable. Les grandes banques s'y sont donc tout naturellement branchées.

L'interface d'une terminologie avec le traitement de texte, comme on la trouve sur les dictionnaires électroniques proprement dits, fait gagner beaucoup de temps aux traducteurs et rédacteurs. C'est elle qui est à l'origine du succès des gestionnaires de vocabulaires, des dictionnaires sur DON, du commerce. C'est que le traitement de texte est le pivot de l'informatique de la traduction comme de toute la bureautique, d'ailleurs.

Progressivement, les autres logiciels de traduction assistée se greffent sur lui, malgré les reprogrammations en chaîne requises par la sortie annuelle d'une mise à jour de ces traitements de texte. Simplifié par la disparition progressive de tous les concurrents de Word, ce lien étroit avec le traitement de texte accroîtrait l'ergonomie des banques. Cette interface se doit cependant d'être plus performante que la simple utilisation du presse-papiers de Windows adoptée dans la dernière version de Termium.

Sauf celle de Dicosoft et de Multiterm, les interfaces des dictionnaires électroniques avec le traitement de texte ne sont pas parfaites. Comme d'autres utilisateurs sans doute, nous avons donc ajouté des macrocommandes à Word pour améliorer les consultations du dictionnaire depuis Word 6 et 7.

6.5 Diffusion des banques de terminologie

Les dictionnaires vendus en même temps que les gestionnaires de vocabulaires Multiterm, MTX21...,

donc consultables par eux, sont encore peu nombreux mais leur nombre croît. MTX21 se vend en même temps que de nombreux dictionnaires spécialisés et généraux, mises à jour comprises. Il est sous le Larousse traductif général analysé plus haut et sous le Larousse des affaires. Les gestionnaires de glossaires devraient-ils devenir l'outil de consultation des dictionnaires spécialisés et des banques de terminologie, comme Multiterm le sera sous peu pour Eurodicautom? Banques de termes, glossaires «boîte à chaussures» des traducteurs, traduction assistée et automatique s'intègrent. Ils ont fait du gestionnaire de vocabulaires un passage obligé, après celui par le traitement de texte. En effet, les logiciels appelés (par un calque malheureux!) «mémoires de traduction» sont des exploiters d'anciennes traductions ressemblant par leur style et leurs vocables aux textes à traduire. Un module de gestion de vocabulaires est inclus à ces exploiters, qui consultent eux-mêmes le contenu de ces vocabulaires. À la différence du gestionnaire de vocabulaires de l'exploiteur d'anciennes traductions Translation Manager d'IBM, l'exploiteur de traductions Translator's Workbench de Trados est dissociable de son module de gestion de vocabulaires. Celui-ci est Multiterm. Translator's Workbench peut consulter lui-même et insérer, à la demande ou en arrière-plan, les termes trouvés dans les fichiers de l'utilisateur, de dictionnaires du commerce et, dans un avenir proche, d'Eurodicautom.

Toutes les banques envisagent peut-être de s'installer sur Internet. Leur diffusion sur disque optique par les grandes librairies comme la FNAC (en Belgique, en France), mais aussi les magasins de jeux et de DON de musique, ne serait-elle pas concevable? Les banques prêtes à une diffusion par domaines ou macro-domaines toucheraient ainsi les

traducteurs spécialisés dans ces secteurs.

D'autres catégories d'utilisateurs que les diffuseurs multilingues comme les traducteurs et les journalistes sont accessibles aux grandes banques. Leurs vedettes pourraient servir d'index hypertextuel aux tutoriels de spécialités ou de langues de spécialité. Des dictionnaires définitoires pourraient, en outre, être inclus en arrière-plan à ces tutoriels. Ils seraient consultables par simple clic! de leur vedette, dans le texte du cours lui-même. Au sein de ce texte didactique, cette vedette aurait en effet le statut de bouton hypertextuel. Enfin, pourquoi des manuels destinés aux professeurs de langues spécialisées ne proposeraient-ils pas une série d'exercices de vocabulaire exploitant le contenu des banques de terminologie? Ils pourraient s'appuyer sur la puissance des opérateurs booléens et des renvois hypertextuels de leur logiciel de consultation. Si ces livres étaient publiés dans les collections des manuels de langue, comme celles de l'université de Cambridge, ils complèteraient ceux qui traitent d'autres modes d'enseignement, comme la compréhension à l'audition, la contraction de texte...

Des juristes, dit Snow, citent des définitions de dictionnaires et de banques de termes à des fins juridiques. D'autres usages des banques sont imaginables. Dans les firmes de conseil ou les grandes entreprises, les spécialistes des réunions pourraient systématiquement faire débiter par une série de définitions certains colloques, certains types de réunions. Histoire de faciliter les échanges entre spécialistes en évitant les malentendus, par exemple dans les groupes de travail multidisciplinaires... Déjà, des professeurs de langues donnant cours à des cadres et des commerciaux couplent leur cours (d'anglais) à

l'enseignement des techniques de communication intra-entreprise. Ils pourraient y intégrer des savoir-faire terminologiques.

*Élisabeth Hombrouck,
Centre de terminologie de Bruxelles,
Institut Marie-Haps,
Belgique.*

ANNEXE I

Liste alphabétique des termes français demandés

accéder à	cache burst	développeur
accès à	cadrer	diagonale d'affichage
accessible (sous Windows)	canal	dictionnaire
acheminer	canal DMA (Direct Memory Access)	diffusion de trame
administrateur	caoutchouc mousse	diode laser
adresse	capot	disque
adresse d'entrée/sortie	carte	disque DON-ROM
adresse d'un site	carte d'extension	disque dur
affichage	carte de type II	disque hotplug
affichage Super VGA	carte mère	disque magnétique
aiguille (d'imprimante)	carte multivoie	disque SCSI fast/wide
alimentation	carte VGA	disquette (3,5 pouces)
alimentation redondante	DON-ROM	distant
alliage étain, nickel et cuivre	châssis	DLL
amovible	chiffrage	DMA
annuaire ouvert	chipset	double balayage
ANSI	circuit imprimé (du clavier)	driver
application	circuit vidéo	dual SCSI path
application MFC	classe «collection»	dynamic sparing
ASCII	classe MFC 3.0	écran couleur
asynchrone	clavier	éjection d'une carte
axe	clavier à micro-contacts	électro-aimant
bac (à feuilles)	clavier capacitif	enduit magnétique
balayage	clé de chiffrage	enregistrer
barre d'outils	clé de session	entrant (fax - ou donnée -)
barre d'outils détachable	clic	environnement
barrette	clic sonore	Ethernet
base de données	client-serveur	extension Plug and Play
base de données client-serveur	CMS (Centralized Management System)	faisceau de lumière
bibliothèque (Win32s)	code	fenêtre
bibliothèque statique	code ASCII	fenêtre parente
BIOS	code XOFF	fibre optique
bit	compatibilité	firmware
bobine de focalisation	compatible	fonctionnalité
boîte aux lettres	compilateur	format
boîte de dialogue	compression	format A4
boîte de dialogue à onglets	condensateur	format de fichier
boîte de dialogue multipages	configuration	format légal
bouton	configuration en concurrence	format lettre
bps	connecter (se -)	frappe
bras	connecteur	fréquence de rafraîchissement
bulle	consulter	ftp
bulle d'aide	couplé à	Gb
bus	courrier vocal	générer
bus local	coussinet	gérer
bus local VESA	cristaux liquides	gestion
bus PCI	cryptographique	gestion centralisée
câble	DAC	gestion de données
câble d'alimentation allume-cigares	détecteur	gestion des pages de garde
cache		groupe de travail

groupe RAID distant	mémoire vive	piste
groupe RAID indépendant	messagerie	piste concentrique
groupware	messagerie électronique	piste en spirale
hôte	MFC	pixel
hôte indépendant	mettre à jour	plate-forme
hot spare	MHz	plongeur métalloplastique
http (HyperText Transport Protocol)	micro	Plug and Play
hypertexte	mise à jour	police
hytelnet	Mo	police résidente
Hz	mode asynchrone	port
image numérisée	mode d'émulation	port d'entrée/sortie
implémentation	mode hibernation	port d'extension
imprimante à sublimation	mode synchrone	port parallèle
imprimante à transfert thermique	mode veille	port PCMCIA
imprimante matricielle	modem	port série
imprimante matricielle à transfert thermique	module	portable
impulsion magnétique	module API	portable distant
indicateur de source	mot de passe	posterizing
installation	moteur pas à pas	pouce
instancier	multispin	préinstaller
interface	multithreading	processeur
interface enfichable	multithreading préemptif	processeur parallèle
interface parallèle	mylar	processeur RISC
interface TWAIN	navigation	progiciel
interface utilisateur	netsurfer	programme
interface utilisateur graphique	norme	programme de conversion
interfacer (s-)	norme Plug and Play	programme graphique
IRQ	notebook	protection antivol
ISA	numéro d'identification	protection par mot de passe
Ko	obturateur de disquette	protocole
langage C ++	OCR	rafraîchissement
lecteur	onglet	RAID
lecteur de bandes	on-line	RAM
lecteur de cartes	optimiser	read/write cache
lecteur de DON-ROM	option	réceptionniste électronique
lecteur de disque	optionnel	renvoi
lecteur de disquette	outil de navigation	réplicateur
levier	oxyde de fer	réplicateur de ports
ligne IRQ (Interrupt ReQuest)	page de garde	réseau
liste chaînée	palette	réseau informatique
local	panneau de contrôle	réseau local
logiciel	panneau de contrôle frontal	réseau ouvert
logiciel de fax	paramétrer	ressource
logiciel de fax en réseau	paramètres	ressource système
logiciel intégré	partage (de télécopies)	RISC
loquet	partageable	RNIS
machine de bureau	path	router
macro	PC	routeur
macro try... catch	PCI	SCAM
Mb	PCMCIA	scannage
mémoire	périphérique	scanner
mémoire EDO (Extended Data Output)	périphérique SCSI	score Winstone
mémoire Flash	pi	SCSI
	pilotage direct	SCSI Configuration AutoMagically
	piloter	SCSI path

secteur	synchrone	trame
serveur	syntaxe	transfert
serveur de fichiers	syntaxe try... catch	transfert de données
serveur Web	système ouvert	transport hypertexte
service	système réseau	try... catch
sillon	table	uniformiser
sillon hélicoïdal	tableau	unité centrale
SIMM	tampon	URL (Unified Resource Locator)
site	tampon interne	usenet
site central	téléchargement	utilisateur
site ftp	télécopie	utilisateur local
sous-système	template	utilitaire (Interlink)
spécifications	tension électrique	version
specs	terminal	VESA
stockage	tête de lecture	VGA
stockage distribué	tête de lecture/écriture	vitesse de rotation
stockage distribué ouvert	thread	vitesse linéaire constante
subnotebook	thread d'exécution	Voice Processing
Super VGA	ticket	Web
support	tirage en couleur	web-browser (= browser)
support dual-host	tolérance aux pannes	XOFF
supporter	trackball	

Les banques de terminologie au seuil d'une nouvelle dynamique

Les banques de terminologie se trouvent au cœur d'un virage technologique. Elles doivent adapter leur évolution à l'évolution des contextes actuels et relever de nouveaux défis. Les producteurs et les diffuseurs de ces banques sont tenus de se fixer de nouveaux objectifs et de mettre en place les mécanismes de production et de diffusion pour les atteindre. Les banques de demain pourront alors jouer un tout nouveau rôle dans la société de l'information dont elles font partie.

Termes-clés :
banque de terminologie ;
industrie de la langue ;
autoroute de l'information ;
partenariat.

1 Introduction

La question que je vais aborder avec vous aujourd'hui concerne bien sûr le développement des banques de terminologie. J'aimerais avancer l'hypothèse que les banques de terminologie pourraient bien se trouver aujourd'hui au seuil d'une nouvelle dynamique.

Pour illustrer mon propos, je vous invite à examiner avec moi les nouveaux contextes dans lesquels évoluent les banques de terminologie. Nous verrons comment ces nouveaux contextes feront surgir de nouveaux défis pour les producteurs et les diffuseurs de telles banques. Face à ces nouveaux défis, nous verrons comment les banques de terminologie seront invitées à se fixer de nouveaux objectifs. Enfin, nous tenterons de proposer de nouveaux mécanismes qui viendront précisément soutenir les nouveaux objectifs retenus par celles-ci. J'espère ensuite pouvoir illustrer comment pourrait s'accomplir l'évolution des banques de terminologie d'hier à demain.

2 Les nouveaux contextes

Quels sont ces nouveaux contextes dans lesquels évoluent les banques de terminologie ?

LES NOUVEAUX CONTEXTES

- La révolution des technologies des communications ;
- Le développement des industries de la langue ;
- Le développement des autoroutes de l'information.

Vers une démocratisation de l'information
Vers une économie du savoir

On ne surprendra personne en mentionnant d'abord la révolution des technologies des communications. En effet, à chaque jour, que ce soit dans les journaux, à la télé, au restaurant ou tout simplement dans la rue, on entend parler de quoi ? Eh bien ! on entend parler de réseaux de communication, de câblodistribution, de systèmes de communications par satellite ou encore de transmission sans fil.

Il faut peut-être en conclure que nous vivons désormais dans ce que certains pourraient appeler une incontournable infrastructure technologique. Ce que je vous propose de retenir ici, c'est plutôt le fait que c'est justement cette nouvelle infrastructure qui nous permettra d'envisager la réalisation d'une foule de nouveaux produits et services extraits des banques de terminologie.

Comme deuxième élément de contexte, je mentionnerai le développement des industries de la langue. Parallèlement au développement des technologies des communications, nous assistons à la création de différents outils informatisés dont ceux dits des industries de la langue.

Vous le savez, il s'agit par exemple d'analyseurs de textes, de correcteurs orthographiques ou encore, plus généralement, de différents outils d'aide à la rédaction. On peut penser également à des interfaces en langage naturel, à des didacticiels de langue ou encore à des produits permettant une reconnaissance vocale. Il importe également de souligner que le développement des systèmes de gestion des bases de données terminologiques relève aussi du domaine des industries de la langue.

Bref, ce sont finalement des outils facilitant la recherche de l'information, son analyse, son traitement, voire la production de nouvelles informations et l'édition de celles-ci. Ce sont donc là différents outils auxquels il devient intéressant de recourir lorsqu'il s'agit de produire et de diffuser les banques de terminologie.

Le troisième élément de contexte est lié au développement des autoroutes de l'information. On le sait, le réseau des réseaux connaît une popularité sans doute irréversible. On estime actuellement à 40 millions dans le monde le nombre d'utilisateurs d'Internet. J'espère que les données demeurent valables pendant quelques jours quand on les compte en millions d'utilisateurs...

Aujourd'hui, tout le monde veut savoir ce qu'est Internet, ce qui se passe sur les réseaux et comment on peut s'en servir. Même chez les enfants, ce qui importe maintenant, ce n'est plus tellement de devenir astronautes «quand ils seront grands», il s'agit plutôt de devenir internautes

dès maintenant. Les gouvernements sont conscients des enjeux que représentent les autoroutes de l'information pour la société. Au Québec, comme au Canada, les administrations souhaitent participer activement au développement des autoroutes et les comités spéciaux qui ont été formés pour étudier cette question ont déjà formulé des recommandations concernant l'implantation d'Internet, notamment dans le domaine de la santé et surtout dans celui de l'éducation. Le domaine de l'éducation, on le verra, pourrait bien être celui qui ouvrirait de nouvelles portes aux banques de terminologie.

On peut se demander maintenant quels sont les effets conjugués de cette évolution des contextes?

Par le biais des réseaux d'information et des différents outils qui permettent de les exploiter, il est permis de dire que nous assistons à une véritable démocratisation de l'information. En même temps, nous passons d'une économie basée sur la production à une économie du savoir. Il faut souligner ici, que cette nouvelle économie se définit principalement par les caractéristiques suivantes: d'abord, cette nouvelle économie est fondée sur les idées, ensuite, elle s'appuie sur un investissement dans le capital humain et, enfin, c'est le savoir qui devient déterminant dans ce que l'on appelle l'accroissement de la richesse.

Il faut donc s'attendre à ce que de plus en plus de gens cherchent à acquérir de plus en plus de connaissances et l'on peut s'attendre à ce que, pour ce faire, ils aient recours aux technologies dont il vient d'être question.

Mais, qu'est-ce que cela signifie pour les producteurs des banques de terminologie? Comment sommes-nous concernés par cette quasi-révolution qu'un collègue informaticien et moi avons déjà

appelée «le retour de Gutenberg» dans le dernier numéro de *Terminogramme*, numéro préparé en prévision de la présente table ronde?

Eh bien! pour les producteurs de terminologie comme pour les autres producteurs d'information, il faut comprendre que le développement de cette nouvelle économie est avant tout un appel à l'ouverture, à l'initiative et à la créativité. Retenons-le et voyons ensemble quels sont les nouveaux défis qui se posent maintenant aux banques de terminologie et comment celles-ci peuvent bénéficier du contexte actuel.

3 Les nouveaux défis

Bien sûr, nous pouvons attendre de voir ce qui va se passer et, d'une certaine manière, subir cette évolution. En fait, au rythme où vont les choses... il se pourrait bien que nous n'ayons pas à attendre très longtemps avant d'être bousculés par les événements... On peut aussi décider de devenir des agents de changement en sachant que c'est sans doute la meilleure façon d'en tirer le plus grand profit possible.

Quels sont donc ces nouveaux défis auxquels sont confrontées les banques de terminologie?

LES NOUVEAUX DÉFIS

Accepter le changement ou devenir un agent de changement

- Adapter l'évolution des banques au contexte actuel;
- Revoir la structure des banques;
- Définir de nouvelles stratégies de production;
- Définir de nouvelles stratégies de diffusion.

Eh bien! il faut adapter l'évolution des banques à l'évolution des contextes actuels. Il faut revoir la structure des banques et définir de nouvelles stratégies de production et de diffusion de celles-ci. Pour utiliser des termes de gestion modernes, il faut rien de moins que d'assurer une complète réingénierie des processus de production et de diffusion des banques de terminologie.

La démarche à entreprendre doit s'articuler autour des nouveaux objectifs que chacune des banques de terminologie devra déterminer en fonction de sa mission propre. Pour sa part, la Banque de terminologie du Québec, la BTQ, doit contribuer à la réalisation de la mission de l'Office de la langue française. Elle doit donc être un outil qui favorise l'utilisation des terminologies dans le contexte de la généralisation de l'utilisation du français dans les milieux de travail comme le prévoit la législation linguistique québécoise. Dans les objectifs qui suivent, on retrouvera donc les préoccupations particulières des producteurs de la BTQ.

4 Les nouveaux objectifs

Quels sont donc ces nouveaux objectifs?

LES NOUVEAUX OBJECTIFS

- Optimiser la production et la diffusion des terminologies en utilisant les nouvelles technologies;
- Élargir le marché actuel des banques de terminologie en développant une nouvelle gamme de produits adaptés aux besoins des clientèles;
- Développer de nouveaux rapports avec les clientèles;
- Développer de nouvelles formules de partenariat;
- Contribuer au développement du contenu des autoroutes de l'information.

Dans le contexte actuel de rareté des ressources, on comprendra aisément que l'on souhaite optimiser la production et la diffusion des terminologies en utilisant les nouvelles technologies maintenant disponibles.

Afin de favoriser la diffusion et l'implantation des terminologies pour assurer au français la place qui lui revient dans notre société, on comprendra également que l'on souhaite élargir le marché actuel des banques de terminologie en développant une nouvelle gamme de produits adaptés aux besoins des clientèles.

Le développement des marchés ne saurait avoir le succès espéré sans qu'il soit nécessaire par ailleurs de développer de nouveaux rapports avec les clientèles. Il faut donc également s'intéresser à cette question.

Enfin, il faudra développer de nouvelles formules de partenariat afin de mettre en commun les ressources existantes. Ceci assurera une meilleure coordination des travaux et évitera d'avoir à refaire des travaux qui ont déjà été faits. Il faut également s'assurer de la participation de nouveaux partenaires dont des fournisseurs d'informations diverses pour les associer à la diffusion des terminologies. C'est ainsi que les banques de terminologies pourront contribuer au développement du

contenu des autoroutes de l'information et dans le contexte particulier qui nous intéresse, nous de l'Office de la langue française, au développement d'un contenu d'expression française.

Pour atteindre ces objectifs, je l'ai mentionné plus tôt, il faudra sans doute inventer et mettre en place de nouveaux mécanismes de production et de diffusion. Nous en proposons ici quelques-uns.

5 Les nouveaux mécanismes

5.1 Pour l'optimisation de la production et de la diffusion des terminologies

Pour la réalisation du premier objectif mentionné, soit le fait d'optimiser la production et la diffusion des terminologies, on doit envisager de nouveaux mécanismes qui sont de deux ordres différents.

LES NOUVEAUX MÉCANISMES

- Utiliser les ressources offertes par les réseaux dans une perspective de production terminologique;
- Poursuivre le développement du poste de production terminologique.

Il faudra notamment utiliser les ressources offertes par les réseaux dans une perspective de production terminologique et, dans une démarche conjointe, poursuivre le développement du poste de production terminologique.

On sait que l'on trouve différentes sources d'information dans les réseaux (documentation spécialisée autant que commerciale, normes techniques, périodiques spécialisés, rapports de recherche,

dictionnaires, etc.). De telles sources présentent déjà beaucoup d'intérêt pour les terminologues. Cet intérêt s'accroît encore quand on sait que, dans les autoroutes, on trouve aussi un grand nombre d'informations sur différents projets de recherche en cours, quand on pense que le réseau nous permet aussi de consulter divers spécialistes ou encore de soumettre des propositions terminologiques aux gens du milieu œuvrant dans le domaine de recherche concerné. Il y a donc tout un monde à découvrir mais ce n'est là mon propos et je laisse le soin aux terminologues d'explorer ces différentes possibilités offertes par les réseaux pour en faire le meilleur usage possible. Des travaux dans ce sens ont d'ailleurs déjà été entrepris à l'Office de la langue française.

En ce qui concerne le développement d'un poste de production terminologique, nous savons que plusieurs travaux sont déjà en cours dans ce domaine, notamment à l'Office de la langue française. Le poste de travail que les terminologues de l'Office ainsi que quelques partenaires utilisent, poste appelé SAMI, vous a d'ailleurs été présenté hier par ma collègue, Sylvie Pelletier.

Je me contenterai ici de mentionner simplement certains secteurs dans lesquels les développements doivent se poursuivre. Je pense notamment :

- Aux outils d'assistance à la recherche documentaire dont l'évolution en fonction des besoins spécifiques des terminologues pourraient être grandement appréciée et qui pourraient contribuer à assurer une certaine veille documentaire ;
- Aux outils d'assistance au dépouillement qui viennent soutenir l'établissement des nomenclatures et peuvent contribuer à une certaine veille terminologique comme cela vous a également été présenté hier avec Nomino ;

- Aux différents outils de traitement des données (saisie/mise à jour, correcteurs orthographiques, systèmes de constitution de dossiers terminologiques, systèmes de comparaison de données, assistance à l'établissement de réseaux notionnels, etc.) ;

- Aux systèmes de conversion de données qui viendront permettre autant l'importation que l'exportation de données en vue de l'alimentation des banques, de l'échange de données et de l'intégration de données terminologiques dans des produits conçus à d'autres fins ;
- Aux systèmes d'édition qui viendront soutenir la production de publications imprimées et électroniques favorisant ainsi la diffusion des résultats des travaux.

J'aimerais souligner ici que c'est dans cette même perspective du développement du poste de production terminologique que le Réseau international de néologie et de terminologie (Rint) a créé un groupe de travail qui doit assurer une veille technologique, veille qui permettra d'inventorier et d'examiner des outils terminologiques. Le Groupe de terminologie, puisque c'est comme cela qu'on l'a appelé, doit d'ailleurs faire des démarches auprès des observatoires des industries de la langue afin de coordonner les efforts entrepris dans ce sens.

Voilà donc ce qui est envisagé pour l'atteinte de ce premier objectif énoncé, soit l'optimisation de la production et de la diffusion des terminologies par l'utilisation des nouvelles technologies.

5.2 Pour l'élargissement du marché actuel des banques de terminologie

Quant aux deuxième et troisième objectifs, soit l'élargissement du marché actuel des banques de terminologie en développant une

nouvelle gamme de produits et l'établissement de nouveaux rapports avec les clientèles, j'aimerais d'abord évoquer certains des principes sur lesquels j'estime que devrait s'appuyer toute initiative dans ce sens.

LES NOUVEAUX MÉCANISMES

Quelques principes :

- L'étude des besoins ;
- Le respect de l'environnement culturel ;
- La réalisation d'outils plus... pédagogiques.

Il sera donc question de l'étude des besoins, du respect de l'environnement culturel des utilisateurs et du développement d'outils plus... pédagogiques.

5.2.1 L'étude des besoins

On a parlé tout à l'heure du développement des autoroutes de l'information. On sait maintenant que les gouvernements reconnaissent le caractère stratégique des données concernant l'utilisation d'Internet. D'ailleurs, dans son rapport final, le Comité consultatif sur l'autoroute de l'information recommandait la création d'un «Centre national d'échange d'information» afin d'avoir accès à une information exacte et à jour concernant le développement de l'autoroute de l'information, de son infrastructure ainsi que des applications, des programmes et des initiatives qui y sont liées pour être en mesure de s'ajuster rapidement à l'évolution observée.

Il sera aussi important pour les banques de terminologie d'être au fait de l'évolution des réseaux, autant pour ce qui est du profil des utilisateurs qu'en ce qui concerne la nature des produits et services qu'on y

trouve que pour établir des liens avec d'autres fournisseurs de produits et services. Les systèmes de messagerie font partie de la culture d'Internet et ils nous permettront de connaître également la nature des produits et services que les utilisateurs aimeraient y trouver. Ce sera un grand avantage pour les banques que d'être en relation directe avec les utilisateurs pour faire évoluer leur contenu et leurs systèmes en fonction des besoins exprimés.

5.2.2 Le respect de l'environnement culturel

J'aimerais mentionner ici que le virage qui pourrait être amorcé par les banques de terminologie ne concerne pas exclusivement le développement des autoroutes de l'information même si celle-ci peut présenter des ouvertures particulièrement intéressantes. Plusieurs autres véhicules d'information pourront être utilisés pour diffuser de la terminologie.

J'aimerais soumettre ici l'hypothèse que ce sont les réseaux de communication des milieux visés ainsi que les réseaux d'information utilisés par ces mêmes milieux qu'il nous faudrait mieux connaître. Il ne s'agit pas nécessairement ici de réseaux informatisés mais bien tout simplement des sources d'information utilisées par les différents milieux visés. C'est ainsi et seulement ainsi qu'on assurera le respect de l'environnement culturel des futurs utilisateurs par une offre de produits qui correspondront à leurs habitudes culturelles de consommation de produits d'information et non pas aux formats habituels de diffusion des données terminologiques.

Si l'on offre des banques de terminologie, des dictionnaires ou des lexiques à des traducteurs ou à des terminologues, on comprend facilement qu'ils soient tout à fait à

l'aise avec de tels produits. Si l'on souhaite rejoindre d'autres types de clientèles, il faudra s'assurer de les rejoindre dans le type de documentation qu'elles ont l'habitude de consommer et ce ne sera pas forcément des vocabulaires ou des lexiques. C'est là précisément ce qui nous incite à proposer la réalisation d'outils qui pourraient être qualifiés de plus pédagogiques, surtout si, comme on l'a vu plus tôt, les domaines de l'éducation et de la formation professionnelle occupent de plus en plus de place dans le réseau des réseaux.

5.2.3 La réalisation d'outils plus... pédagogiques

Une fois pris en compte les besoins observés, il se pourrait bien donc que nous nous dirigeons vers la réalisation d'outils plus... pédagogiques, lesquels qui viendraient soutenir nos objectifs de diffusion et d'implantation des terminologies.

Il est bien entendu que les banques de terminologie continueront de trouver leur place sous les formes qu'on leur connaît déjà, soit l'accès direct et la version disque optique compact. On pourra également retrouver les banques dans les autoroutes, et ce, dans leur structure actuelle. Il y aura toujours des utilisateurs pour un tel produit, notamment les langagiers qui sont des partenaires dans la réalisation de notre objectif de diffusion et d'implantation des terminologies, il importe donc de continuer à leur offrir nos produits traditionnels.

On peut toutefois déjà avancer quelques hypothèses quant à une nouvelle façon de présenter des données pour rejoindre d'autres types de clientèles.

LES NOUVEAUX MÉCANISMES

Quelques hypothèses :

- Les vues partielles ;
- La présentation des réseaux notionnels ;
- Les hyperliens ;
- Les banques virtuelles ;
- La terminologie : une valeur ajoutée.

Il s'agit par exemple des suivantes :

En langage informatique, les vues partielles d'une banque constituent en fait une sélection d'une portion de banque. À partir de la structure actuelle d'une banque, il est donc possible de définir de nouveaux découpages d'information et d'offrir ainsi des mini-banques qui peuvent être structurées selon différents critères. De nombreuses combinaisons sont possibles et elles doivent être envisagées en tenant compte des besoins spécifiques d'une clientèle cible donnée.

Il est possible, par exemple, de rassembler la terminologie d'un seul domaine ou encore d'un certain nombre de domaines connexes. Le corpus peut comprendre seulement la terminologie de base et être présenté dans une seule langue ou en version bilingue ou encore multilingue. Il serait aussi possible de joindre des références bibliographiques commentées, et ce, directement à partir des banques documentaires. La constitution d'index par domaine de spécialité, la présentation d'unités phraséologiques ou de contextes d'utilisation ou encore l'inclusion d'illustrations sont d'autres exemples. On peut facilement imaginer que de tels regroupements d'information pourraient venir soutenir un

programme de formation professionnelle.

L'intérêt de la constitution de telles vues partielles devient particulièrement évident dans le contexte des autoroutes de l'information puisque celles-ci pourraient permettre de rejoindre des clientèles qui ne consulteraient pas forcément une banque de terminologie comme celles que l'on connaît mais qui apprécieraient voir regrouper les notions de base d'un domaine. Elles y verraient vraisemblablement un moyen pratique de se familiariser avec un domaine donné.

La présentation des réseaux notionnels peut également constituer une nouvelle façon de présenter les terminologies de spécialité. Même si l'établissement de tels réseaux est à la base de tout travail terminologique, la présentation des données dans les grandes banques ne permet malheureusement pas de présenter toutes les connaissances acquises lors de la recherche terminologique. En effet, la présentation sous format de fiches (une à une, chacune portant sur une notion différente) ne permet pas d'illustrer de façon visuelle les rapports entre les différentes notions traitées si ce n'est par le recours à un système de renvois qui devient rapidement lourd à consulter. La visualisation des index apporte souvent une autre dimension mais ne constitue pas pour autant un outil pédagogique puisqu'il faut déjà connaître le terme pour le chercher dans un index.

Dans un contexte de démocratisation de l'information et tout en sachant que les réseaux deviendront non seulement des outils de familiarisation avec une discipline mais encore des sources de formation professionnelle, la représentation des réseaux notionnels devient ainsi une nouvelle voie fort intéressante à explorer puisqu'elle permet d'avoir une vue d'ensemble d'un concept ou

d'un objet et de mieux comprendre les rapports entre les différentes composantes et les termes qui les expriment.

Une nouvelle présentation de la terminologie dans une forme mieux adaptée aux besoins des utilisateurs des réseaux pourrait ainsi favoriser la connaissance des termes de base d'une discipline, et les producteurs de la Banque de terminologie du Québec y voient déjà une nouvelle façon de favoriser l'implantation des termes français auprès des futurs travailleurs.

Toute fréquentation des réseaux met l'utilisateur en contact avec les liens hypertextes et hypermédias. Dans le domaine de la terminologie, l'utilisation de tels liens permet d'envisager une toute autre présentation des informations. Nous donnerons ici simplement quelques exemples qui viendront illustrer notre propos et donner un bref aperçu des possibilités nouvelles qu'offrent les hyperliens.

C'est ainsi qu'il devient possible, à partir d'un terme donné et suivant les besoins de l'utilisateur, d'obtenir soit la définition du terme, soit un contexte d'utilisation, soit une note linguistique concernant son usage, soit son domaine d'appartenance, soit son équivalent dans une autre langue ou encore le réseau notionnel auquel il appartient.

Les relations entre les termes qui désignent des notions connexes trouvent également des précisions dans leur mise en relation par des liens hypertextes. L'accès à des dossiers encyclopédiques permet de poursuivre une recherche plus poussée tandis qu'une illustration pourra être consultée en permettant l'accès aux définitions de chacune de ses composantes.

Un des atouts parmi les plus importants des hyperliens demeure le fait qu'il est possible d'établir des relations entre des sites différents mais qui peuvent présenter de l'intérêt pour un profil d'utilisateurs donné. Il

pourrait ainsi devenir possible de donner accès à la terminologie de base d'un domaine par le biais d'un site concernant les normes techniques ou faisant état de la formation disponible dans ce même domaine.

Le contenu des banques de terminologie demeura toujours le résultat des recherches terminologiques mais la redéfinition de leur structure permettra un accès aux données complètement modifié. Dans cette démarche, il faudra bien sûr tenir compte des besoins des nouveaux utilisateurs pour développer le type de produit qui correspond le mieux à leur contexte de travail.

La notion de banques virtuelles pourrait bien constituer par ailleurs une réponse aux problèmes de mise à jour des données que vivent les producteurs quand ils souhaitent ajouter de nouvelles données aux nombreuses fiches qu'on trouve déjà dans une grande banque.

Il s'agirait en fait de mettre en relation différentes banques ou encore des extraits de différentes banques afin d'offrir aux utilisateurs un ensemble de ressources, et ce, dans un seul et même accès. Cette façon d'envisager les choses permettrait de donner naissance à ce que l'on pourrait appeler des bibliothèques virtuelles spécialisées.

Vous aurez compris qu'une telle aventure nécessite non seulement des ententes entre les producteurs d'information, mais également des développements informatiques qui permettraient d'établir les liens désirés entre les différentes banques et de mettre au point les interfaces qui permettront un accès convivial et efficace à ces différentes sources d'information.

Qu'en est-il maintenant de la valeur ajoutée que pourrait présenter la diffusion de la terminologie dans une nouvelle gamme de produits adaptés aux besoins des clientèles? Il faut dire ici que c'est le principe même du respect de l'environnement

culturel des utilisateurs qui nous incite à avancer l'idée que la terminologie pourrait bien constituer une valeur ajoutée importante si l'on retenait la possibilité de la diffuser dans des véhicules d'information non terminologiques, c'est-à-dire dans des publications imprimées ou électroniques ou encore dans des sites conçus à d'autres fins que celle de la diffusion de la terminologie.

Il est impossible de décrire ici toutes les possibilités qu'il faudrait envisager mais il y aura un exercice important à faire dans ce sens. D'abord, il faudra déterminer dans quels types de sources d'information il serait le plus envisageable de diffuser de la terminologie. On pense déjà au matériel didactique, à la documentation technique accompagnant différents produits, aux nombreux disques optiques compacts qui occupent maintenant le marché dans divers domaines d'intérêt sans compter le grand nombre de sites W3 qui voient le jour. Il faudra ensuite établir des ententes avec les fournisseurs d'information concernés... Ceci nous conduit d'ailleurs directement au point suivant.

5.3 Pour la mise au point de nouvelles formules de partenariat

Quant à la réalisation du quatrième objectif, on a déjà compris qu'il sera nécessaire de mettre au point de nouvelles formules de partenariat autant avec les producteurs de terminologie qu'avec d'autres types de fournisseurs.

LES NOUVEAUX MÉCANISMES

Développer de nouvelles formules de partenariat :

- Offrir des outils de production;
- Établir un réseau de fournisseurs de données;
- Établir des relations technologiques entre les banques de terminologie;
- Ouvrir le partenariat à des producteurs d'autres types d'information.

S'il est possible de rejoindre de nouvelles clientèles par le biais des réseaux, il est aussi possible d'assurer une meilleure coordination des travaux de production terminologique. Pour ce faire, l'Office de la langue française à l'intention d'offrir des outils de production qui faciliteront le travail des partenaires, il pense pour l'instant à son poste de travail SAMI. Cela pourra se faire autant pour des projets de production spécifiques à une organisation que pour une contribution au développement du contenu de la Banque de terminologie du Québec.

L'Office compte ainsi établir un réseau de fournisseurs de données, ce qui permettra d'utiliser au mieux les rares ressources qui sont encore disponibles et de favoriser la diffusion des données terminologiques. On sait en effet que certaines données connaissent une diffusion fort limitée quand pourtant elles présentent un intérêt certain pour la collectivité. L'inventaire qui a été réalisé récemment concernant les banques de terminologie informatisées confirme d'ailleurs la richesse des ressources existantes.

La mise en relation de différentes banques, comme cela a été évoqué plus tôt, nécessite pour sa part l'obligation d'établir ce que j'appellerais des relations technologiques entre les banques. Je n'entrerai pas dans les détails ici mais

chacun peut déjà imaginer les défis qu'il faudra relever.

Enfin, les hypothèses de diffusion de la terminologie dans des productions non terminologiques implique le fait qu'il faudra étendre le partenariat à des producteurs d'autres types d'information. Tout cela reste toutefois encore à définir mais plusieurs y trouveraient sûrement leur intérêt. C'est probablement ainsi d'ailleurs que nous arriverons tous ensemble à atteindre le sixième objectif qui est de contribuer au développement du contenu des autoroutes de l'information et notamment du contenu d'expression française.

6 Les banques de terminologie d'hier à demain

Il nous est donc permis de croire que l'évolution des banques de terminologie passera par une redéfinition de plusieurs de leurs composantes actuelles. Nous présentons ici des tableaux mettant en évidence les différences entre les banques d'hier et celles de demain.

Les banques de terminologie pourront jouer un tout nouveau rôle dans la société de l'information dont elles font partie :

LE RÔLE D'HIER	LE RÔLE DE DEMAIN
<ul style="list-style-type: none"> - Un outil de compilation et de diffusion de données documentaires et terminologiques. - Assistance à la rédaction; - Assistance à la traduction. 	<ul style="list-style-type: none"> - Un outil de production et de diffusion de données; - Un outil de coordination; - Un outil de mise en relation de différentes sources; - Un outil de création de différents produits d'information; - Un outil de rétroinformation. - Assistance à la rédaction; - Assistance à la traduction; - Assistance au développement des connaissances; - Assistance à la formation; - Assistance à la réalisation d'autres types de produits; - Contribution au développement du contenu des autoroutes, notamment du contenu d'expression française.

Les banques de terminologie pourront élargir la gamme de leurs utilisateurs :

LES UTILISATEURS D'HIER	LES UTILISATEURS DE DEMAIN
<ul style="list-style-type: none"> - Les langagiers: terminologues, traducteurs, réviseurs; - Les rédacteurs. 	<ul style="list-style-type: none"> - Les langagiers; - Les rédacteurs; - Les enseignants; - Les formateurs; - Les chercheurs; - Les professionnels de toutes disciplines; - Les étudiants; - Les concepteurs de matériel didactique; - Les concepteurs de produits divers; - Les concepteurs de sites Internet; - Les amateurs de réseaux.

Les banques de terminologie verront leur structure modifiée:

LA STRUCTURE D'HIER	LA STRUCTURE DE DEMAIN
<ul style="list-style-type: none"> - Une structure fixe; - Une présentation des fiches par notion. 	<ul style="list-style-type: none"> - Une structure souple, adaptable, multiformes; - Une présentation des réseaux notionnels; - Une utilisation des hyperliens.

Les banques de terminologie développeront de nouveaux mécanismes d'alimentation:

LES MÉCANISMES D'ALIMENTATION D'HIER	LES MÉCANISMES D'ALIMENTATION DE DEMAIN
<ul style="list-style-type: none"> - Les résultats des recherches terminologiques effectuées pour l'organisme; - Les publications imprimées; - Les publications électroniques; - Le transfert de fichiers. 	<ul style="list-style-type: none"> - Les mécanismes déjà connus; - Le développement d'un réseau de fournisseurs; - La création de banques virtuelles.

Les banques de terminologie pourront multiplier leurs mécanismes de diffusion des données terminologiques :

LES MÉCANISMES DE DIFFUSION D'HIER	LES MÉCANISMES DE DIFFUSION DE DEMAIN
<ul style="list-style-type: none"> - L'accès direct; - Les disques optiques compacts. 	<ul style="list-style-type: none"> - Les mécanismes déjà connus; - La diffusion dans les autoroutes de l'information; - La mise en relation des sites; - La présentation sous plusieurs formes adaptées; - L'exploitation du contenu des banques pour la production d'autres types d'information.

Conclusion

Pour terminer, je dirai simplement qu'il est évident pour tous que les banques de terminologie se trouvent au cœur d'un virage technologique. J'ajouterai cependant qu'un véritable virage philosophique est indispensable pour que les banques de terminologie puissent exploiter pleinement leurs ressources et connaître un nouvel essor.

S'il est vrai que nous assistons à la démocratisation de l'information, il est sans doute temps que nous mettions tout en place pour assurer aussi la démocratisation de la terminologie. C'est sans doute ainsi que les banques de terminologie apporteront une contribution active au développement de cette nouvelle économie qu'on dit du savoir.

*Mireille Lacasse,
Office de la langue française,
Québec.*

Des banques de données terminologiques en Afrique francophone

Le projet d'implantation et de développement de banques de données terminologiques en Afrique est un programme d'action du Rint. La présente étude préliminaire balise la question en clarifiant les notions de base de données *versus* banque de données, en justifiant le projet, en décrivant les besoins à quoi ce projet devra répondre, en traçant enfin les activités à mener dans les institutions pour que la banque soit correctement gérée et rentabilisée. Ces activités concernent la collecte, le traitement, l'enregistrement, l'exploitation, la publication et la diffusion des données, aussi bien que la promotion de l'activité terminologique et traductionnelle dans la communauté linguistique.

Termes-clés :
banques de données ;
terminologie ; Afrique ;
développement terminologique.

1 Quelques précisions terminologiques

Dans la perspective d'une action concertée du Rint en vue de soutenir le développement de l'activité terminologique en Afrique francophone, nous posons ici les premiers jalons d'une réflexion à poursuivre sur les conditions de création de banques de données terminologiques dans ces pays. Il n'est pas inutile, pour éviter toute équivoque, de rappeler quelques précisions terminologiques.

L'expression *Afrique francophone* désigne l'ensemble des pays d'Afrique, y compris les îles adjacentes politiquement indépendantes, où le français fait partie des langues en usage.

Si une base de données et une banque de données sont toutes les deux des systèmes de gestion de données structurées et soigneusement enregistrées dans des fichiers, elles renvoient à deux concepts différents lorsqu'on les met en contraste. La base de données est d'abord utilisée par le terminologue/lexicographe pour collecter un maximum de données et d'informations sur ces données. Ces informations peuvent être incomplètes ou revêtir un caractère confidentiel, personnel, ou provisoire, elles seront consignées dans la base de données ; tandis que, pour une banque de données, il est essentiel que toutes les informations enregistrées soit complètes, correctes, et validées. En outre, la banque de

données est destinée à une exploitation extensive par des utilisateurs extérieurs à l'institution qui abrite la banque, à savoir les professionnels du langage et ceux des domaines de spécialité et, en dernier ressort, le grand public désireux de connaître le mot juste. On peut donc dire qu'en tant que système informatif, la banque de données est à la base de données ce qu'est le dictionnaire publié par rapport aux boîtes de fiches du lexicographe.

Ces différences caractéristiques entre base et banque de données sont naturellement traduites dans les types de logiciels proposés sur le marché informatique. Les gestionnaires de bases de données, que nous qualifierions de « classiques », sont très largement orientés vers la collecte et l'organisation de données avec un minimum de fonctions d'import/export. Elles répondent plutôt aux besoins du lexicographe et peuvent le plus souvent tenir sur un poste de travail individuel, tandis que les gestionnaires de banques de données intègrent de nombreuses fonctions pour faciliter l'utilisation de la banque par les usagers les plus divers, tout en protégeant davantage l'intégrité des données. Il est donc clair que de tels logiciels nécessitent davantage une installation réseau et un serveur pour une exploitation optimale. C'est pourquoi on peut se demander si l'infrastructure informatique et les conditions de travail en Afrique permettent d'envisager la création de banques de données et pour quoi faire ?

2 L'analyse des motivations

Il faut d'abord souligner que toute la recherche terminologique, au Nord comme au Sud, a pour objectif le développement contrôlé de la langue en vue d'un développement économique, social et culturel de la communauté linguistique tout entière. En Afrique francophone en particulier, où le français n'est la langue maternelle d'aucune communauté autochtone tout en jouissant d'un statut dominant en tant que langue du pouvoir et du savoir, les besoins en communication de masse, tant pour l'information et l'éducation des populations que pour la formation des agents de développement, le transfert des connaissances et des technologies, nécessitent, de façon tout à fait incontournable, l'utilisation des langues africaines.

C'est pourquoi l'essentiel du travail terminologique effectué dans ces pays porte en priorité, mais sans exclusive, sur les langues africaines. Il s'agit bien souvent de doter celles-ci du vocabulaire technique dont elles ont besoin pour exprimer des connaissances scientifiques et techniques déjà bien établies et considérées comme classiques dans le Nord afin de les rendre plus accessibles à la population au moyen de programmes de formation de masse, tandis que dans les pays du Nord, l'activité terminologique concerne essentiellement les domaines de recherches avancés et des technologies nouvelles, voire de pointe. Elle laisse aux dictionnaires généraux ou professionnels le soin de consigner les termes classiques des connaissances de base. Il s'ensuit que, dans la plupart des cas, on est amené à constituer un corpus spécifique de référence en français dans le cadre même du travail de recherche terminologique sur les langues africaines. C'est dire combien il est

nécessaire de créer des banques de terminologie adaptées aux besoins propres du Sud et qu'il faut implanter au Sud, près des utilisateurs. Le fait que certaines banques du Nord se donnent assez de souplesse pour accueillir des données dans quelques langues africaines, ne constitue absolument pas une solution de rechange.

En effet, l'implantation d'une banque de données dans un institut en Afrique dynamisera nécessairement le travail terminologique de cet institut et son rayonnement dans l'environnement social. La réalisation de produits terminologiques à partir de la banque de données (produits pratiques et variés tels que lexiques, plaquettes, dépliants, vocabulaires, affiches, tout support de diffusion de la terminologie) est d'autant plus essentielle que le support papier est de loin le plus répandu et le plus pratique des supports en Afrique. Une telle activité ne peut se développer à distance tant que les coûts de communication resteront prohibitifs pour la majorité des pays du Sud. C'est pourquoi, il est préférable de créer des banques de terminologie sur place, quitte à les équiper progressivement d'un accès Internet au fur et mesure que le réseau de télécommunication local le permettra.

3 Le site institutionnel

Idéalement, dans chaque pays francophone du Sud où existe une structure de travail en terminologie, il devrait à terme se créer une banque de données terminologiques sur décision de la structure elle-même. On en est encore loin. Une solution intermédiaire envisagée un moment par l'ACCT, avait consisté en la création de laboratoires régionaux comme support technique aux Offices des langues. On y formerait leur

personnel (assistants de recherche, techniciens, informaticiens et administratifs) et on y rassemblerait les équipements collectifs. Le projet avait envisagé quatre laboratoires : pour l'Afrique Centrale, l'Afrique du Nord, l'Afrique de l'Ouest et l'Océan indien.

Il faut observer qu'il s'agit là d'institutions régionales à statut international et que de telles institutions existent déjà en Afrique (CELTHO [Centre d'études en littérature, histoire et tradition orale], CERDOTOLA [Centre régional de documentation sur les traditions orales et les langues africaines], BASE [Bureau d'études en sciences de l'éducation], EACROTANAL [Eastern African Centre for Research Oral Traditions and African National Language]) et connaissent toutes un fonctionnement pour le moins difficile, essentiellement par manque de support financier. C'est pourquoi nous sommes plutôt favorables à un plan de développement progressif de toutes les institutions nationales de terminologie, en soutenant de façon plus appuyée celles qui font preuve d'un dynamisme notoire ou qui ont à leur charge une responsabilité collective comme d'assurer la coordination d'un module collectif du Rint.

Dans cette hypothèse et sous toutes réserves, nous suggérons comme bons candidats à l'implantation d'une banque de données terminologiques : l'IERA (Institut d'études et de recherche pour l'arabisation [Maroc]), le CLAD (Centre de linguistique appliquée de Dakar [Sénégal]), Le Centre des Langues de l'Académie malgache (Madagascar), l'ILA de Bangui (Centrafrique), le CELTA (Centre de linguistique théorique et appliquée [Zaire]). Une deuxième étape pourrait concerner les institutions des pays d'Afrique de l'Ouest qui travaillent sur des langues de grande diffusion : Niger, Mali,

Burkina, Mauritanie, Côte d'Ivoire, Guinée. Pour cette dernière liste, l'implantation de la banque de données est à étudier cas par cas en tenant compte des réalités locales de chaque pays.

4 L'analyse des besoins

4.1 Multilinguisme

La situation sociolinguistique africaine est toujours multilingue, même dans les rares cas où le pays n'a qu'une seule langue locale (Madagascar, Burundi, Rwanda), le français fait toujours partie du paysage, surtout dans les domaines de spécialité qui nous intéressent ici. Il s'ensuit que la banque de données doit être multilingue.

4.2 Typographie multilingue

La gestion des caractères spéciaux (phonétiques, arabes, etc.) est très importante et doit absolument figurer en bonne place dans les critères déterminant le choix d'un système intégré de logiciels. Par exemple, dans plusieurs institutions d'Afrique de l'Ouest et du Nord, il est essentiel de pouvoir mélanger dans le même texte des caractères latins, phonétiques et arabes, avec changement automatique de sens de l'écriture.

4.3 Communication

Plusieurs langues africaines sont parlées dans plusieurs pays où elles jouent un rôle véhiculaire notoire. Elles sont définies par l'ACCT comme «langues transnationales» (ex. : kiswahili, lingala, hausa, fulfulde, bambara, wolof, etc.) et sont l'objet d'un soutien prioritaire auquel sont associées les langues

d'importance nationale ou officielle (ex. : malgache, sângö, kirundi, kinyarwanda, etc.). La coopération entre les pays travaillant sur les mêmes langues doit être préservée, voire intensifiée par le développement des banques de données terminologiques. Une attention particulière devra être portée à l'instauration de dispositifs multiples de communication entre les banques, compte tenu des avantages et des contraintes de chaque dispositif : échange de documents primaires par courrier, échange de disquettes, télécopie, réseaux (Syfed, Rio, Internet).

4.4 Équipement informatique

Le parc informatique en Afrique, on le sait, est sans commune mesure avec ce qu'il est dans les pays développés du Nord. Dans les modules Rint du Sud, l'équipement informatique est, dans la plupart des institutions, non seulement minimal mais souvent obsolète. Tout projet de création d'une banque de données terminologiques devrait donc comporter un volet équipement informatique dont le détail serait à étudier pour chaque cas. Pour l'essentiel, il faudra probablement prévoir au moins une machine puissante de type serveur. Une étude technique de prospection devrait pouvoir conduire à des propositions concrètes.

4.5 Formation

Les institutions qui font de la terminologie comptent, bien sûr, des terminologues dans leur personnel. Plusieurs d'entre eux ont suivi ou doivent encore suivre des stages de formation pour accroître leur compétence ou acquérir de nouvelles connaissances utiles à leur travail. Avec l'installation d'une banque de

données, il est très utile qu'au moins un intégrateur de technologies informatiques fasse partie du personnel. Il pourrait assurer l'assistance technique des utilisateurs de la banque, la gestion du parc informatique et des communications interbanques, le suivi des logiciels, la formation permanente des agents, etc. Il faudrait élaborer un plan de formation pour les trois catégories d'agents suivants : les terminologues, les administratifs (gestionnaires de l'activité terminologique) et les techniciens (intégrateurs, informaticiens gestionnaires de réseaux, documentalistes). Un tel plan de formation pourrait être préparé et réalisé en concertation avec d'autres organismes partenaires du Rint (ACCT, Riofil, l'Aupelf-Uref, etc.) dans le seul souci d'optimiser les moyens qu'apportent les uns et les autres.

4.6 Dimensions juridiques

Étant donné que les banques de données doivent échanger entre elles, il n'est pas superflu d'intégrer d'emblée tous les aspects juridiques dans les études préparatoires du plan et d'en tenir compte dans le plan lui-même.

5 Le développement de l'activité terminologique

5.1 Au sein de l'institution

L'implantation d'une banque de terminologie représente un développement certain de l'activité terminologique au sein d'une institution. On peut schématiquement décrire comme suit les différentes phases de cette activité : collecte, traitement, archivage,

récupération, publication, diffusion. À chacune de ces phases correspond des tâches particulières requérant des logiciels et du matériel spécifiques.

5.1.1 La Collecte des données

Elle comporte aussi bien l'enregistrement de documents oraux (interviews de praticiens, techniciens et autres spécialistes) que le dépouillement de documents écrits. La constitution d'une collection d'objets à des fins d'identification (minéraux, végétaux, pièces détachées, échantillons, photos, dessins, cartes, etc.) fait partie de cette phase. Selon la tâche, on peut recourir à un appareil photo, une caméra, un scanner, un traitement de texte, un logiciel de dépouillement comme Termino, etc.

5.1.2 Le traitement des données

Les données collectées sont généralement élaguées, nettoyées, caractérisées, classées, annotées, décrites, discutées, éventuellement corrigées et, enfin, consignées dans des bases de données. La base de données est certainement le logiciel le plus essentiel pour cette phase, avec ses utilitaires d'édition, de comparaison, de concordance, de recherche, d'intégration, etc. Plusieurs cycles d'opérations font généralement partie du traitement.

5.1.3 L'archivage des données

On peut certes archiver des données brutes en vue de les traiter plus tard, mais ce n'est pas de cet archivage-là qu'il s'agit ici. Une fois les données traitées et validées, elles peuvent être archivées dans une banque de données pour être mises au service des utilisateurs. Le cœur d'une banque de données est naturellement de même nature qu'une base de données. Mais ici, l'ouverture sur la clientèle extérieure est essentielle.

5.1.4 La récupération des données

Cette tâche est généralement servie par une ou plusieurs fonctions implémentées dans les systèmes de gestion des bases de données. Cependant, il arrive souvent qu'on ait besoin de logiciels spécialisés dans cette opération lorsqu'il ne s'agit pas seulement d'appliquer quelques filtres assortis de requêtes simples. Ces logiciels peuvent se montrer assez intelligents pour gérer des accès multiples et simultanés à une même banque de données, optimiser les parcours de recherche, conserver la mémoire des sessions, voire dialoguer oralement avec l'utilisateur si l'équipement s'y prête. Mais, d'ordinaire, les points les plus importants à vérifier sont : la capacité de convertir les fichiers dans la majorité des formats standard existants, la capacité de conserver les caractères spéciaux lors de la récupération ou des conversions, la capacité de préserver l'intégrité des données.

5.1.5 La publication des résultats

Toute l'activité terminologique devra aboutir à des publications sous quelque forme que ce soit. Il faut bien que l'institution s'en préoccupe. Et comme nous l'avons mentionné plus haut, une banque de données est la source toujours renouvelée de nombreuses publications, des plus anodines aux plus encyclopédiques selon les objectifs de l'aménagement linguistique ou la demande du marché. Beaucoup de ces travaux requièrent seulement une bonne formation en PAO (publication assistée par ordinateur) assortie d'un équipement standard tant en logiciels qu'en matériel. Mais pour des travaux plus sophistiqués, il vaut mieux s'adresser à des éditeurs professionnels.

5.1.6 La diffusion de la terminologie

Toutes les stratégies d'implantation des terminologies normalisées relèvent de cette opération de diffusion. Il est de la responsabilité de l'institution de l'organiser par tous les moyens de communication en direction de tous les utilisateurs. Un service de documentation devrait exister au sein de l'institution pour s'en occuper.

5.2 À l'extérieur de l'institution

5.2.1 Groupes de terminologies

Il appartient aux institutions gérant une banque de données de déployer tous les efforts nécessaires pour organiser le développement de l'activité terminologique dans d'autres institutions du pays ou de la région. Par exemple, dans le cas d'une action à portée nationale, l'institution pourrait susciter, organiser, ou animer des groupes de terminologie, quels que soient leurs noms, dans des structures telles que les ministères, les partis politiques, les parlements, les mairies, les tribunaux, les hôpitaux, les ONG, les associations, les centres religieux, les syndicats, les corporations professionnelles, les presses, les radios et télévisions, les universités, les centres de recherche, etc.

Le développement des groupes de terminologie est l'un des facteurs favorables à la création d'une banque de données terminologiques là où il n'y en a pas. Inversement, si la banque de terminologie est créée en premier, il faudrait tout mettre en œuvre pour susciter la création et la multiplication de tels groupes. Ceci revient, en réalité, à promouvoir la terminologie dans toutes les institutions publiques et privées du pays, de façon à créer les conditions d'une implantation rapide et durable des termes normalisés.

5.2.2 Groupes de traduction

L'activité terminologique trouve son aboutissement dans la traduction massive et continue d'ouvrages de techniques, pédagogiques, littéraires ou utilitaires de toutes sortes. C'est une condition essentielle de modernisation et de diffusion du savoir et du savoir-faire. L'activité traductionnelle ne devrait pas être laissée au hasard des initiatives individuelles, mais bien au contraire, elle devrait être organisée et conduite de façon systématique dans des institutions publiques spécialisées (ex. : institut de traduction), ou tout au moins des services où œuvrent des groupes de traducteurs professionnels. De tels groupes sont les clients privilégiés d'une banque de données terminologiques.

6 Les actions à suivre

Au terme de cette étude préliminaire, nous retiendrons ceci : il est nécessaire de créer des banques de

données terminologiques en Afrique francophone, pour soutenir l'activité terminologique des modules du Sud. À cette fin, il convient de mener une étude plus approfondie devant aboutir à l'élaboration :

- D'un plan d'implantation progressive des banques de données terminologiques. Ce plan devrait établir, en accord avec les modules, une ou plusieurs listes ordonnées de sites d'actions prioritaires ;
- D'un plan d'équipement en logiciels essentiels et en matériel informatique nécessaire pour chaque banque implantée, ainsi qu'une étude plus détaillée des dispositifs de communication entre banques, ou entre institutions... ;
- D'un plan de formation pour les ressources humaines, si possible en concertation avec d'autres organismes francophones.

*Marcel Diki-Kidiri,
LLACAN,
CNRS,
France.*

Une base de données multimédia : le dictionnaire descriptif et visuel de Parcs Canada

L'article décrit la méthode utilisée dans l'élaboration du dictionnaire visuel de Parcs Canada qui est un thésaurus destiné au classement des artefacts conservés dans les sites historiques de cette institution. Le mode de rédaction des définitions est analysé dans le cadre de l'approche systémique utilisée pour mieux identifier les incluants et maintenir la cohérence de la nomenclature. Les problèmes théoriques soulevés par l'application des principes de la terminologie sont mentionnés.

Termes-clés :
terminologie de la culture matérielle ; vocabulaires contrôlés ; documentation muséologique ; définition terminologique.

1 La nomenclature

Le *Dictionnaire descriptif et visuel de la collection des objets du patrimoine/The Descriptive and Visual Dictionary for Historical Collections* de Parcs Canada, dont la version anglaise est établie par un groupe de recherche interdisciplinaire à l'Université d'Ottawa, et la version française à Québec, a pour but d'améliorer la gestion des collections historiques de cette institution. Ces collections comportent plus de 500 000 artefacts appartenant à ce qu'on appelle communément la culture matérielle, c'est-à-dire l'ensemble des objets façonnés par l'homme pour contrôler son environnement physique et humain.

Le dictionnaire est aussi un thésaurus destiné au classement des artefacts de la culture matérielle à partir d'une nomenclature établie par Parcs Canada (Service canadien des parcs, 1992; Canadian Parks Service, 1992). Il a donc une fonction documentaire qui s'ajoute à sa fonction terminologique ou lexicographique. Le présent article concerne la méthode utilisée pour l'établissement de la version anglaise du dictionnaire.

Le système de classification, reproduit ci-après (Service canadien des parcs, 1992: p. 18-20), est fondé sur la nomenclature destinée à la classification des artefacts établie par Chenhall (1978), et révisée par Blackaby (1988) ainsi que ses collaborateurs. Dans la mesure où les

artefacts se définissent essentiellement par leur fonction, il repose sur une hiérarchisation des dénominations d'objets à partir du critère de la fonction originelle de ces derniers.

Le système se subdivise en trois niveaux: les catégories, les classes et sous-classes ainsi que les dénominations d'objets. Dénominations des notions individuelles, classes et catégories correspondent à des fonctions particulières. C'est ainsi que la *corde à linge* sert à pendre des vêtements à sécher. Ce terme fait partie de la classe des *outils et équipements servant à l'entretien*, c'est-à-dire aux activités de lavage, classe qui fait elle-même partie de la catégorie des *outils et équipements de science et de technologie*. Cette dernière comporte les outils, équipements et fournitures utilisés pour l'observation des phénomènes naturels ou pour l'application de connaissances tirées de l'observation et, dans le cas de la *corde à linge*, les outils élaborés en réponse à un ensemble de connaissances sur la façon de prendre soin des objets, en l'occurrence les vêtements.

L'ensemble des 10 catégories de la nomenclature comprend près de 7000 termes préférentiels sélectionnés lors d'un processus de normalisation accompli par Blackaby et Chenhall, préalablement à la constitution du dictionnaire, et adaptés à la réalité canadienne.

Comme toute nomenclature destinée aux musées, le système de classification de Parcs Canada a essentiellement une double fonction. D'une part, elle fournit aux conservateurs et aux archivistes un instrument de recherche leur

facilitant la connaissance de l'objet et, par conséquent, les opérations d'identification, de classement et de catalogage; d'autre part, elle fournit aux gestionnaires des collections un outil de gestion permettant de rationaliser les activités documentaires et les procédures relatives au contrôle et à la circulation des objets. L'usage de la nomenclature de termes normalisés rend aussi possible la centralisation de l'information muséologique par l'informatisation des données, opération qui est nécessaire en raison de la multiplicité des sites historiques de Parcs Canada, lequel comprend six entrepôts, 114 sites historiques et plus de 500 000 objets disséminés à travers le Canada.

Par rapport à la perspective de la terminologie, la nomenclature de Parcs Canada doit être envisagée comme un système de classification documentaire, établi par déduction à partir du critère de la fonctionnalité, et non comme un arbre de domaine établi par induction à partir des notions individuelles.

2 La fiche terminologique

Les entrées du dictionnaire ont été constituées selon les principes généralement admis en terminologie voulant que chaque notion fait l'objet d'une seule fiche en fonction de la règle de l'uninotationalité. La fiche terminologique, qui a déjà été décrite dans le numéro 11 de la revue *Terminologies nouvelles* de juin 1994: p. 19 à 22 (Pepermans, 1994), et dont nous présentons deux exemples à la fin de cet article, comporte trois types de définitions correspondant à trois champs de la fiche:

– La définition ostensive qui consiste en une représentation picturale de l'objet; il s'agit soit de photographies provenant des fiches catalographiques, soit d'illustrations représentant les

traits notionnels de la notion figurant sur la fiche;

– La définition analytique, par genre prochain (ou incluant) et différence spécifique, qui identifie les caractéristiques morphologiques ou physiques de l'objet défini (les traits intrinsèques) de la notion;

– La définition synthétique qui identifie les caractéristiques extrinsèques de la notion, sa relation à un autre objet, par la description du but, de la fonction, de l'origine ou du domaine se rapportant à l'objet défini. La définition fonctionnelle révèle la dimension la plus importante des notions appartenant à la nomenclature, car non seulement il s'agit d'artefacts, mais de plus elle correspond étroitement aux critères sur lesquels cette dernière est fondée.

La fiche terminologique de Parcs Canada comporte, en plus des champs définitoires, des sections donnant à l'utilisateur des renseignements sur la classe et la catégorie auxquels appartient l'objet défini, avec les codes correspondants, ainsi qu'un champ pour la consignation des synonymes. De plus, un champ «exclusions» est destiné à accueillir les termes à ne pas confondre avec celui qui figure en vedette.

3 La méthodologie de la recherche

Comme c'est le cas pour la plupart des projets terminologiques, la méthodologie de la recherche obéit à une double dimension, à la fois sémasiologique et onomasiologique. Ces deux perspectives interviennent à des moments divers de la recherche en fonction des particularités de la documentation et l'état d'avancement des travaux.

Les documents lexicographiques (dictionnaires généraux, vocabulaires et glossaires spécialisés) ainsi que les

documents non lexicographiques (manuels, catalogues, encyclopédies, monographies et fiches catalographiques) sont traités à partir de l'unité terminologique figurant sur la fiche terminologique, c'est-à-dire en sémasiologie. Cependant, ces documents comportent des renseignements importants sur la réalité extra-linguistique représentée par chaque notion. C'est notamment le cas des renseignements encyclopédiques fournis par les descriptions d'objets, et surtout les photographies d'artefacts qui se trouvent sur les fiches de catalogage. Le définisseur dispose ainsi de documents iconographiques qui lui permettent de compléter ses définitions, dans une perspective onomasiologique, à partir de l'observation de l'objet.

Les définitions du dictionnaire sont terminologiques en ce sens qu'il s'agit de définir des objets qui se trouvent dans les magasins de Parcs Canada et non de concepts généraux. Les notions sont, par conséquent, et dans la mesure du possible, traitées à partir de la dimension extra-linguistique et comme appartenant aux différents domaines de la connaissance technique. À ce propos, les fiches catalographiques qui contiennent non seulement une photographie de l'objet, mais aussi sa description, conformément aux normes de l'ICOM (International Council of Museums), jouent un rôle important comme sources de l'activité définitionnelle.

Les traits essentiels des notions de la culture matérielle sont la fonction, l'usage, la forme, la dimension et, dans le cas des machines ou des outils, le fonctionnement, ce dernier étant envisagé par rapport aux critères intrinsèques ou extrinsèques de la notion. Dans le cas des critères intrinsèques, le fonctionnement participe de la description physique: il est mentionné dans la définition

analytique; et dans le cas des critères extrinsèques, il participe de la fonction et de son rapport à un objet qui lui est extérieur: il est donc mentionné dans la définition synthétique. Le style de l'objet n'est pas retenu comme critère définitoire.

Un soin particulier est apporté à la cohérence de la nomenclature tant du point de vue de la définition analytique que de la définition fonctionnelle. L'approche systémique permet en effet de définir les notions selon les systèmes conceptuels et de mieux identifier les incluants lors d'une première rédaction des définitions analytiques. Ces dernières doivent refléter dans la mesure du possible les rapports logiques et ontologiques (relations génériques et partitives) existant entre les notions de la culture matérielle. Cette systématisation n'est cependant pas toujours possible à réaliser de façon satisfaisante en raison de l'état des collections de Parcs Canada qui ne comportent pas tous les objets susceptibles de faire partie d'un même réseau notionnel. D'autre part, il existe dans la nomenclature des termes isolés qui ne figurent apparemment dans aucun réseau. À moins de pouvoir identifier le réseau dont les notions font partie, une approche lexicographique plus traditionnelle s'impose alors au définisseur.

Lors d'une seconde rédaction, les définitions sont systématisées au niveau de l'ensemble d'une classe, à partir des notions individuelles, sous la forme d'un arbre de domaine. Cette méthode permet d'identifier des niveaux intermédiaires entre les rubriques fonctionnelles de la nomenclature et les réseaux notionnels logiques et ontologiques établis lors de la première rédaction. Elle permet aussi de compléter par une approche inductive les classes et les catégories du thésaurus documentaire. Le modèle utilisé pour

établir les rubriques intermédiaires est le *Art & Architecture Thesaurus* (1994) de The Getty Art History Information Program qui prévoit l'établissement de rubriques à l'aide du critère de la fonctionnalité. Ce dernier est compatible avec celui qui fut utilisé pour déterminer les descripteurs du thésaurus de Parcs Canada. L'arbre de domaine consiste donc à établir un lien entre les approches inductive et déductive par la constitution de rubriques permettant d'établir des raccords entre, d'une part, les réseaux notionnels fondés sur des critères logiques et ontologiques et, d'autre part, les descripteurs fonctionnels du thésaurus documentaire. Les notions et les définitions correspondantes sont donc traitées en amont et en aval en vue d'obtenir un maximum de précision dans la détermination des incluants.

4 Les problèmes théoriques

Un ensemble de problèmes théoriques qui ont surgi au cours de la recherche, principalement en raison des caractéristiques lexicosémantiques du vocabulaire de la culture matérielle, retiennent l'attention de l'équipe chargée du projet.

Comme nous l'avons déjà montré dans un article précédent (Pepermans: 1995), la règle de l'unioctionnalité n'est pas toujours suivie à la lettre lors de la constitution des entrées. En effet, la rigidité du système de codage des termes en vigueur, en raison des impératifs de la normalisation, ne permet pas l'établissement de plusieurs fiches pour les termes polysémiques ou pour rendre compte des variations dans le temps et dans l'espace d'une même notion. Des fiches multiconceptuelles sont donc rédigées dans certains cas. De même, la fiche monosémique ne

rend pas compte d'une manière satisfaisante du phénomène de la synonymie complexe; les vrais synonymes, les quasi-synonymes sont consignés dans le même champ en raison du manque de flexibilité du principe de la monosémie. Cependant, les faux synonymes sont dans la mesure du possible consignés dans le champ «exclusions».

En outre, à propos du principe de la fonctionnalité de l'objet, il n'est pas toujours facile d'établir une distinction entre usage et fonction. Une solution à ce problème dépendra d'une analyse à la pièce de toutes les définitions synthétiques rédigées par les définisseurs et d'une étude de la littérature anthropologique portant sur le sujet.

*Raymond Pepermans,
École de traduction et d'interprétation,
Université d'Ottawa,
Canada.*

5 Bibliographie

- Art and Architecture Thesaurus*, 1994, New York, Oxford University Press.
- Blackaby (James A.), Greeno (Patricia) and the Nomenclature Committee, 1988: *The Revised Nomenclature for Museum Cataloging. A Revised and Expanded Version of Robert G. Chenhall's System for Classifying Man-Made Objects*, Nashville, Tennessee, American Association for State and Local History.
- Canadian Parks Service, 1992: *The Canadian Parks Service Classification System for Historical Collections*, Ottawa, Canada Communication Group.
- Chenhall (Robert G.), 1978: *Nomenclature for Museum Cataloging. A System for Classifying Man-Made Objects*, Nashville, Tennessee, American Association for State and Local History.
- Pepermans (Raymond), 1994: «La terminologie au service de la communauté muséale: le dictionnaire visuel du Service canadien des parcs»,

dans *Terminologies nouvelles*, n° 11,
p. 19-22.

Pepermans (Raymond), 1995: «Material Culture in the Computer Age. An Assessment of the Parks Canada Terminology Record as a Means of Storing Conceptual Data», dans Bearman (David) (Ed.), *Multimedia Computing and Museums: Selected Papers from the Third International Conference on Hypermedia and Interactivity in Museums (ICHIM'95 — MCN'95)* (San Diego, California, October 9-13, 1995), Archives and Museum Informatics, Pittsburgh, p. 310-321.

Service canadien des parcs, 1992: *Système de classification des collections du Service canadien des parcs*, Ottawa, Groupe Communication Canada.

Les banques de terminologie de services linguistiques

De nombreux services linguistiques et des traducteurs en pratique privée désirent posséder leur propre banque de terminologie pour y consigner les recherches terminologiques qu'ils ont effectuées. Le présent article porte sur les critères à considérer dans le choix d'un logiciel et sur la création de la fiche terminologique en fonction des caractéristiques du logiciel.

Termes-clés :
banque de terminologie ;
service linguistique ;
fiche terminologique.

1 Introduction

L'atelier de ce matin est consacré au développement et à la diffusion des banques de terminologie, un vaste sujet qu'il est difficile d'explorer à fond en quelques heures. J'ai choisi d'aborder la question des banques de terminologie de services linguistiques en me limitant, toutefois, au choix du logiciel et à la création de la fiche terminologique. J'aimerais tout d'abord souligner que dans le présent exposé, l'expression « banques de terminologie de services linguistiques » désigne les banques des services linguistiques dans les entreprises et celles des traducteurs en pratique privée.

2 Les banques de terminologie : outil indispensable

Dans les services linguistiques au Québec et au Canada, l'existence des banques de terminologie remonte au début des années 80, peu après l'apparition des grandes banques publiques que sont Termium, des Services gouvernementaux du Canada, et la BTQ (Banque de terminologie du Québec). Leur nombre n'a cessé d'augmenter et un récent sondage m'a permis de constater que les banques sont devenues l'outil de base du traducteur. Celui-ci les consulte avant d'explorer toute autre source.

Plusieurs raisons expliquent ce phénomène : la quantité et la variété des textes à traduire, les nombreux nouveaux concepts à nommer, le virage informatique bien amorcé dans les services linguistiques, les délais de traduction très courts et la nature des textes qui déborde souvent le domaine précisé par le client. Par exemple, il y a quelques semaines, un traducteur qui avait accepté de traduire un texte en économie a constaté qu'il y était question en bonne partie de téléphonie cellulaire, domaine qui lui était peu familier. Aux raisons que je viens d'énumérer, s'ajoutent le rétrécissement du réseau de personnes-ressources que le traducteur avait établi au fil des ans et le nombre de plus en plus restreint de services de consultation des organismes publics. Si l'information n'est pas à portée de clavier, les traducteurs doivent effectuer eux-mêmes leur recherche, ce qui n'est pas toujours facile étant donné la diversité des domaines et les délais.

Il n'est donc pas étonnant que la consignation de la terminologie qui a fait l'objet d'une recherche soit devenue pratique presque courante. Dans les services linguistiques, la banque de terminologie est un outil complémentaire aux grandes banques. En général, on n'y consigne pas l'information que l'on trouve dans ces dernières.

Je fais une distinction entre les grandes banques, par exemple la BTQ ou Termium, et les banques des services linguistiques. Les premières contiennent plus d'un million de termes appartenant à de nombreux domaines. On y a accès par abonnement. Lorsqu'on les consulte,

on souhaite y trouver, outre l'équivalent, tous les renseignements utiles, notamment une définition ou un contexte, une source fiable et des marques d'usage (régionalisme, néologisme, anglicisme...), s'il y a lieu.

Les banques de services linguistiques sont beaucoup plus petites. Elles permettent de retrouver rapidement un terme déjà employé, ou la terminologie propre à l'entreprise ou à un service client, afin d'assurer l'uniformité si le texte est réparti entre plusieurs personnes ou s'il est confié à un traducteur qui connaît moins bien le domaine. Certaines banques contiennent plus de 50 000 fiches, mais la plupart sont de taille plus modeste. Le nombre de fiches varie généralement en fonction du nombre de terminologues et de traducteurs que compte le service linguistique.

Certains traducteurs en pratique privée ont, eux aussi, de « belles » banques de terminologie qui contiennent plusieurs milliers de fiches, souvent dans des domaines de pointe. Ils y stockent de nombreux vocabulaires et lexiques de manière à trouver rapidement leur terminologie, sans avoir à quitter leur poste de travail.

Lorsqu'un service linguistique ou un traducteur décide de créer sa propre banque de terminologie, il doit bien analyser deux aspects en particulier: le choix du logiciel et la création de la fiche terminologique.

3 Le choix d'un logiciel

Le choix d'un logiciel n'est pas une mince affaire. Même s'il en existe un plus grand nombre sur le marché qu'il y a dix ans, le choix est difficile. Il faut s'y connaître pour les comparer et... le logiciel parfait n'existe pas.

Il faut trouver le logiciel qui répond aux critères établis: le prix; la

recherche plein texte; la saisie en différé; l'importation des données sur un autre support; et l'élaboration de lexiques.

Dans le prix du logiciel, il faut tenir compte des frais relatifs au soutien technique, à l'achat de nouveau matériel et à la formation.

La recherche plein texte simplifie l'interrogation car elle porte sur toute la fiche. Si le terme qui fait l'objet d'une recherche se trouve dans le champ synonyme et non dans le champ vedette, il n'y a pas lieu de poser deux fois la question, c'est-à-dire une fois dans le champ vedette et une autre fois dans le champ synonyme. Par contre, avec ce type de logiciel, il faut créer des index pour restreindre la recherche lorsque le nombre d'occurrences est élevé. Par exemple, les mots *finances* ou *financiers* apparaîtront sur des centaines de fiches d'une institution financière. Si l'on cherche ces mots dans un champ en particulier, le nombre d'occurrences sera moins élevé. Mais, avant de choisir un logiciel de recherche plein texte, il faut s'assurer que l'ordinateur est assez puissant car les index exigent beaucoup d'espace disque.

La possibilité d'effectuer la saisie de fiches en différé, c'est-à-dire dans un fichier de traitement de texte est une caractéristique qui plaît aux terminologues et aux traducteurs. Au moyen d'une macrocommande, on crée un masque de saisie permettant d'entrer l'information dans un fichier de traitement de texte. La rédaction de fiches peut donc être effectuée par plusieurs personnes et à partir de différents postes de travail. La saisie de données identiques, qui se répètent d'une fiche à l'autre, (p. ex.: source, domaine, auteur) peut être faite automatiquement. En outre, avec la saisie en différé, il est possible de faire circuler les fiches parmi les traducteurs avant de les stocker.

L'importation des données au moyen des fonctions copier/coller est

une caractéristique recherchée, entre autres, lorsqu'il y a de longs contextes à transcrire. La saisie est plus rapide et moins fastidieuse.

Enfin, il est essentiel de pouvoir extraire des données pour élaborer des lexiques facilement, sans devoir effectuer de mise en page complexe.

D'autres critères sont également à prendre en compte au moment de l'achat du logiciel: le soutien technique, le nombre d'années d'existence de l'entreprise, la fréquence des mises à niveau, le nombre d'utilisateurs du logiciel et la convivialité.

Les responsables de services linguistiques n'ont pas toujours leur mot à dire dans le choix du logiciel. Dans certains cas, la direction d'une entreprise achète un logiciel pour d'autres besoins et offre au service linguistique de l'utiliser pour son fichier terminologique. Il faut alors savoir adapter le logiciel. Et souvent, après utilisation, on constate qu'il s'agit d'un bon outil. Le moment venu, il sera toujours possible de transférer les fiches sur un autre support informatique.

4 La création de la fiche terminologique

4.1 L'arbre de domaine

Lorsqu'on prévoit créer un fichier informatisé, la première étape consiste, avant même l'achat du logiciel, à créer un arbre de domaine ou à modifier celui qui existe déjà et qu'il est peut-être temps de mettre à jour. Dans un service linguistique où il n'y a pas de terminologue et où le fichier n'est pas informatisé, il existe rarement un arbre de domaine tel qu'on le conçoit en terminologie. L'arbre de domaine correspond

souvent au système mis en place pour la gestion des dossiers.

Un arbre de domaine d'une petite banque est souvent peu élaboré et il vaut mieux qu'il en soit ainsi, surtout si la gestion de la base de données n'est pas confiée à un spécialiste ou si aucun thésaurus de domaines n'existe. Plus un arbre comprend de branches, plus il est difficile d'attribuer les domaines de façon uniforme et d'extraire les données pour un domaine en particulier.

4.2 La fiche terminologique

L'étape suivante est la création de la fiche en fonction des caractéristiques du logiciel choisi. Par exemple, si le logiciel n'est pas un logiciel de recherche plein texte, il est préférable de restreindre le nombre de champs en regroupant, entre autres, les champs synonyme, variante et abréviation, les champs domaine et sous-domaine ou les champs auteur et réviseur. Les utilisateurs n'aiment pas poser plus d'une fois la même question et ne cherchent pas dans tous les champs.

Si les conventions sont bien expliquées dans un protocole de rédaction et d'interrogation et que chaque utilisateur a son exemplaire et le consulte, il sera facile de bien interpréter les données consignées sur les fiches et de trouver l'information cherchée.

Les fiches terminologiques des fichiers créés récemment comprennent moins de champs que les fiches des grandes banques ou de gros services linguistiques rédigées à l'époque où les ressources humaines et financières étaient importantes et où il y avait des terminologues dans la plupart des services. La source est souvent le numéro de la demande de traduction ou le nom du fichier de traitement de texte. Ce renseignement suffit la plupart du temps parce qu'il

permet de retracer rapidement le texte dans lequel figure le terme et, dans certains cas, de marquer des passages identiques déjà traduits pour les insérer dans le nouveau texte. Lorsque la fiche contient un contexte, c'est plutôt le contexte dans la langue de départ qui y est consigné parce que le contexte dans la langue d'arrivée est souvent le texte traduit et que le traducteur y a accès à partir de son poste de travail.

Sur une fiche de service linguistique, il est souvent important d'ajouter certains champs qu'on ne trouve pas sur les fiches des grandes banques. Un champ client, ou demandeur, permet d'extraire la terminologie propre à un client et de la remettre au traducteur pour assurer l'uniformité de la terminologie dans tous les textes ou au client pour qu'il puisse la valider.

Il est préférable d'entrer les sources dans un champ distinct, d'une part pour simplifier la publication de lexiques destinés aux traducteurs ou aux clients et, d'autre part, pour permettre, éventuellement, l'utilisation d'un logiciel de traduction assisté par ordinateur ou l'insertion de la terminologie dans un texte à traduire.

5 Prospectives

Depuis quelques années, les besoins des services linguistiques ont changé, comme on l'a vu au début du présent exposé (virage informatique, délais de traduction très courts...). Les banques de terminologie, conçues selon les règles établies en terminologie, sont devenues l'outil de base des traducteurs dans les services linguistiques et des traducteurs en pratique privée.

D'autres besoins pointent à l'horizon, notamment la création de fichiers de phraséologie, le dépouillement automatique des

textes, la consultation des dossiers de traduction et l'accès à de nombreuses bases de données par Internet, à partir du poste de travail du traducteur.

Un grand nombre de banques de terminologie existent. Elles constituent une mine de renseignements, qui ont fait l'objet d'une recherche poussée, souvent en collaboration avec des spécialistes, et qui ne sont pas consignés dans les grandes banques. Peu de personnes ont accès à ces renseignements. Dans certains cas, cette information est même perdue à tout jamais lorsqu'un service linguistique disparaît.

Dans un proche avenir, les banques de terminologie utiliseront des données SGML (Standard Generalized Markup Language). L'échange de données dans un format normalisé et universel sera réalité. Il faudrait se pencher dès maintenant sur la question des échanges de données entre services linguistiques. Nous pourrions ainsi assurer l'uniformité de la terminologie, cesser de réinventer constamment la roue et assurer la protection du patrimoine terminologique.

*Marielle Hébert,
Services linguistiques Hébert,
Québec.*

L'inventaire des ressources terminologiques du Rint

L'importance et la variété des besoins terminologiques manifestés dans la francophonie a conduit le Rint à instituer un inventaire permanent des ressources terminologiques du français. D'abord limité aux dictionnaires terminologiques, aux ouvrages paraterminologiques et aux travaux en cours, l'inventaire s'étend maintenant aux banques de terminologie. L'analyse des premiers résultats de ce nouveau volet de l'inventaire démontre qu'à côté des grandes banques de terminologie généralistes, il existe un foisonnement de banques spécialisées de taille moyenne dont la diffusion est parfois très restreinte. La conduite de l'inventaire par le biais d'Internet et la diffusion des résultats seront sans doute favorables à une meilleure diffusion des données terminologiques.

Termes-clés :
dictionnaire terminologique ;
banque de terminologie.

Les besoins terminologiques, qui caractérisent toutes les langues vivantes, varient selon le rythme de développement des connaissances dans tous les domaines. C'est devenu un lieu commun que de parler de l'accélération du développement du savoir, de la multiplication des communications et du volume accru du retard terminologique de certaines langues par rapport à d'autres, par exemple, du français par rapport à l'anglais. Ces besoins terminologiques nouveaux concernent tout autant l'évolution des terminologies existantes que des besoins dénominationnels des champs d'expérience nouveaux qui se créent, notamment par l'interdisciplinarité, évoluent et se diversifient.

Les ressources terminologiques du français

Au cours des années récentes, les projets de veille néologique du Rint (Réseau international de néologie et de terminologie) ont mis en évidence la nécessité de dresser un inventaire des ressources terminologiques du monde francophone afin de planifier adéquatement et efficacement le recensement des terminologies nouvelles. Ce recensement a d'abord porté sur les ouvrages terminologiques.

Inventaire des travaux terminologiques

L'Inventaire des travaux terminologiques est mené par l'Office de la langue française depuis 1975, puis en collaboration avec le Rint après 1986. Le fichier informatisé regroupe plus de 45 000 titres de dictionnaires terminologiques traitant le français, qu'il s'agisse d'ouvrages publiés ou de travaux en cours. Les fiches les plus récentes (environ 30 000) sont diffusées sur disque optique compact avec la Banque de terminologie du Québec. De plus, le Rint a publié plusieurs répertoires dont le dernier, paru en 1995, regroupe plus de 2500 notices présentant les travaux les plus récents ⁽¹⁾. Cet inventaire est alimenté par des enquêtes permanentes menées dans chaque module du Rint et par un certain nombre de correspondants dans d'autres réseaux (par exemple, Realiter, Nordterm, etc.).

Inventaire des banques de terminologie

Cependant, nous assistons à la montée de l'édition électronique et à la multiplication des banques de terminologie. Afin de compléter son travail de recensement des ressources terminologiques, le Rint a mené en 1995-1996 une première enquête afin de dresser un inventaire des banques de terminologie traitant le français.

Cette enquête a été menée à l'aide d'un questionnaire (voir ci-dessous) qui a été conçu de manière à recueillir pour chaque banque recensée des données sur les points

(1) Québec (Gouvernement). Office de la langue française. Direction des travaux linguistiques. *Inventaire des travaux de terminologies récents 1990-1993*, préparé par Marie-Ève Varin, avec la coll. de John Humbley et autres, pour le réseau international de néologie et de terminologie, 3^e édition, Québec, Office de la langue française, 1994, 768 p.

suiuants: les objectifs des différentes banques, leur contenu terminologique, la nature et l'organisation des données qui y sont traitées, le système informatique qui les supporte, les modes d'alimentation et de diffusion.

Chaque banque répertoriée a fait l'objet d'une fiche descriptive la plus complète possible de façon à encourager la concertation et les échanges entre les différentes banques, mais aussi de faire connaître aux usagers potentiels les différents «gisements» terminologiques qu'ils peuvent mettre à profit dans la recherche d'une communication plus efficace. Le Rint a publié en 1996 une première version du répertoire des banques (déjà épuisée). Les membres du Rint estiment que la diffusion de ce répertoire des banques de terminologie traitant le français et d'autres langues sera de nature à favoriser une meilleure concertation dans les travaux terminologiques menés de part et d'autre, ce qui devrait conduire à des économies et à un développement davantage harmonisé du français scientifique et technique en interaction avec d'autres langues. Une diffusion de cet inventaire sur Internet est actuellement envisagée.

Dans cet inventaire, nous entendons par «banque de terminologie» tout système d'information informatisé permettant la consignation, le traitement et la diffusion de données de langues de spécialité et plus particulièrement, des données terminologiques.

Cette définition ne fait pas référence au volume des données stockées. Au contraire, l'inventaire du Rint s'intéresse à toutes les banques publiques ou privées, grandes ou petites, qu'elles fassent l'objet d'une certaine diffusion, même restreinte, ou non.

Même si l'enquête sur les banques n'est pas terminée, en voici quelques résultats.

Nombre de fiches	Banques généralistes	Banques spécialisées
100 000 et +	diffusion publique: 4	diffusion publique: 1
	diffusion restreinte: 5	diffusion restreinte: 1
50 000 à 100 000	diffusion publique: 3	diffusion publique: 6
	diffusion restreinte: 7	diffusion restreinte: 7
5000 à 50 000	diffusion publique: 0	diffusion publique: 2
	diffusion restreinte: 1	diffusion restreinte: 9
1000 à 5000	diffusion publique: 2	diffusion publique: 3
	diffusion restreinte: 1	diffusion restreinte: 8
moins de 1000	diffusion publique: 0	diffusion publique: 1
	diffusion restreinte: 0	diffusion restreinte: 8
Total:	23	46
Grand total:	69	

La première remarque concernant cette enquête est que, contrairement à l'enquête sur les ouvrages terminologiques, nous sommes loin de l'exhaustivité. En effet, si l'on peut être certain d'avoir rejoint les grandes banques de terminologie, celles qui sont davantage connues, il est beaucoup plus difficile de dénicher les petites banques de terminologie surtout lorsqu'elles font l'objet d'une diffusion restreinte ou lorsqu'il s'agit de banques de terminologie internes (entreprises ou organismes). C'est parfois le hasard qui met l'enquêteur sur la piste d'une banque, car l'existence des banques n'est pas toujours officielle ou alors elle peut aussi n'être connue que de quelques personnes dans une entreprise ou un organisme de grande taille. De plus, certaines banques de terminologie font partie du secret industriel: ainsi, certaines entreprises n'ont pas souhaité que l'existence de leur banque soit rendue publique.

Une seconde remarque sur le tableau de l'enquête au sujet de la répartition des banques. Nous assistons, à côté du développement continu des quelques grandes banques (Eurodicautom, BTQ, Termium,

Normaterm, etc.), à la prolifération de nombreuses petites banques qui, pour la plupart, se spécialisent dans un nombre restreint de domaines qu'elles couvrent parfois avec un fort degré d'exhaustivité et d'actualité.

On peut déplorer toutefois le fait que la plupart de ces banques fassent l'objet d'une diffusion privée ou très limitée, alors que, compte tenu des besoins terminologiques nombreux et diversifiés, il serait impératif de multiplier les échanges et de mettre l'ensemble des banques en réseau.

Avec la multiplication des enquêtes de ce genre, à la condition qu'elles soient permanentes, on peut penser qu'à moyen terme, l'essentiel des ressources terminologiques existantes, pour ce qui est du français, sera connu. Un tel inventaire constitue non seulement une source d'information et d'accès aux terminologies, mais également un outil de gestion qui permet de connaître ce qui existe avant de mettre en route des recherches terminologiques, évitant ainsi de créer des chevauchements inutiles et coûteux.

C'est pourquoi le Rint poursuivra ses enquêtes sur les ressources de façon permanente.

L'inventaire permanent des ressources terminologiques se fera par Internet à compter de la fin de 1996, de façon à rejoindre le plus grand nombre de producteurs de terminologie.

*Louis-Jean Rousseau,
Secrétaire général du Rint.*

Inventaire des banques de terminologie

Questionnaire d'enquête

1. Identification de la banque

1.1 Nom de la banque:

.....

1.2 Nom et adresse du producteur (Organisme ou entreprise)

.....

.....

1.3 Objectifs visés par la banque:

.....

1.4 Public cible (catégories):

Interne:

.....

Externe:

.....

1.5 Nombre d'utilisateurs:

2. Contenu de la banque

2.1 Nature des fichiers:

- terminologique
- documentaire
- autres (énumérer)

2.2 Langues traitées:

- français - catalan
- allemand - espagnol
- anglais - italien
- arabe - portugais
- autres (préciser)

2.3. Fichiers terminologiques

2.3.1 Liste des données terminologiques

	donnée obligatoire	donnée facultative
terme (entrée principale)	<input type="checkbox"/>	<input type="checkbox"/>
synonyme(s)	<input type="checkbox"/>	<input type="checkbox"/>
variante(s)	<input type="checkbox"/>	<input type="checkbox"/>
indicatif de langue	<input type="checkbox"/>	<input type="checkbox"/>
indicatif de pays	<input type="checkbox"/>	<input type="checkbox"/>
abréviation	<input type="checkbox"/>	<input type="checkbox"/>
domaine	<input type="checkbox"/>	<input type="checkbox"/>
marque d'usage (ex. terme désuet) indicatif	<input type="checkbox"/>	<input type="checkbox"/>
de grammaire (ex. n.m.; n.f.; adj.)	<input type="checkbox"/>	<input type="checkbox"/>

pondération (ex. terme normalisé, terme rejeté, etc.)	<input type="checkbox"/>	<input type="checkbox"/>
définition	<input type="checkbox"/>	<input type="checkbox"/>
contexte	<input type="checkbox"/>	<input type="checkbox"/>
donnée		
phraséologique	<input type="checkbox"/>	<input type="checkbox"/>
note	<input type="checkbox"/>	<input type="checkbox"/>
illustration	<input type="checkbox"/>	<input type="checkbox"/>
source	<input type="checkbox"/>	<input type="checkbox"/>
rédateur de la fiche	<input type="checkbox"/>	<input type="checkbox"/>
date de la rédaction	<input type="checkbox"/>	<input type="checkbox"/>
autres (énumérer):		

2.3.2 Système de classement:

- Système de domaines préétabli
- Système de domaines «ad hoc»
- Brève description:

.....

.....

2.3.3 Nombre de fiches:

2.3.4 Domaines couverts (principaux):

- Sciences humaines
- Sciences pures
- Sciences appliquées
- Arts
- Autres (préciser)

2.3.6 Modes d'interrogation: Donner la liste des champs interrogeables

.....

.....

2.3.7 Modes d'accès

- Accès direct par lien télématique
- Disque optique compact
- Disquette
- Bande magnétique
- Produits imprimés (dictionnaires, etc.)
- Service SVP
- Autres (énumérer)

.....

2.3.8 Modes d'alimentation

- Volume annuel de fiches nouvelles:
- Travaux terminologiques internes.....
- Échanges de donnée.....
- Acquisition de données externes.....

2.3.9 Mise à jour

Continue

Mensuelle

Autre (préciser)

2.3.10 Abonnement

Mode d'abonnement (conditions):

.....

Tarification:

.....

.....

3. Aspects techniques

3.1 Topologie d'exploitation

• Mode autonome

• Mode serveur

3.2 Configuration matérielle d'exploitation minimale requise

• Type d'ordinateur:

.....

• Processeur:

.....

• Mémoire centrale:

.....

• Espace-disque:

.....

• Interface de communication:

.....

• Système d'exploitation:

.....

3.3 Logiciels d'exploitation des données
(Énumération et brève description)

.....

.....

.....

.....

.....

.....

.....

.....

.....

L'inventaire français des banques de terminologie effectué en 1995 dans le cadre du Réseau international de néologie et de terminologie (Rint)

Cette communication traite de l'inventaire des banques de terminologie qui a été réalisé en France pour le Rint en 1995. L'un des enjeux était de faire un panorama de l'existant pour ce support afin de ne pas dupliquer ce qui était déjà réalisé et de choisir des orientations adéquates. Il a permis de mettre en place une typologie des principaux types d'acteurs, d'évaluer les besoins et de proposer des solutions pour faciliter les échanges de données terminologiques.

Termes-clés :
inventaire; banque de terminologie;
échange de données terminologiques;
diffusion des données.

La première enquête concernant les banques de terminologie remonte en France à 1979, et il ne s'agissait alors que de créer une grande banque française de terminologie, projet qui n'a jamais vu le jour. Depuis cette époque, des banques de données terminologiques se sont créées en France dans différents organismes, particulièrement en entreprise, afin de répondre au plus près aux besoins. Mais le recensement n'en avait jamais été véritablement fait.

L'inventaire réalisé en France pour le Rint (Réseau international de néologie et de terminologie francophone) intervient à la suite de plusieurs enquêtes effectuées dans les années antérieures, notamment dans le cadre du Centre Jacques Amyot, et du Centre de néologie et de terminologie du CNRS. Mais ces enquêtes ne touchaient pas directement la question du recensement des banques de terminologie. Il était, dans le premier cas, question de créer un réseau de terminologie interentreprise et, dans ce cadre, d'appréhender les possibilités offertes à cette idée dans certains secteurs d'activité. Dans le second, il s'agissait d'effectuer un recensement des acteurs de la terminologie en France. Par la suite, les efforts déployés pour essayer de mettre en place un réseau national de terminologie n'ont pas abouti, en raison principalement de l'absence de nécessité ressentie par les principaux partenaires pressentis. Il faut cependant bien considérer que certains secteurs industriels, comme le

secteur de l'aéronautique, travaillaient déjà en collaboration sous l'égide du Gifas (Groupement des Industries Françaises Aéronautiques et Spatiales). Dans le cadre d'un réseau national de terminologie, seules les universités étaient prêtes à mettre des ressources à la disposition des partenaires, sur un réseau téléinformatique de type Renater. Mais, en quelques mois, l'apparition de réseaux aussi puissants qu'Internet a fait éclater les schémas antérieurs sur lesquels s'appuyait cette volonté d'ordre politique de mettre en commun des données terminologiques. Au lieu de réseaux d'échanges de données effectués dans un cadre relativement contraignant et limité, il est possible aujourd'hui de parier sur des réseaux d'échanges très larges et largement ouverts, dont l'absence de contrainte et de limite fait toute l'efficacité.

Pour pallier les difficultés rencontrées et répondre aux besoins exprimés depuis de longues années, il a paru utile de commencer par faire un inventaire de l'existant en matière de ressources terminologiques. Ce dernier est au moins de deux sortes : travaux terminologiques en cours, et supports de données terminologiques existants (support papier, disque optique compact et bases de données). Un grand travail a été fait depuis plusieurs années par le Rint pour répertorier ces éléments (particulièrement dans l'inventaire permanent des travaux de terminologie récents).

C'est dans le prolongement de cette réflexion que le Rint a lancé un inventaire des banques de données terminologiques, chacun des modules

du Rint ayant en charge la conduite de cet inventaire pour sa zone géographique.

L'inventaire effectué en France, intitulé *Inventaire des banques de terminologie*, a été réalisé par M. Christophe Riondet, dans le cadre d'un mémoire de maîtrise en anglais – Recherche et Informatique à l'Université Paris IV – Sorbonne, sous la conduite de M. Loïc Depecker. Il a été fait grâce à l'apport du Crettal (Centre d'études et de recherches en terminologie, terminologie et aménagement des langues) situé à l'Université Paris III – Sorbonne nouvelle, et à celui du Centre de terminologie et de néologie du CNRS (Université Paris XIII). Cette enquête s'inscrit dans le développement des activités du Rint, qui s'est notamment donné pour mission de recueillir les informations disponibles en matière de terminologie, notamment en effectuant des inventaires de documents existants (*Inventaires des travaux de terminologie récents* publiés depuis 1989).

L'un des enjeux de l'inventaire des banques de terminologie effectué en 1995 est, précisément, de faire pour ce support, un panorama de l'existant. La méthode de l'inventaire est indispensable dans les domaines de recherche, et particulièrement dans celui-ci afin de bénéficier des travaux déjà existants, de ne pas dupliquer ce qui est déjà réalisé, et de choisir des orientations adéquates. La méthode poursuivie a été celle d'un questionnaire envoyé par courrier, suivi d'appels téléphoniques. Dans la mesure du possible, ce questionnaire servait également de support à une entrevue avec les participants.

Les différentes voies qui s'ouvrent à partir de là sont notamment :

- De dresser un inventaire permanent des banques de terminologie des secteurs public et privé ;
- De faire se correspondre offre et demande de terminologies ;

- D'encourager l'harmonisation des terminologies de secteurs homogènes ;
- De travailler à l'amélioration des formats d'échanges afin de faciliter les transferts de données.

Le présent inventaire a intégré des informations sur la France, en l'étendant aux institutions européennes et aux organisations internationales en Europe.

On s'attardera, dans les lignes qui suivent, sur les réactions qui ont accompagné cette enquête, particulièrement sur les plans :

- Des besoins exprimés ;
- Des attitudes ;
- Des solutions susceptibles d'être apportées.

Cette réflexion pourrait permettre de tracer les voies visant à décloisonner les terminologies utiles aux spécialistes.

Les types de «banques de données terminologiques»

La notion même de banque de terminologie n'est pas entièrement claire. L'éventail de ce qui se rencontre en ce domaine est vaste : simple fichier sur traitement de texte, logiciel de terminologie, logiciel documentaire, logiciel d'aide à la traduction, logiciel maison ou banque de données structurée, développée sous SGBD. Des uns aux autres existe tout un ensemble de réalisations : création de produits multimédias (Matra Marconi Space, ENAC), serveur de nomenclatures (INSEE), etc.

L'inventaire réalisé pour la France comporte donc aussi bien des fichiers terminologiques «embryonnaires», que des «produits terminologiques» développés en interne et utilisant des supports multimédias ou l'intelligence artificielle, ou encore des banques de données proprement dites.

Il faut considérer cependant que cet inventaire ne saurait encore être exhaustif et entièrement précis, en raison tout d'abord des nécessités de confidentialité.

Réactions des interrogés

L'une des premières questions prévisibles est en effet celle de la confidentialité des données. La terminologie est le reflet du savoir-faire d'un organisme et oriente facilement, dans ses contenus néologiques, vers les recherches que développe celui-ci. Au point que parfois la confidentialité ne porte pas seulement sur les données elles-mêmes, mais sur le fait même de dévoiler l'existence d'une banque de données terminologiques. Sur ce point, plusieurs entreprises de certains secteurs ont refusé de coopérer à cet inventaire, comme le secteur de la construction aéronautique.

Mais là encore, il convient de nuancer : dans le domaine de l'aérien, les données restent confidentielles pour les constructeurs, mais non forcément chez les sociétés de service aérien telles qu'Air France, qui est plus disposée à une coopération avec d'autres partenaires.

Les nécessités de confidentialité ne rendent donc pas *a priori* caduque toute démarche d'inventaire. D'autant que celle-ci est souvent nouvelle pour les personnes interrogées. Au point que, par exemple, la personne contactée au laboratoire de Sanofi Recherche, qui traite de terminologies pharmaceutiques, était vivement intéressée par la démarche d'inventaire des banques de terminologie, car elle lui donnait l'occasion de sortir d'un relatif isolement. Il en va de même pour la chaîne de télévision franco-allemande ARTE, qui souhaite développer ses échanges de données terminologiques

avec des entreprises du secteur audiovisuel à travers l'Europe.

Cet ensemble d'expériences recueillies permet de mettre en place une typologie des acteurs, des attitudes et des solutions à proposer pour faciliter les échanges de données terminologiques.

Types d'acteurs

Parmi les principaux groupes d'acteurs, il convient de distinguer :

- Les entreprises publiques (ministères, institutions, organisations) qui constituent des terminologies ayant vocation à être diffusées;
- Les entreprises privées qui traitent des terminologies propres à leur domaine de spécialité, en particulier la terminologie « maison »;
- Les entreprises de traduction et d'ingénierie linguistique qui disposent de terminologies destinées, la plupart du temps, à des sociétés clientes.

Les attitudes en matière d'échanges terminologiques diffèrent évidemment en fonction du type d'acteur :

- Dans le secteur public, les échanges terminologiques sont ponctuels et se limitent aux organismes publics nationaux et internationaux. Néanmoins, on notera le souhait de certains organismes d'étendre leurs échanges au secteur privé (tel l'Onu à Genève). En France, dans le domaine de la finance, un projet de coopération terminologique a vu le jour entre la Banque de France et diverses banques privées;
- Dans le secteur privé, les échanges terminologiques sont rares en raison principalement du problème de la confidentialité des données. La question des droits d'auteur constitue également un frein à la coopération. Pourtant, chez la plupart des entreprises contactées, ce projet a suscité un vif intérêt, car il

permettrait, entre autres, de réduire les coûts liés à la terminologie. Ce constat est particulièrement saisissant dans les domaines pharmaceutique (ex. : Sanofi, Pasteur Mérieux Sérums et Vaccins) et financier (ex. : Axone Services et Développement).

Besoins et solutions possibles

En conséquence, les principaux besoins exprimés par les participants s'orientent autour des questions suivantes :

- L'acquisition de terminologies spécialisées multilingues;
- L'information sur les terminologies existantes, les formats de données et les logiciels utilisés, la question des droits d'auteur.

Au vu de ces problèmes, on peut envisager les solutions suivantes :

- La diffusion d'un répertoire des fichiers et banques terminologiques des entreprises qui ont participé au programme;
- La coordination des échanges assuré par un organisme qui garantirait la protection juridique et la confidentialité des terminologies;
- L'adoption d'un format d'échange de données terminologiques compatible;
- L'organisation de réunions régulières entre participants.

Conclusion

L'expérience que constitue cet inventaire fournit un champ d'expériences riches de perspectives.

Parmi les différentes composantes évoquées plus haut, les données techniques sont capitales en vue d'un échange facile et à moindre coût. C'est là que l'élaboration de formats de données terminologiques est important.

L'inventaire étend de fait la réflexion vers la question des formats d'échange de données terminologiques : car il n'y a pas que la décision d'ordre politique, de vouloir ou de ne pas vouloir échanger des données, qui entre en ligne de compte. La technique qui doit être mise en œuvre pour échanger des données pose un certain nombre de problèmes.

Ainsi, pour faciliter les échanges de données terminologiques entre concepteurs de banques de terminologie, l'utilisation d'un format de transfert, ou d'échange adéquat, reste plus que jamais nécessaire. Cela passe par l'adoption d'une norme d'échange de type SGML ou HTML, telle qu'elle est actuellement en cours de fixation au sein du groupe ISO TC 37, conduit pour la France par l'Afnor et l'Université de Rennes 2.

Par ailleurs, en ce qui concerne la diffusion des données, il sera utile que l'ensemble des données terminologiques soit interrogeable à distance. Pour cela, l'organisme ayant en charge la collecte et gestion des terminologies devrait disposer d'une « base de cumul » mis en accès sur un serveur. D'autres supports de diffusion comme les disques optiques compacts (DOC) complèteraient cette diffusion.

*Loïc Depecker,
Délégation générale à la langue
française,
France.
Christophe Riondet,
Université de Paris IV,
France.*

L'utilisation d'Internet dans la commercialisation des produits de l'industrie de la langue

Cet article vise à faire connaître le type de données linguistiques accessibles sur Internet. En outre, il propose des outils conçus par l'industrie de la langue qui peuvent servir aux langagiers. Pour aiguiller les recherches sur le mode de diffusion des banques de terminologie, l'article fait état des avantages et des désavantages de la commercialisation sur Internet tels que perçus par Machina Sapiens. En effet, l'entreprise, qui se spécialise aussi en terminotique, utilise Internet pour diffuser son produit, un outil de recherche incluant un extracteur de mots-clés.

Termes-clés :
Internet; terminotique; extracteur de mots-clés; banque de terminologie; outil de recherche; industrie de la langue.

Machina Sapiens est une entreprise de haute technologie qui mise sur l'innovation, l'excellence de ses produits et l'expertise de son personnel. L'entreprise a commercialisé une dizaine de logiciels, dont des logiciels de formation, des outils de programmation par objets et des outils de traitement de la langue dont le Correcteur 101. Machina Sapiens a remporté le Prix du « Meilleur concept logiciel » en 1986, le prix Mica pour la recherche en 1988, le Prix de la ministre de l'Enseignement supérieur et de la Science en 1991, l'Octas 93 de la technologie et le Prix du français en informatique 1994, décerné par l'Office de la langue française.

Dans le cadre d'une démarche terminologique, il est important de tenir compte de l'inexistence de normes sur Internet. En effet, Internet ne prescrit aucun processus d'édition. Il est donc difficile pour le terminologue de connaître l'identité de la source et de s'assurer de l'exactitude des données. Le terminologue doit être critique face à la qualité de la source.

À titre d'exemple, Machina Sapiens a trouvé les produits suivants :

- Corpus de textes en langue de spécialité;
- Listes de mots;
- Listes de mots fréquents;
- Dictionnaires de synonymes;
- Traducteurs bilingues (accès informatique à un dictionnaire bilingue)

Shake and bake demo FTP.
Directory: members-only/tools/
translation/shake and bake/more
info: info/0098.

De plus, certains groupes mettent en commun leurs données linguistiques et les offrent à différents prix sur Internet. C'est le cas du Consortium for Lexical Research (CLR) lexical@crl.nmsu.edu. La liste suivante est un ensemble de sites, de répertoires et de fichiers à consulter sur Internet.

1. Information linguistique sur Internet

Internet regorge de données linguistiques. Pour le terminologue averti, il est possible de trouver toute l'information nécessaire pour bâtir une fiche terminologique. Bon nombre de monographies portant sur des domaines de spécialité peuvent être consultées sur Internet.

L'accès à ces données est parfois gratuit, parfois très coûteux, selon le site qui l'offre. En outre, l'information offerte varie rapidement. Par exemple, ce qui était disponible gratuitement hier, ne l'est plus aujourd'hui, ou tout simplement n'existe plus sur Internet.

Sites FTP

Obtenus du serveur Archie (telnetarchie.rutgers.edu, loginarchie, commandes «find english», «find grammar», etc.)

Site	Répertoire(s) et fichier(s)
gum.isi.edu	/share/pub/english-corpus-1
ftp.isi.edu	/pub/english-corpus-1
ra.msstate.edu	/pub/doc/words-1/Funnies pronouncing.english
ftp.cc.gatech.edu	/pub/coc/graduate
-	info/INTERNATIONAL/intensive.english
epas.utoronto.ca	/pub/cch/italian
	/pub/cch/english
	/pub/cch/english/dictionaries/roget13a.zip
	/pub/cch/italian/wordlists/words.italian.Z
ftp.iro.umontreal.ca	/lude-iro/ispell3.1.08/src/orig/languages/english
ftp.cc.utexas.edu	/pub/english
ftp.tex.ac.uk	/pub/archive/language/english
	/pub/archive/info/spanish-tutorial
	/pub/archive/language/italian
ftp.cac.psu.edu	/pub/courses/english
kekule.osc.edu	/pub/central-eastern
-	europe/russian/windows/cyrwin/spanish.kbd
qiclab.scn.rain.com	/pub/wordlists/italian
	/pub/wordlists/spanish
	/pub/wordlists/german
	/pub/wordlists/dictionaries
black.ox.ac.uk	/wordlists/italian
	/wordlists/spanish
	/wordlists/dictionaries/Roget.words.Z
ftp.cs.tulane.edu	/pub/rpg/new-words.spanish
cs.nyu.edu	/pub/local/grishman/spanish-proteus-tar.t
	/pub/nlp/dictionaries/spanish-uni.dict
	/pub/nlp/grammar/comm-spanish.gram
	/pub/nlp/grammar/spanish.gram
vax.cs.pitt.edu	/pub/editors/ue310e/SPANISH.H
ftp.cs.yale.edu	/pub/lojban/etymology/spanish.voc
achilles.doc.ic.ac.uk	/tex/aston/languages/babel/italian.doc
	/tex/aston/languages/babel/spanish.doc
freebsd.cdrom.com	/.5/gutenberg/freenet/german-surrender
midway.uchicago.edu	/pub/tex/macintosh/utilities/excalibur/italian-dictionary-s... x
	aussi NewGermanDictionary-...
sol.cs.bucknell.edu	/pub/mac/Excalibur-dictionaries/NewGermanDictionary (etc.)
relay.cs.toronto.edu	/doc/Dictionaries/words.italian.Z
oes.orst.edu	/pub/data/etext/roget.p2
sonata.cc.purdue.edu	/pub/next
	/Literature/Gutenberg/etext91/roget11...
sumex-aim.stanford.edu	/info-mac/info/nms/roget-1911-thesaurus...
mrcnext.cso.uiuc.edu	/pub/info-mac/info/nms/roget-1911-thesauru...
	/pub/etext/etext91/roget13.txt
ftp.spies.com	/Library/Classic/roget.txt
ftp.apple.com	/alug/hypercard/roget.1.cpt.hqx
sun.soe.clarkson.edu	/pub/src/getgrammar
	/pub/src/dictionary
cmns.think.com	/archives/learning/grammar
ics.uci.edu	/pub/machine-learning-
databases	/prodigy/domains/grammar
ftp.latrobe.edu.au	/pub/celia/english/grammar

2 L'application de la terminotique dans les outils de recherche sur Internet.

Certains logiciels de recherche en vente sur Internet offrent des outils qui résultent de la terminotique.

2.1 InfoScan, un outil de recherche conçu par Machina Sapiens inc.

À quoi sert InfoScan? InfoScan est un assistant personnel qui lit, filtre et classe les messages reçus d'Internet. Il aide les utilisateurs qui sont submergés par les nouvelles (*News*) et le courrier électronique. Il repère pour eux les documents les plus susceptibles de les intéresser. InfoScan donne immédiatement accès aux documents les plus pertinents; l'utilisateur n'est donc plus obligé de lire tous les documents, ni leurs titres souvent peu évocateurs, ni même un seul mot de leurs textes. Les documents intéressants sont visibles d'un coup d'œil, ce qui évite des explorations fastidieuses et des lectures inutiles.

Comment fonctionne InfoScan? Pour décrire ses intérêts à InfoScan, l'utilisateur doit simplement taper des mots-clés dans des filtres. La particularité des filtres d'InfoScan est qu'ils permettent d'accorder une pondération aux catégories de mots-clés. Par conséquent, non seulement l'utilisateur choisit le domaine de recherche, mais il décide aussi des relations entre les termes par rapport au domaine de sa recherche. De cette façon, InfoScan fait des recherches plus pointues avec plus d'exactitude.

D'abord, InfoScan cherche les mots-clés dans l'ensemble des

documents que lui indique l'utilisateur. Puis, InfoScan affiche les résultats sur un écran de radar qui permet de consulter les documents les plus pertinents. Le terminologue peut ainsi fouiller plus à fond les sources terminologiques qu'il possède sur support informatique. Notamment, InfoScan est doté d'outils de tri, de recherche textuelle et de lecture de documents.

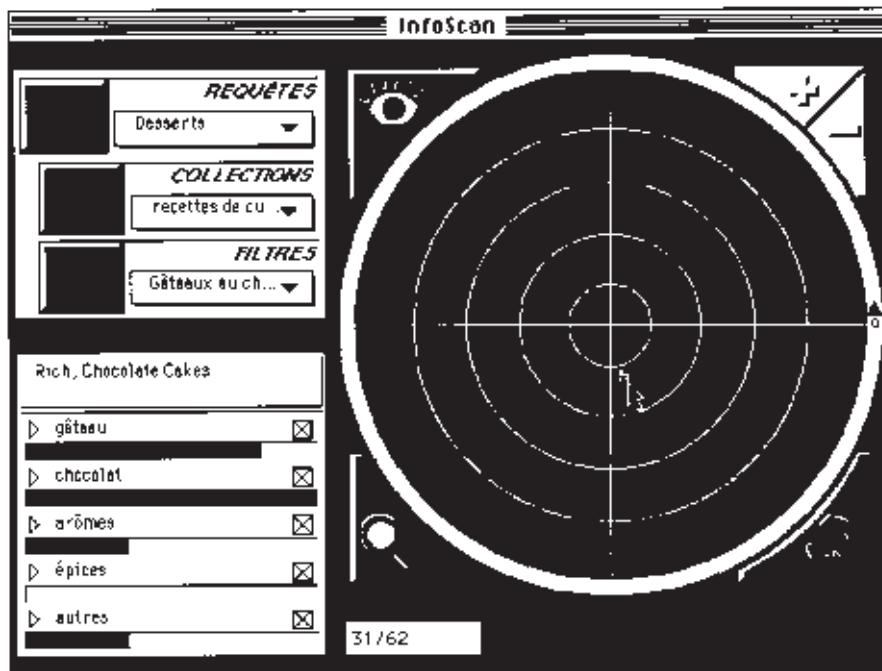
À quoi ressemble InfoScan?

L'interface d'InfoScan se divise en trois parties. D'abord, il est composé principalement d'un outil de visualisation graphique, inspiré de la métaphore du radar, qui permet de repérer rapidement les documents les plus pertinents, sans même lire un seul mot.

Puis, il contient une zone du profil de l'utilisateur qui permet de définir et de sélectionner des filtres, des sources d'information ou collections et de définir des requêtes. Une requête est une association entre un filtre (termes d'un domaine) et une collection de documents (sources d'information). Enfin, son écran principal est aussi doté d'une zone de retour d'information qui contient le titre du document courant ainsi qu'un aperçu de son contenu.

Où peut-on se procurer InfoScan? La version 1.0 d'InfoScan est offerte sur le site W3 de Machina Sapiens : <http://www.MachinaSapiens.qc.ca>. Après avoir téléchargé le logiciel, on peut l'utiliser gratuitement en version de démonstration. Toutefois, toutes les fonctions de recherche du logiciel ne sont pas offertes en mode de démonstration. InfoScan deviendra pleinement fonctionnel une fois que l'utilisateur sera enregistré et qu'il aura payé la somme requise.

Qui a contribué à la réalisation d'InfoScan? Le consortium Canarie (Canadian Network for the Advancement of Research, Industry and Education) a participé financièrement à la réalisation



L'écran de visualisation d'InfoScan après une requête. Les meilleurs articles sont situés au centre du radar.

d'InfoScan. Le CEFRIO (Centre francophone de recherche en informatisation des organisations) a contribué au transfert d'expertise d'une équipe de recherche universitaire.

L'outil d'extraction de mots-clés

InfoScan offrira dans une version ultérieure un extracteur de mots-clés trilingue. Cet outil cherche dans les textes les mots-clés et les inclut automatiquement dans les filtres de recherche. L'outil d'extraction de mots-clés permettra aux utilisateurs de remplir automatiquement leurs filtres de recherche. Parfois, l'utilisateur ne connaît pas suffisamment les domaines des recherches qu'il désire effectuer sur Internet. Afin d'entrer correctement les mots-clés qu'il recherche dans les filtres d'InfoScan, il peut utiliser l'extracteur de mots-clés. Pour ce faire, il doit fournir à InfoScan un texte portant sur le domaine de

recherche en question. À l'aide de ce texte, l'outil d'extraction de mots-clés repérera automatiquement les mots-clés qui s'y trouvent et les inclura dans le filtre de recherche. Le nouveau filtre ainsi créé sera plus représentatif du domaine et servira à trouver de l'information pertinente qui intéresse vraiment l'utilisateur. Les textes repérés par *InfoScan* à l'aide de ce filtre traiteront du même domaine de recherche que le texte de départ, c'est-à-dire qu'ils contiendront les mêmes mots-clés.

Quant à la façon d'entrer les mots-clés dans le filtre de recherche, l'extracteur de mots-clés connaît leur valeur conceptuelle par rapport au texte et leur accorde automatiquement une pondération. Cet outil est capable d'extraire des termes simples et complexes à partir d'écrits français, anglais et espagnols.

2.2 AnchorPage

D'autres entreprises aux États-Unis telles qu'Iconovex ont choisi d'offrir aux utilisateurs d'Internet un extracteur de concepts pour faire de l'indexation. Cet outil repère dans les textes les mots représentant des concepts et en fait des index de recherche.

3 Mode de commercialisation sur Internet

3.1 Les avantages de la commercialisation des produits de l'industrie de la langue sur Internet :

- Accès international instantané à une clientèle de millions d'internautes;
- Rapidité de distribution (quasi instantanée);
- Élimination des frais de transport, d'emballage, de support magnétique et d'impression de manuels;
- Réduction des frais de soutien technique, meilleure qualité du service;
- Cycles de production et de diffusion raccourcis;
- Bêta testeurs facilement accessibles;
- Publicité à faible coût directement sur Internet;
- Possibilité d'automatisation poussée des transactions;
- Prix plus bas pour les utilisateurs.

3.2 Les problèmes auxquels est confrontée l'entreprise Machina Sapiens dans l'établissement d'une procédure de vente sur Internet :

- Manque de sécurité des transactions, il n'existe toujours pas

de normes de sécurité pour échanger des renseignements confidentiels;

- Manque de confiance de la part des clients et des institutions financières;
- Clientèle de Internet habituée aux logiciels gratuits;
- Besoin de normes communes pour favoriser le commerce sur Internet entre différents pays;
- Bruit sur Internet (l'information est plus disponible, par contre elle est moins facile à trouver que dans une bibliothèque par exemple);
- Diffusion de l'information publicitaire difficile (cette démarche requiert un certain tact, que l'on appelle dans le jargon anglophone de Internet, le *Netiquette*).

4 Mode d'acquisition d'InfoScan

Modalités d'acquisition du logiciel

Une version de démonstration limitée au traitement de vingt documents est offerte gratuitement sur Internet. Pour rendre cette version complètement fonctionnelle, le client doit :

- Remplir un formulaire d'enregistrement qui se trouve sur le site. Le client recevra dans les vingt-quatre heures un numéro de commande personnel calculé par InfoScan à partir de son adresse Internet et de la configuration de son ordinateur;
- Commander le produit par télécopieur, par téléphone ou par la poste en indiquant le numéro de commande reçu. Une fois le paiement validé, un jeton est envoyé au client pour rendre la version de démonstration pleinement fonctionnelle. Le jeton est libellé au nom du client et de son adresse électronique. C'est là un moyen de protection de logiciel davantage dissuasif, car les coordonnées du

client apparaissent sur chaque copie piratée.

À l'avenir, si les normes de sécurité sur Internet entrent en vigueur, il sera possible d'instaurer un service commercial sécuritaire sur le réseau Internet. Machina Sapiens fera alors appel à une technologie adoptée par Netscape et ses partenaires dont Bank of America et MasterCard.

5 Protocole standard de communication entre une banque de terminologie et des programmes informatiques (pour la diffusion des données sur disque optique compact)

Une entreprise privée, ou même un gouvernement, peut permettre aux fabricants de logiciels d'utiliser sa banque de terminologie moyennant le paiement d'une licence par exemplaire du logiciel vendu. Pour donner accès à des banques de terminologie, il faut établir un protocole de communication standard entre la banque et l'interface des divers programmes d'exploitation. Ce procédé permet à différents diffuseurs de concevoir des applications variées qui utilisent le contenu de la banque de terminologie. Ce procédé permet aussi une utilisation diversifiée des données terminologiques tout en protégeant le format et le contenu des fiches terminologiques sous format crypté. Dans le domaine commercial, deux types de consultations pourraient coexister. D'une part, la banque de terminologie en entier serait disponible sur un disque optique compact.

Quant aux vocabulaires spécialisés et à l'utilisation restreinte des données qui apparaissent sur la fiche, ils pourraient être accessibles sur Internet aux utilisateurs qui possèdent un mot de passe. On peut

obtenir ce dernier soit par abonnement ou moyennant des frais de consultation.

6 Question ouverte: L'accès aux banques de terminologie doit-il être gratuit?

Il serait présomptueux de répondre trop rapidement et sans équivoque à cette question. Toutefois, les renseignements suivants porteront sans doute à réflexion. D'abord, il faut constater que toute information gratuite qui se trouve sur Internet peut servir au développement de produits commerciaux de tous genres, et ce, partout dans le monde. Les industries de la langue sont friandes des données linguistiques qui alimentent leurs produits. Pour éviter que les grandes entreprises qui profitent de budgets colossaux ne viennent rivaliser avec l'entreprise de la région, ou les services gouvernementaux qui ont eux-mêmes mis sur pied la banque de terminologie, deux solutions sont proposées:

- Réclamer des frais d'accès;
- Surveiller l'identité de celui qui télécharge les données terminologiques (ce qui est difficile à effectuer de façon certaine).

*Nathalie Côté,
Les Logiciels Machina Sapiens,
Québec.*

M. Louis-Jean Rousseau signale aux participants que l'inventaire mené par le Rint a permis de décrire environ 70 banques de terminologie, tant en Europe qu'au Québec et au Canada. Il constate que peu d'entre elles sont diffusées publiquement. On ne semble se rendre compte ni de la valeur des informations terminologiques ni des possibilités de diffusion: il y a une demande, mais l'offre ne suit pas. Selon M^{me} Marielle Hébert, deux raisons expliquent ce phénomène. La première, c'est le travail en vase clos dans l'entreprise, la terminologie maison, la confidentialité des données et la concurrence. La seconde raison tient au fait que la terminologie en entreprise est faite à des fins internes liées aux besoins de traduction et de communication. On n'a ni les moyens ni le temps d'organiser la diffusion de la terminologie. Pour ce qui est des cabinets de pigistes, il faut noter que la terminologie produite par le traducteur ou le terminologue appartient au client qui pourrait ne pas souhaiter sa diffusion.

M. Jean-Guy Meunier fait remarquer que, tout comme la fiche bibliographique, la fiche terminologique actuelle est fondée sur la description linéaire, les champs, etc. La série de communications de ce matin nous

invitent à réfléchir profondément à cette question. Les banques de données terminologiques ne disparaîtront pas, mais le format même conceptuel, et non le support matériel, va commencer à se modifier fort probablement en une fiche terminologique virtuelle comme l'accès aux programmes Java qui peuvent, sur des bases de données textuelles, aller chercher des bases d'information spécifiques. On va travailler davantage, un peu comme on a illustré ce matin dans InfoScan, avec des programmes *ad hoc*, des logiciels de pointe qui vont aller chercher l'information qu'on retrouvait normalement dans la fiche terminologique. Je pense qu'il y a une réflexion à faire sur la transformation même du concept de fiches terminologiques en raison du nouveau support matériel qui est en train de se créer et des modes de diffusion qui sont en jeu, mais aussi de la structure conceptuelle du mode d'accès à cette information.

M. Chérif Mbodj intervient pour souligner l'exposé de M. Marcel Diki-Kidiri sur l'implantation des banques de terminologie en Afrique. Il a parlé d'équipement et de formation, évidemment ce qu'on entend souvent dans les tribunes c'est que la francophonie est constituée de

l'ensemble des communautés qui partagent la langue française. Il faut savoir ce que l'on partage réellement en ce qui concerne le français, car dans les pays du Sud, ceux qui parlent réellement le français, quand on est tout à fait optimiste, c'est 20% de la population. Mais il y a 80% de la population qui ne parle pas le français du tout et le développement des langues partenaires permet à ce 80% de la population du Sud de se développer. Il ne s'agit pas de développer seulement 20% de cette population. Or le français, justement, a un rôle tout à fait positif qui est de permettre, en développant les langues partenaires, de développer souvent ce 80%. Il y a donc là un problème qu'il faut poser réellement: c'est de savoir si en réalité on veut ou non aider le Sud à se développer.

Le secrétaire général du Rint, M. Louis-Jean Rousseau, souhaite que l'on traite très sérieusement de cette question dans les instances de la francophonie. Il informe les participants qu'au cours de la dernière réunion du Comité d'orientation du Rint, le réseau a amorcé un virage important qui s'était préparé en 1994-1995 par le soutien, de plus en plus grand, aux travaux terminologiques menés par des équipes africaines de terminologie dont plusieurs mènent leurs

Comptes rendus des débats

recherches sur des vocabulaires en différentes langues transnationales. C'est un très modeste effort et il faut aller beaucoup plus loin effectivement pour assurer un transfert des technologies terminotiques et faire en sorte que les banques de terminologie puissent trouver leurs fonctions en d'autres lieux et d'autres langues, et bien entendu je parle ici des pays du Sud. L'ébauche d'étude que nous a présentée M. Marcel Diki-Kidiri est intéressante dans cette perspective. On peut constater que cette année, au sein du Rint, la réflexion sur les banques de terminologie prend forme et, à travers cette réflexion et d'autres actions, on pourra aller beaucoup plus loin. Le Rint devrait sensibiliser l'Agence de coopération culturelle et technique et d'autres instances de la francophonie aux besoins des pays du Sud en matière d'équipement informatique pour les travaux de développement terminologique des langues nationales et transnationales. Ce sont des réflexions qui ont été faites à l'occasion d'un stage organisé par le Rint à l'École internationale de Bordeaux tenu en octobre dernier. Il s'agissait d'un stage d'initiation à certains logiciels terminotiques, et la grande préoccupation des participants était la suivante : pourrions-nous mettre en pratique dans nos pays ce que nous avons appris ici, compte tenu des lacunes de l'équipement ? Cette question est d'autant plus épineuse que les logiciels terminotiques exigent des ordinateurs de plus en plus puissants. Il serait nécessaire de créer, dans les instances de la francophonie, un fonds de dotation en équipement et en logiciels pour les équipes de terminologues des pays du Sud où se font des travaux d'aménagement terminologique. Tout cela doit faire l'objet d'évaluation. M. Marcel Diki-Kidiri proposait ce matin un certain nombre de critères pour le choix des sites où l'on pourrait implanter des banques de terminologie ou des bases

de données terminologiques, en privilégiant la présence de compétences en matière de terminologie et la présence de projets d'aménagement terminologique. Il reste à souhaiter que ces quelques idées se transforment en action.

Prospectives et avenues de collaboration

Présidente de séance :
Tina Célestin
(Office de la langue française, Québec)

Thème 4

La Banque internationale d'information sur les États francophones (BIEF)

La Banque internationale d'information sur les États francophones, aussi reconnue sous le sigle BIEF, est depuis décembre 1991, un programme décentralisé de l'Agence de coopération culturelle et technique qui œuvre dans les domaines de la bibliothéconomie, des archives et des sciences de l'information.

Termes-clés :
BIEF ; Francophonie ; inforoute ;
communication ; information.

C'est en 1986 lorsque les chefs d'États des pays francophones se rencontrent pour la première fois à Versailles afin d'établir les grandes stratégies de la coopération francophone qu'ils déterminent l'information scientifique et le développement technologique comme l'une des cinq priorités. À cette époque, on demande de dresser un inventaire des informations bibliographiques informatisées concernant les États francophones et la restitution de ces informations aux États membres. Le Gouvernement du Canada accepte de diriger cette opération et la confie à des spécialistes de l'information. C'est ainsi qu'est née la BIEF.

Ce catalogue collectif multilatéral de la Francophonie existe toujours d'ailleurs. Anciennement reconnu comme le volume 2 de la collection « CDThèque Francophonie », il a fait peau neuve et vient tout juste de sortir (depuis quelques semaines) sous un nouveau nom *Ensemble*. Il est disponible sur disque compact.

Au Sommet de Québec en 1987, l'ACCT demande la reconduction du programme de la BIEF et l'élargissement de son mandat :

– Par le renforcement des institutions documentaires nationales,

c'est-à-dire les archives et les bibliothèques nationales, les centres nationaux de documentation et les bibliothèques publiques et scolaires ;

– Par la structuration, modernisation et consolidation des politiques et systèmes nationaux d'information ;

– Par la formation et le perfectionnement des professionnels dans le domaine ;

– Par la conservation et la préservation du patrimoine documentaire ;

– Par la démocratisation de l'accès à l'information.

Présentement la BIEF compte un réseau de 47 centres serveurs, donc un point de service dans tous les pays membres de la Francophonie. La fonction « centre serveur » est installée auprès des institutions documentaires nationales telles que les bibliothèques nationales, les archives nationales et les centres nationaux de documentation. Ce sont eux qui sont les points focaux de la BIEF dans ces pays. Ils représentent la BIEF au sein de leur pays et contribuent au renforcement du système d'information nationale. De son côté, le Gouvernement du pays met à la disposition de la BIEF un spécialiste dans le domaine de la bibliothéconomie ou des archives. Je reviendrai plus tard sur l'importance des centres serveurs en matière d'inforoute.

Comme je vous le mentionnais un peu plus tôt, la BIEF est un programme de l'ACCT, organisation intergouvernementale de la Francophonie. L'ACCT a des buts et un rôle à jouer en matière d'information et de communication. La BIEF est là pour répondre à ces préoccupations. C'est la raison pour laquelle je vous exposerai les grands objectifs de la Francophonie plutôt que vous énumérer les actions pointues de la BIEF.

En 1993, au Sommet de Maurice, les chefs d'États déterminent une prospective construite en quatre espaces :

- Un espace de savoir et de progrès : enseigner le français, langue seconde, créer un cadre de scolarisation, d'éducation de base et de formation professionnelle ;
- Un espace de culture et de communication : prendre en compte les nouveaux défis technologiques créés par les inforoutes ;
- Un espace de liberté et de démocratie : construction de l'état de droit, l'enseignement de la démocratie et des droits de la personne, la liberté de presse et la formation des journalistes, la création de base de données juridiques ;
- Un espace Francophonie dans le monde : mailler en réseau la Francophonie, consolider l'existence et l'usage du français en tant que langue de travail à l'Onu et dans les grandes instances et conférences internationales.

Avec l'arrivée des nouveaux défis créés par les inforoutes, la Francophonie a renforcé ces actions en matière de communication et d'information. La BIEF, TV5 et TV5 Afrique, l'UREF (Université des réseaux d'expression française) ; le Forum francophone des télécommunications et des technologies de l'information, etc., ont donc pris encore plus d'ampleur.

Lors du Sommet de Cotonou en décembre dernier, les chefs d'États

s'entendaient en matière d'information pour que des efforts soient déployés en priorité sur les actions suivantes :

- La mise en place de serveurs du Sud ;
- La formation des ressources humaines au développement et à l'utilisation des serveurs et des sites ;
- La normalisation des protocoles de communication ;
- La mise en place de vitrines pour les pays du Sud ;
- Le développement de contenus en français.

À ce Sommet, la programmation de l'ACCT, pour le *biennium* 1996-97, a été présentée sous les mêmes quatre axes d'intervention qui avaient été déterminés lors du Sommet de Maurice en 1993.

C'est sous la priorité *Espace de culture et de communication*, que l'on retrouve l'alliance entre les télécommunications, l'informatique et l'audiovisuel, dans le cadre de ce qui est communément appelé « nouvelles technologies de l'information et de la communication ».

L'ACCT s'est donc fixé comme objectif premier de promouvoir, de produire et d'enrichir d'une façon significative les contenus de langue française.

À l'heure actuelle, l'information qui existe et circule sur des réseaux comme Internet est majoritairement en langue anglaise. Par contre, il en existe beaucoup plus en langue française que l'on croit. Pour répondre aux soucis de l'ACCT, la BIEF a publié en mars 1995 un annuaire de base de données pour la Francophonie. Cet annuaire nous démontre qu'il existe des contenus francophones intéressants actuellement interrogeables par les experts de la Francophonie. Cet outil de référence recense environ 900 bases de données disponibles sur disques compacts, en télématique et en accès kiosque (Minitel),

dont l'une des langues de contenu ou d'interrogation est le français.

Une seconde édition de cet annuaire est d'ailleurs prévue pour le premier semestre de 1996. Dans cette nouvelle édition, nous retrouverons aussi l'information concernant les bases de données disponibles par Internet.

Un disque compact de « répertoires/annuaires » est également en cours de production. Ce disque regroupera tous les répertoires et annuaires publiés par l'ACCT tels que le *Répertoire des fournisseurs d'information francophones pour le développement*, le *Répertoire des sigles et acronymes en usage dans la Francophonie*, les *Profilis géo-documentaires des États et Gouvernements membres des Sommets francophones*, le *Répertoire d'information sur les droits de la personne en Francophonie*, l'*Annuaire des centres de lectures et d'animation culturelle (CLAC)*...

Deuxième priorité : brancher la Francophonie sur Internet et brancher Internet sur la Francophonie, c'est-à-dire installer une passerelle Internet avec serveur et adresse. Ce qui est fait. Sur ce serveur Internet de l'ACCT, nous retrouvons des pages d'information sur l'organisation, ses réseaux et ses partenaires, il nous fournit également l'adresse Internet des systèmes d'information spécialisés comme la BIEF, l'IEPF, le CIFDI, etc.

Dans cette philosophie des inforoutes, il faut s'assurer que les productions de l'ACCT et de ses partenaires sont conformes aux normes internationales et qu'elles sont inscrites dans les grandes bases de données et répertoires internationaux. C'est ainsi que nous garantissons l'accès et la promotion des productions francophones auprès des consommateurs.

L'ACCT vise également à aider les pays de la Francophonie, notamment ceux du Sud et de l'Est, à

se doter d'équipements performants dans le domaine des télécommunications et des technologies de l'information (avec 12 % de la population mondiale, l'Afrique n'a que 2 % des lignes téléphoniques). La BIEF a déjà contribué à l'enrichissement des infrastructures, équipements informatiques et autres matériels technologiques à travers son réseau de centres serveurs. L'ACCT utilise donc les points BIEF pour favoriser le développement de réseaux locaux et leur raccordement à des réseaux nationaux, régionaux et internationaux. Le lien technologique entre les centres serveurs nous permettra ainsi d'avoir accès aux banques de données des pays du Sud (Afrique, Europe de l'Est, Asie).

Bien que la Francophonie pave la voie à une autoroute francophone de l'information, elle ne met pas pour autant de côté des supports traditionnels comme l'imprimé qui demeure extrêmement important et présent. Un équilibre et une complémentarité de supports sont essentiels à la démocratisation de l'accès à l'information.

Les enjeux culturels et les enjeux économiques sont importants. L'urgence d'agir, de se positionner, d'être présente en français sur Internet et les inforoutes, de préserver et aussi de promouvoir la diversité des identités culturelles nationales, doivent être plus que des slogans. L'ACCT a le devoir de conscientiser les États membres comme elle l'a d'ailleurs fait en décembre dernier lors du Sommet de Cotonou. Elle agit en fédérant ses programmes documentaires, les axant sur la production de contenus, en renforçant des programmes comme la BIEF et plus particulièrement son réseau de centres serveurs.

L'ACCT compose avec le phénomène des inforoutes; elle contribue à assurer la présence de la Francophonie sur Internet et sur les

grands réseaux d'information. Pour ce faire, elle souhaite s'allier et collaborer avec ses partenaires pour confirmer cette présence.

Comme l'indique notre brochure descriptive *Clin d'œil sur la BIEF, une complicité multilatérale*, il est préférable de joindre nos efforts plutôt que de reconstruire la roue chacun de notre côté.

Comme vous l'avez sans doute remarqué, j'ai concentré ma présentation sur la BIEF et le phénomène des inforoutes. Mais la BIEF, c'est beaucoup plus que ça. Je vous invite à prendre connaissance du contenu de notre bulletin d'information et de liaison *L'Écluse*. Ce bulletin est publié quatre fois par année et expose les activités de la BIEF en cours. Nous retrouvons également une « Chronique des partenaires ». Ce sont des textes envoyés par des partenaires de la BIEF, plus particulièrement ceux du Sud. Sous la rubrique « Horizon documentaire », vous y retrouverez les nouveautés en matière de publications, de disques compacts, etc.

Nous avons publié deux numéros spéciaux :

- « Écoles des sciences de l'information en Francophonie »;
- « Les bibliothèques des grandes organisations internationales » telles que la Bibliothèque des Nations Unies à Genève, la Bibliothèque du Bureau international du travail. *L'Écluse* est diffusé partout à travers le monde et rejoint les passionnés de la Francophonie.

*Lucie Lépine,
Banque internationale d'information
sur les États francophones (BIEF).*

Les avenues de collaboration entre les banques de terminologie

Le rôle du Bureau de la traduction du Canada s'inscrit dans la situation sociolinguistique qui prévaut en Amérique du Nord. Il consiste à effectuer le transfert linguistique de la réalité anglophone à la réalité francophone, notamment à l'aide de Termium®, banque de données linguistiques du Gouvernement du Canada. En dépit de certaines barrières (spécificité des banques, situation commerciale, problématique de gestion et considérations juridiques), les échanges interbanques permettraient de combler des lacunes dans certaines langues de spécialité et d'enrichir le contenu multilingue des banques concernées.

Termes-clés :
Termium®; Bureau de la traduction du Canada; barrières à l'échange; avenues de collaboration interbanques; avenir des banques de terminologie.

1 Situation sociolinguistique

En Amérique du Nord, la situation sociolinguistique se caractérise par une présence marquée de l'anglais et une présence minoritaire du français. Les innovations technologiques et industrielles qui marquent cette fin du XX^e siècle sont légion. Il en découle une apparition continue de concepts divers qui connaissent rapidement une large diffusion. Il s'ensuit inévitablement une intense activité de création lexicale et terminologique qui accompagne l'émergence de ces nouveaux concepts. Les États-Unis et le Canada, pays où les progrès technologiques sont particulièrement importants, sont très touchés par le mouvement d'ensemble qui transforme langue et société. Dans ce contexte, le rôle du Bureau de la traduction de Travaux publics et Services gouvernementaux Canada consiste à effectuer un énorme transfert traductionnel, terminologique et linguistique de la réalité anglophone canado-américaine à la réalité francophone canadienne à l'aide de Termium®, banque de données linguistiques du Gouvernement du Canada.

2 Bureau de la traduction

2.1 Présentation

Créé en 1934, le Bureau de la traduction du Gouvernement du Canada, qui relève du ministère de Travaux publics et Services gouvernementaux Canada, compte quelque 800 traducteurs, terminologues et interprètes, répartis en divers points de service à travers le Canada. Chaque année, le Bureau de la traduction traduit environ 300 millions de mots dans tous les domaines de l'activité humaine, c'est-à-dire à peu près un million de pages.

Pour sa part, la Direction de la terminologie et de la documentation du Bureau de la traduction affecte ses ressources à diverses activités : alimentation de Termium®, gestion de Termium®, documentation, consultations terminologiques, publications, échanges et francophonie ainsi que normalisation.

2.2 Demande en traduction

En vertu de la Loi sur les langues officielles, édictée en 1969 et modifiée en 1988, la majorité des documents produits par le gouvernement fédéral doivent être disponibles dans les deux langues officielles du Canada, soit le français et l'anglais.

Or, en raison de la proportion d'anglophones et de francophones au Canada, ces documents sont généralement rédigés en anglais et traduits vers le français. Cette situation se reflète au Bureau de la traduction dans la répartition de la demande de traduction par langue.

Ainsi, la majorité des textes sont traduits de l'anglais vers le français dans une proportion de 82 %, du français vers l'anglais dans une proportion de 12 % et, enfin, de l'anglais et du français vers quelque 60 autres langues et vice versa dans une proportion de 6 %.

Parmi les 60 autres langues traduites par le Service de traduction multilingue du Bureau de la traduction, les langues où le volume de traduction est le plus important sont l'allemand dans une proportion de 17 %, le russe dans une proportion de 16 % et l'espagnol dans une proportion de 11 %.

Avec l'Aléna (Accord de libre-échange nord-américain), la traduction vers l'espagnol connaît une augmentation sans cesse croissante.

2.3 Produits et services

En terminologie, le Bureau de la traduction dispose de trois produits principaux bien connus dans les milieux linguistiques au Canada et à l'étranger, soit Termium®, les publications terminologiques spécialisées et *L'Actualité terminologique*. En plus de ces produits, le Bureau offre également un service de consultation terminologique.

Ces produits et services, qui ont pour objectif d'améliorer la communication dans les deux langues officielles du Canada, visent une clientèle vaste et diversifiée :

- L'administration fédérale;
- Les langagiers du secteur public et privé;
- Les gouvernements provinciaux, municipaux et étrangers;
- Les établissements d'enseignements canadiens et étrangers;
- Les entreprises privées canadiennes et étrangères;
- Les organismes internationaux.

Termium®, banque de données linguistiques du Gouvernement du Canada, offerte sur disque optique

compact, compte quelque trois millions de termes dans tous les domaines de la connaissance incluant des appellations officielles, et accorde un traitement équivalent au français et à l'anglais. On y trouve autant la terminologie fondamentale que la terminologie de pointe de divers domaines. Certaines fiches comportent des données phraséologiques en plus des données qui figurent habituellement sur les fiches de terminologie (définitions, contextes, exemples d'utilisation, observations et notes).

Depuis près d'une trentaine d'années, mais surtout depuis 1980, le Bureau de la traduction a une abondante productivité en matière de publications terminologiques. Ainsi, les collections *Petits lexiques* et *Bulletins de terminologie* comptent quelque 150 lexiques et vocabulaires bilingues anglais-français dans une quinzaine de grands domaines.

Publié depuis 1967, *L'Actualité terminologique*, périodique trimestriel qui célébrera son trentième anniversaire en 1997, fait le point sur l'activité langagière en français et en anglais.

Enfin, le service de consultation terminologique bilingue du Bureau de la traduction traite, bon an mal an, quelque 100 000 demandes, incluant les requêtes relatives aux appellations officielles en français et en anglais.

3 La banque de données linguistiques du Gouvernement du Canada

À la fin de l'exercice financier 1994-1995, les données emmagasinées dans Termium® provenaient des activités de dépouillement, de recherche, de publication, de consultation terminologique et de normalisation menées par la Direction de la terminologie et de la documentation,

la plus grande partie des fiches résultant toutefois des projets menés pour le compte de clients.

Pour ce qui est du contenu de la banque, on note un accroissement notable du nombre de fiches dans les domaines suivants entre 1988 et 1995 :

- Appellations officielles (80 000 fiches);
- Industries (25 000 fiches);
- Sciences humaines (40 000 fiches);
- Militaire et paramilitaire (30 000 fiches);
- Électronique et informatique (30 000 fiches);
- Droit et justice (20 000 fiches);
- Sciences du milieu (10 000 fiches).

Dans le contexte canadien, le contenu de Termium® est en majorité bilingue anglais-français. Toutefois, en raison de la mondialisation des marchés et des échanges, les besoins se font de plus en plus pressants en terminologie multilingue.

4 Avenues de collaboration

4.1 Situation actuelle

Le Bureau de la traduction est en voie de renégocier une vingtaine d'ententes visant soit l'échange de données terminologiques, soit la réalisation de projets conjoints. De plus, il y a possibilité de relancer d'autres ententes, d'où un certain apport nouveau à Termium®.

Dans la plupart de ces ententes, le Bureau de la traduction remet des exemplaires de *Termium® sur CD-ROM* ou de ses publications terminologiques en échange des contributions des collaborateurs extérieurs qui peuvent prendre diverses formes.

4.2 Barrières à l'échange

Désireux de poursuivre ses échanges, le Bureau de la traduction

constate que des barrières peuvent freiner la collaboration complète entre les banques de terminologie. Essayons de relever certaines d'entre elles afin de mesurer l'ampleur des hésitations qui pourraient se poser entre exploitants.

À l'heure actuelle, les principales barrières à l'échange que connaissent les banques de terminologie sont les suivantes :

- Spécificité de la vocation des banques ;
- Situation commerciale ;
- Problématique de gestion ;
- Considérations juridiques.

4.2.1 Caractère spécifique des banques

Les banques sont mises au point en fonction d'un mandat, d'une mission et d'objectifs bien précis, ce qui donne lieu à l'exploitation des champs de spécialité propres à la clientèle desservie.

Les échanges interbanques doivent donc viser à répondre aux besoins spécifiques des exploitants de chaque banque. Ainsi, des données, intéressantes en elles-mêmes, mais étrangères aux champs d'activité propres à une banque, pourraient constituer un fardeau inutile pour cette dernière. En raison de la clientèle assez précise qu'elle dessert généralement, chacune des banques n'est pas toujours immédiatement intéressée par l'importation d'autres langues ou d'autres données appartenant à des champs d'activités se situant trop loin de ses préoccupations premières.

De fait, les exploitants de banques voudront peut-être conserver et cultiver ces éléments de différenciation, qui deviennent des avantages marqués en situation de vente, en situation de marketing.

Si la spécificité des banques de terminologie constitue, pour les exploitants, un atout en fonction d'une clientèle visée, il faut noter au passage qu'elle peut être une barrière en termes de coûts pour la clientèle

qui doit alors s'abonner à plus d'une banque de terminologie pour répondre à des besoins plus diversifiés.

4.2.2 Situation commerciale

En raison des limites que leur imposent les mandats respectifs des organismes dont elles relèvent, les banques de terminologie institutionnelles pourraient connaître des difficultés à augmenter le nombre de domaines d'application.

Certaines banques de terminologie commencent à œuvrer dans des environnements de type commercial. La propension des gouvernements vers le recouvrement accru des coûts touche aussi les exploitants de banques de terminologie. Jusqu'à quel point les possibilités de collaboration seront-elles assombries par la volonté de protéger des « secrets d'entreprise » ? La réponse à cette question dictera, dans une large mesure, les politiques d'échanges et de collaboration interbanques.

4.2.3 Problématique de gestion

La taille des banques est un facteur dont il faut tenir compte. Il n'est pas toujours facile de trouver des mécanismes de mise à jour appropriés qui nous permettent de garantir la qualité et l'harmonisation des données emmagasinées.

Des coûts de traitement relativement élevés doivent donc être consentis pour faire en sorte que de nouvelles données puissent être intégrées en toute harmonie avec les données existantes. Faute de ce travail, le fonds terminologique risque vite de devenir encombré par des données provenant de l'extérieur.

Par ailleurs, dans le cas de la terminologie multilingue, peut se poser pour une banque de terminologie, la question des ressources disponibles dans les langues traitées, ce qui accroît encore davantage le coût du traitement des

nouvelles fiches à emmagasiner. L'acquisition d'une terminologie nouvelle est une chose, l'intégration de cette terminologie au contenu d'une banque en est une autre, du point de vue de la gestion et des coûts. La difficulté des échanges se situe presque toujours à ce niveau.

4.2.4 Considérations juridiques

Les questions relatives à l'utilisation raisonnable des sources de référence dans les banques de terminologie et à la diffusion du contenu de ces mêmes banques sont parmi les questions les plus importantes qui surgissent dans l'exploitation des grandes banques.

Les avis des spécialistes en droit, s'ils vont toujours dans le même sens, diffèrent quant à la portée à donner à la protection des auteurs et quant à la définition du caractère original d'une œuvre.

4.3 Avenues de collaboration

En dépit de toutes ces difficultés, il existe des possibilités de collaboration interbanques intéressantes et elles doivent être prises en considération par les exploitants concernés. L'échange de personnel entre banques de terminologie pour fins de partage d'expertise, d'enrichissement mutuel et de perfectionnement est d'application facile.

Compte tenu notamment de l'Aléna en Amérique, du Traité de l'Union européenne, de la place de l'Afrique dans la Francophonie ainsi que de l'intérêt commercial de certains continents pour l'Asie, il est à prévoir que le XXI^e siècle sera celui de la terminologie multilingue. L'essor du multilinguisme est un justificatif de taille pour inciter à la collaboration et au partage. S'il revêt un intérêt marqué, le multilinguisme devra constituer un des principaux axes de développement dans les années à venir, malgré ses coûts élevés

d'implantation et d'exploitation pour les banques de terminologie.

5 Avenir des banques de terminologie

En dépit des contraintes soulevées précédemment, on peut soutenir l'idée que l'exploitation des grandes banques de terminologie, Termium® en étant un exemple, se fera plutôt à la fois par l'échange de données pertinentes entre bases de données existantes et par l'acquisition de fonds terminologiques limités et spécialisés. L'utilité des banques de terminologie sera toujours fonction des éléments suivants :

- Qualité de l'information offerte;
- Quantité des données emmagasinées;
- Rapidité et facilité d'accès à ces données.

Enfin, on peut déceler un mouvement de gestion de l'information générale de masse et, dans cette optique, les clés d'accès à cette information deviennent d'une importance capitale. En effet, de plus en plus, les organisations aussi bien gouvernementales que privées mettent sur pied d'énormes fonds d'information. Il est facile de se perdre dans cette mer de données, et la seule façon d'utiliser efficacement ces renseignements est d'avoir des clés d'accès à la fois significatives et précises. Ces clés sont des termes, des expressions justes dont la définition relève de l'activité terminologique. Une terminologie juste permet l'accès plus ciblé à cette information. D'où la possibilité de nouveaux créneaux à exploiter pour les organismes de terminologie et des débouchés en perspective pour l'avenir de la profession. Peut-être est-ce un signe de renouveau des banques de terminologie...

*Louis-Claude Tremblay,
Bureau de la traduction,
Canada.*

Vers la fiche virtuelle en terminologie

Le travail du terminologue exige une grande polyvalence et se transforme depuis l'introduction des documents textuels électroniques. Sa tâche passera de la production classique de la fiche de terminologie à une gestion de plus en plus générale de l'information de nature linguistique.

Sous l'influence de la technique, la forme matérielle de la fiche actuelle, et même informatique, disparaîtra peut-être au profit d'une fiche virtuelle.

Termes-clés :
fiche virtuelle; fiche terminologique;
traitement informatique de la langue;
chaîne de traitement.

1 Le travail terminologique

En raison de l'introduction de plus en plus importante des documents textuels électroniques, de nouvelles requêtes sont faites au terminologue.

Certes, on continue à lui réclamer des fiches terminologiques mais on veut aussi, par exemple, qu'il :

- Détermine le vocabulaire spécifique d'une entreprise ou d'un service (vocabulaire de spécialité);
- S'assure que le vocabulaire des écrits d'une entreprise est stable tant dans sa publicité, ses écrits techniques, sa correspondance officielle;
- Ajuste le vocabulaire aux normes internationales et nationales;
- Assiste à la construction d'instruments documentaires et de gestion de l'information (thésaurus, dictionnaire, lexique, aide à la traduction, etc.);
- Lie son lexique à des bases de connaissances, etc.

Bien que le terminologue, par sa formation linguistique, ne soit pas préparé à une telle variété de tâches, le milieu économique et la technologie exigeront de lui une plus grande polyvalence. Dans une entreprise, on voudra de moins en moins assumer de manière séparée les coûts reliés à la préparation des fiches terminologiques, des index et des thésaurus, parce qu'ils seront de plus en plus liés à une même base d'information. Le problème sera d'autant plus ressenti que

l'information sera dynamique, c'est-à-dire en constante évolution. Pensons ici à Internet et Intranet qui livreront l'information dans un flux perpétuel où l'indexation, la terminologie, les thésaurus seront en perpétuelle modification.

Bref, le terminologue verra sa tâche passer lentement de la production classique de la fiche terminologique à une exploration et à une gestion de plus en plus générales de l'information de nature linguistique.

Dans leur état actuel, les technologies de l'information qu'utilisent les terminologues ne permettent pas de saisir cette éventuelle transformation. La majorité des logiciels sont soit des instruments de traitement de texte ou, dans le meilleur des cas, des logiciels de bases de données spécialisés dans la cueillette et l'archivage des données terminologiques. Cette technologie est périphérique au travail traditionnel, et il n'existe que très peu de logiciels sur l'assistance à la cueillette et à l'analyse terminologique elle-même; leurs liens avec les logiciels de gestion de l'information de nature linguistique sont relativement limités. On acquiert donc une impression que ce type de travail classique sera éternel. Il sera simplement mieux assisté et peut-être accéléré par cette technologie. Pourtant, l'architecture même de la technologie se modifie grandement et elle ne sera pas sans influencer le travail du terminologue.

2 Vers une compréhension cognitive du travail du terminologue

Le travail du terminologue est habituellement défini en termes d'une discipline spécifique : à savoir la linguistique et surtout d'une technologie particulière : à savoir le papier et le crayon. Mais on peut aussi le penser autrement. Ce travail peut, en effet, être analysé en termes cognitifs, c'est-à-dire en regard des opérations intellectuelles qui y sont accomplies. Dans cette perspective, on verra que le terminologue effectue une chaîne de traitement complexe sur de l'information de divers types. Ainsi, il manipulera des textes qu'il interprétera et dans lesquels il reconnaîtra des termes spécifiques. Certains de ces termes seront retenus et recueillis pour être analysés, définis et, ultimement, déposés dans une base de données. Cette chaîne de traitement ne sera pas directe mais constituée d'allers et de retours jusqu'à une consolidation finale. Celle-ci sera alors consignée dans des fiches terminologiques. En ce sens, la fiche n'est que la pointe de l'iceberg de la chaîne de traitement. Elle n'en est qu'une forme de manifestation particulière.

En termes plus synthétiques, on peut schématiser cette chaîne de traitement en quatre moments :

- Une manipulation matérielle;
- Une classification;
- Une sélection;
- Une reconfiguration.

Vu ainsi, le travail du terminologue apparaît comme une modalité particulière de traitement cognitif de l'information. Seule la nature linguistique de cette information le distingue des autres types de traitement de l'information. Le travail est complexe, certes, mais il est décomposable en des opérations et des sous-opérations. Cependant, en raison même de cette complexité de

la tâche, il n'est pas question de la robotiser. L'expérience montre que ce travail ne peut être automatisé dans son ensemble. Le terminologue doit toujours contrôler le déroulement des opérations, mais il est concevable que certaines de ses composantes soient assistées par ordinateur.

3 La modularité dans le traitement de l'information

Depuis une dizaine d'années, la conception même des logiciels se modifie. Les avancées du génie logiciel et le développement des langages de programmation permettent de penser, de plus en plus, les chaînes de traitement comme des assemblages de modules autonomes, c'est-à-dire comme des « objets » ou « processus » autonomes qu'on peut appeler individuellement pour réaliser des composantes des tâches complexes. Les nouvelles plates-formes et les systèmes d'exploitation des mini-ordinateurs nous ont habitués à opérer dans ce type de cadre. Aujourd'hui, un bon logiciel de traitement de texte n'est pas un robot qui écrit un texte pour un utilisateur, mais une boîte à outils, ou un atelier, qui met des modules à la disposition de l'utilisateur. C'est lui qui décide de la chaîne de traitement qu'il produira selon ses besoins.

Dans cette même perspective, la recherche et le développement, dans le domaine du traitement informatique de l'information, commencent à offrir des systèmes modulaires de plus en plus complexes qui sont intégrés dans des plates-formes qu'un utilisateur peut contrôler. On s'éloigne donc des logiciels boîtes noires pour aller vers de la modulation et de l'interaction, etc. De plus en plus, la recherche et le développement, dans le génie logiciel et linguistique, offrent des modules qui pourront s'intégrer les uns aux autres. On trouve déjà sur le marché :

- Des analyseurs linguistiques : lemmatiseurs, analyseurs

morphologiques, identificateurs de mots composés ;

- Des outils linguistiques : dictionnaires, thésaurus, lexiques, encyclopédies, bases de données, etc ;
- Des outils de fouille et d'analyse : logiciels de gestion et d'analyse textuelle, etc ;
- Des outils numériques : logiciels statistiques, classifieurs, etc.

Tous ces nouveaux modules permettent de plus en plus de construire des cheminements spécialisés et adaptés aux besoins des spécialistes du traitement informatique de la langue.

4 Vers la fiche virtuelle

En raison de la technologie plume et papier utilisée habituellement par le terminologue, il lui a fallu développer des chaînes de traitement qui construisaient des fiches. Celles-ci devaient être stables, composées de champs fixes et, finalement, elles devaient être déposées dans des banques aux fins de conservation. Mais avec une technologie changeante alliée à une meilleure compréhension de la tâche du terminologue, rien n'exige que les résultats de son travail prennent une telle forme. Certes, il y aura toujours la nécessité de donner toute l'information que cette fiche présentait, mais il faut se demander si elle doit encore prendre cette forme et être encore mise de cette manière dans des banques spécialisées. Doit-on encore construire des fiches terminologiques qui décrivent tous les termes avant même qu'on sache si on a vraiment besoin? Pourquoi ne pas créer la fiche sur demande, selon le besoin?

D'un point de vue formel, la fiche n'est qu'une matrice constituée d'informations de divers types (champs issus de plusieurs sources). Ne pourrait-on pas penser que chaque type d'information puisse être

construit et rappelé sur demande par des modules spécialisés dans ces tâches? Dans cette vision des choses, la fiche devient virtuelle. Sa forme matérielle, et même informatique actuelle, disparaît mais l'information qu'elle comprenait est retraceable au moyen des stratégies de fouilles et d'analyse, etc.

L'idée peut sembler lointaine, mais il faut ici comparer la situation avec la fiche bibliographique de nos bibliothèques. Où sont les grands tiroirs bibliographiques d'antan? Même les bases de données bibliographiques tendent à disparaître. Code, titre, auteurs, prix, éditeurs, nombres de pages, etc., ne sont plus réunis dans une même fiche. Ils sont rappelés sur demande selon les requêtes qu'on construit autour de l'univers du livre. La fiche dans ce domaine devient de plus en plus virtuelle.

En terminologie, on peut accepter que toute l'information pertinente nécessaire à la construction de la fiche classique existe ailleurs sous d'autres formes : textes, dictionnaires, encyclopédies, lexiques, mémoires institutionnelles. Des chaînes de traitement spécialisées pourraient récupérer sur demande l'information pertinente à des requêtes. Il n'y aurait plus besoin de construire des banques de fiches. L'information terminologique serait toujours accessible, mais elle se présenterait sous des formes de plus en plus diversifiées. En ce sens, la fiche virtuelle est une idée logiquement possible bien que concrètement encore lointaine. Mais elle n'est pas aussi loin qu'on le croit.

*Jean-Guy Meunier,
Laboratoire d'analyse cognitive de
l'information (LANCI),
Université du Québec à Montréal,
Québec.*

Évolution des banques de données terminologiques

Comprendre les raisons qui font évoluer les banques de données terminologiques et les nouvelles orientations qu'elles prennent constitue un enjeu particulièrement sérieux pour quiconque produit, utilise, exploite, ou conçoit des ressources terminologiques à une échelle qui se veut industrielle ou, pour le moins, résolument « professionnelle ». Le point de vue exposé ici s'appuie sur une multitude d'approches de la terminologie et des banques : l'approche du producteur, l'approche du multiutilisateur, l'approche du chercheur, l'approche du concepteur de systèmes, et, pourrait-on dire, l'approche du « vendeur »

Termes-clés :
banque de terminologie; module de banque; producteur de données terminologiques; chaîne terminographique.

(1) Dès l'instant où les facteurs technologiques agissent sur l'offre et la demande, ils ont des effets économiques. Cependant, ils ne peuvent être considérés, à proprement parler, comme de vrais facteurs économiques.

Facteurs d'évolution

Si l'on excepte le facteur politique, l'évolution actuelle et future des banques de terminologie est poussée ou tirée par trois grandes séries de facteurs : les facteurs idéologiques, les facteurs technologiques et les facteurs économiques.

Les facteurs idéologiques tiennent essentiellement à la représentation que les opérateurs de banques de terminologie peuvent avoir de leur produit. Pendant très longtemps, la terminologie a été réduite à une aide à la traduction. Il fallait bien, à l'époque, aligner le discours sur le seul argument économique sérieux. Aujourd'hui, la terminologie se pousse du col (à moins que ce ne soient les « nouveaux terminologues ») et revendique une place privilégiée dans les contextes d'intelligence artificielle (base de connaissances), d'aide à la rédaction, d'ingénierie linguistique, et ainsi de suite. Et c'est précisément là que l'idéologie intervient pour réclamer un droit à l'autonomie d'une discipline, d'une pratique, et de produits terminologiques qui, du point de vue des bailleurs de fonds, ne sont qu'une subdivision de l'ingénierie linguistique. L'un des enjeux, pour les opérateurs de banques et, en amont, les concepteurs et pourvoyeurs de ces banques, est de déterminer la place (le positionnement, disent les commerciaux) du produit terminologique sur le marché. Pour que les banques de terminologie continuent d'exister et de se

développer dans de bonnes conditions, il faut que leur importance continue d'être reconnue par ceux qui la reconnaissaient déjà et soit enfin reconnue par ceux dont l'intérêt serait d'orienter les investissements vers d'autres produits que la terminologie proprement dite, ou d'autres programmes que les systèmes de gestion et d'exploitation de la seule terminologie.

En ce qui concerne la technologie, toutes les révolutions sont en marche. La puissance de traitement des machines et des logiciels, les possibilités accrues de diffusion (CD-ROM, réseaux intra- et interentreprises, réseau de réseaux) et la généralisation des logiciels consommateurs de terminologie contribuent, chacune à leur manière, à modifier les conditions d'exploitation des banques. Elles influencent parallèlement la demande terminologique (en ouvrant et en diversifiant les marchés) et l'offre destinée à répondre à la demande nouvelle. Modifier l'offre, c'est, à terme, modifier les contenus, les matériels, les programmes et la conception même des banques de terminologie et l'on peut aujourd'hui prédire, sans grand risque d'être démenti par les faits, que les moyens technologiques modifieront les procédures de travail des terminographes, le champ des ressources qu'ils exploitent, les modes de gestion de leurs données, le champ et les modalités de diffusion de ces données, et les modes d'exploitation des contenus des banques.

Mais l'influence la plus forte est celle des facteurs que l'on pourrait dire purement économiques⁽¹⁾.

Ces facteurs n'ont rien de bien original. Ils ont pour noms : rentabilité et productivité. Tout opérateur de banque de terminologie sait bien, en effet, qu'on attend de lui qu'il rentabilise l'investissement passé, présent et à venir.

Pour rentabiliser l'investissement passé, il faut le « commercialiser », aller vers les clients en leur proposant un produit de qualité irréprochable répondant strictement à leurs besoins. Il faut donc viser le zéro défaut terminologique (vaste programme!), *relooker* au besoin le produit, et diversifier l'offre et les supports. Pour rentabiliser l'investissement présent et à venir, il faut respecter les mêmes conditions que pour le passé en y ajoutant, autant que possible, des gains de productivité et des gains de qualité. Il faut donc optimiser les procédures et, bien entendu, automatiser tout ce qui peut l'être. Et c'est là qu'intervient l'intéressante conjonction du jour : l'idéologie confirme l'intérêt économique de la terminologie, la technologie offre des capacités inégalées de gestion et de diffusion de terminologies, et la demande se manifeste de plus en plus fortement et se diversifie. Il reste simplement aux opérateurs de banques terminologiques à faire évoluer leur offre.

Une offre plus large et plus proche des utilisateurs

Vendre les contenus des banques terminologiques, c'est d'abord susciter de nouvelles utilisations de la terminologie. Ainsi sommes-nous passés d'une terminologie essentiellement traductive à une terminologie qui se veut outil d'aménagement en entreprise, aide essentielle à la rédaction, outil de formation, clé d'indexation, moteur de nomenclatures, poumon de la gestion électronique de documents,

condition nécessaire de la veille technologique et commerciale, aliment essentiel (carburant) des automates linguistiques de toutes sorte et, bien entendu, base même de toute activité significative et, singulièrement, des activités pilotées par l'intelligence artificielle. Il est admis que les banques de terminologie doivent proposer des éléments ou produits susceptibles de répondre, singulièrement ou globalement, à l'ensemble des « besoins » ci-dessus. Il va donc de soi qu'elles devront s'adapter pour y répondre et que leur adaptation sera d'abord une adaptation de leurs contenus. En même temps, les banques terminologiques doivent se rapprocher physiquement de leurs clients. Cette dernière partie du programme est aujourd'hui réalisable. Les contenus des banques sont disponibles – à des coûts très modestes – sur des supports diversifiés, conviviaux, aisément accessibles, et facilitant un accroissement des volumes disponibles. Ces supports sont les CD-ROM, les réseaux intra-entreprises, les réseaux interentreprises, Internet, et tout ce dont nous n'avons même pas encore idée.

Il est intéressant de noter que l'objectif de diversification de l'offre et des supports a provoqué, chez les « banquiers terminologiques », une réflexion forte sur l'ergonomie de consultation comme sur l'adaptation de l'offre à une multitude de besoins particuliers. Le produit lui-même change.

Un produit adapté

Quels que soient les supports et modes de diffusion, il va de soi que le contenu même du produit terminologique doit évoluer pour répondre aux nouvelles demandes de

la clientèle, aux exigences des terminologues eux-mêmes, aux appels d'air générés par les avancées technologiques et, singulièrement, par les automates consommateurs. Les contenus terminologiques proposés par les banques doivent aujourd'hui se spécialiser afin de répondre aussi étroitement que possible à une combinaison de trois variables : la variable « métiers », la variable « fonctions » et la variable « application/exploitation ». En d'autres termes, il faut que chaque produit terminologique corresponde strictement à un domaine ou aux domaines d'activité voulus (le ou les métiers), à la fonction qu'il doit permettre de remplir (traduction, rédaction, etc.) et, de plus en plus, aux spécifications du logiciel d'exécution de la tâche ou d'aide à l'exécution de la tâche.

L'adaptation à la fonction pose des exigences nouvelles, que confirme et renforce la puissance de traitement des machines : il ne suffit plus que l'offre soit terminologique, elle doit, dans le cadre de sa spécialisation, s'étendre pour inclure la dimension phraséologique (absolument nécessaire), la dimension hypertextuelle et, par le biais des informations et analyses conceptuelles et interconceptuelles, la dimension ontologique. Le produit s'enrichit.

Micro- ou mégabanque ?

Du point de vue du producteur, l'option mégabanque et l'option microbanque demeurent également ouvertes. L'option microbanque est celle du petit producteur traitant un domaine très pointu, le traitant de manière extrêmement poussée, tant sur le plan linguistique que sur le plan conceptuel/notionnel, et développant sa part de marché en adaptant un même contenu à une variété de plates-formes et

d'applications. On voit ainsi se multiplier les microbanques hyperspécialisées (dont les approches terminographiques sont souvent fortement novatrices) directement injectables dans tous les logiciels existants ou même constituées en cartouches de programme.

L'option mégabanque est celle des opérateurs multidomains disposant déjà de stocks impressionnants de terminologies et déterminés à continuer d'alimenter leurs mastodontes. Les actuelles mégabanques sont connues et, contrairement à ce que l'on imagine, de nouvelles pourraient bien apparaître. Le choix est affaire de moyens. Associée à d'autres microbanques, toute microbanque terminologique peut jouer dans la cour des grandes et, inversement, une mégabanque terminologique divisée ou découpée en ensembles homogènes (en « collections ») n'est rien d'autre qu'un conglomérat de microbanques virtuelles. À très brève échéance, le poids des banques se mesurera non pas en nombre de termes mais en nombre de modules complets, opérationnels, définis comme la terminologie d'un secteur donné adaptée à une fonction particulière et à un outil logiciel spécifique sur tel ou tel support.

Vers des banques modulaires ou modularisables

Quels que soient les volumes stockés et quelles que soient les conditions du développement des banques, le souci d'efficacité et le souci d'exhaustivité de traitement conduiront tous les producteurs à raisonner en termes de modules de banques.

Pour compléter les éléments évoqués plus tôt, nous dirons qu'un module de banque (défini par le concepteur d'une mégabanque ou

imaginé par le concepteur d'une microbanque) est un ensemble terminologique complet (dans son inventaire des éléments à traiter), fiable, traité complètement à la date de réalisation du module (en fonction des objectifs définis) et mis à jour dans des délais raccourcis, et directement exploitable. Ce module de banque traite systématiquement, et non pas de manière aléatoire, un objet, une collection d'objets, un champ d'activité ou une activité complète. Pareille définition du module de banque garantit que chaque module est faisable en termes de temps, ressources humaines, ressources matérielles, proximité et disponibilité des compétences, conditions de contrôle de qualité et validation. Elle permet également les tests d'acceptation des utilisateurs.

La construction de banques par modules favorise le traitement en profondeur, l'exhaustivité, l'adaptation aux besoins des clients, la diversité fonctionnelle. Elle permet aussi, pour peu que l'on prenne les dispositions techniques voulues, de partager l'investissement. En effet, la banque terminologique peut se construire par conjonction de modules obéissant pour l'essentiel à des principes d'élaboration homogènes et dans lesquels chaque producteur-utilisateur trouve la monnaie d'échange qui lui permet l'accès à tous les autres modules fédérés. Dans un schéma modulaire, coopératif, chacun dépose à la banque le contenu de son ou ses modules en échange d'un droit de tirage sur les modules des autres métiers. En-deçà de ce schéma optimal (et peut-être optimiste) plusieurs fournisseurs peuvent diffuser leurs ressources terminologiques sur un même support, créant ainsi les conditions de gestion en banque, étant entendu que les mégabanques intègrent par nature une série complète de modules conçus ou produits par un opérateur unique ou des opérateurs multiples.

Petite ou grande, la banque peut se constituer par fusion de modules (fusion de fichiers), importation de modules sur un même serveur central, interconnexion de serveurs, gestion de dictionnaires hétérogènes au moyen d'un même logiciel d'interrogation sur CD-ROM, ou par tout autre moyen. Les questions posées aux opérateurs conjoints sont, indépendamment de la nécessité de définir et d'adopter un standard, des questions de droits et redevances et des questions de technologie et de génie logiciel. La définition d'un standard est affaire de bon sens, les droits et redevances se régissent par l'adoption des conventions voulues et les développements technologiques semblent se préciser au détour du chemin (ou des autoroutes de l'information).

L'un des éléments nouveaux aujourd'hui est précisément que tout producteur de données terminologiques doit tenir compte de la logique de banque et considérer son ou ses produits comme autant de modules potentiellement intégrables à une collection ou à une banque. À défaut, le distributeur devra (comme c'est le cas avec Lexpro-CD Databank) développer un logiciel d'interrogation capable de gérer des modules hétérogènes.

Des modules alimentés et exploités par une chaîne terminographique

Comme nous l'avons vu, les producteurs de banques de données sont confrontés à un problème de rentabilité-productivité-qualité. Pour résoudre ce problème complexe, ils sont nécessairement amenés à exploiter un ensemble d'outils exécutant ou facilitant chacun une tâche précise au fil du processus de construction et exploitation des banques. L'évolution la plus marquée

de ces dernières années a ainsi été la modularisation (et l'automatisation au moins partielle) de chaque fonction en amont ou en aval de la banque. On peut désormais – et les présentations de ces deux journées l'ont abondamment démontré – disposer de modules de collecte automatisée de données terminogènes à partir de corpus, mettre en œuvre des extracteurs, faire exclure les candidats termes par des anti-dictionnaires, organiser des traitements automatiques de séries de fiches ou de fichiers entiers, détecter les doublons, regrouper les dossiers présentant des caractéristiques identiques et, singulièrement, des carences identiques, vérifier la conformité des données à la DTD, exploiter des modalités de stockage particulières, échanger automatiquement des données avec d'autres producteurs ou consommateurs, basculer des collections dans un logiciel d'exploitation particulier, sélectionner toutes les données dont la dernière mise à jour remonte à un nombre déterminé de mois ou d'années, et, bien entendu, multiplier les canaux et supports de diffusion. Chaque producteur de banque construit ainsi sa chaîne terminographique en fonction de ses besoins et de ses moyens et maint chercheur rêve de développer « la » chaîne de production terminographique qui se passerait du terminographe (pour la collecte), du terminologue (pour l'organisation des données), du terminographe à nouveau (pour la préparation des applications), et de l'opérateur de banque, puisqu'il suffirait de disposer à la fois du corpus et de la chaîne automatique de production terminographique pour générer le résultat voulu à la demande. L'avenir radieux de la banque terminologique serait alors un monde dans lequel ne subsisteraient que les informaticiens (pour développer et améliorer la chaîne), le

comptable (pour percevoir les redevances) et le banquier, au sens propre, cette fois (pour financer le tout et en tirer le profit économique).

Des modules terminologiques dans des banques linguistiques/ontologiques

Les modules terminologiques constitués et exploités par une chaîne terminographique et terminotique modulaire s'inscriront, à très court terme, dans des banques de ressources linguistiques/ontologiques. Dans ces banques élargies cohabiteront des couches superposées de matériaux de degrés d'ouvrison variables exploités par des modules fonctionnels adaptés. La couche la plus basse sera constituée des corpus documentaires bruts, exploitables par des agents recherchistes à partir d'équations de recherche de type « clé terminologique + filtre + cible » et générant des contextes définis comme probablement pertinents pour l'utilisateur concerné. La deuxième couche sera constituée par des corpus documentaires indexés et, probablement, alignés – au sens des bitextes actuels – l'indexation s'effectuant par des séries de clés et filtres définis non plus par les contextes mais par les tables du terminologue. La troisième couche sera constituée par des répertoires contextuels correspondant aux précurseurs terminologiques proprement dits. On y trouvera une ressource essentiellement phraséologique et accessoirement terminologique mais aussi, pour peu que les contextes soient triés par des automates performants, des informations conceptuelles-notionnelles significatives. La quatrième couche sera constituée par de l'information terminologique

organisée, indexée, construite, mais pas nécessairement parvenue aux stades ultimes du traitement terminographique habituel. La cinquième couche sera constituée par des collections ou répertoires validés et la sixième et dernière couche sera constituée par un ensemble de matériaux homologués en vue d'une application spécifique de métier-maison-outil. Pour les opérateurs de banques terminologiques, la solution à leurs problèmes de mise à jour, de main-d'œuvre, de productivité et d'adaptation à la clientèle passe par la conjonction, sur un même site et un même support, d'une matière première et de tous les produits dérivés intermédiaires jusqu'au produit fini (provisoirement). Quel que soit le statut futur de la terminologie, il apparaît à l'évidence qu'elle ne doit pas (ne peut pas?) demeurer un produit isolé. Elle est l'une des composantes de la ressource linguistique et « sa » banque doit devenir une banque de ressources linguistiques. Il est sans doute temps de penser en termes de banques de données linguistiques-conceptuelles dans lesquelles la terminologie (au sens où nous l'entendons traditionnellement) apparaîtrait dans tous ses états et dans tous ses environnements, même et surtout provisoires.

Des données terminologiques individualisées

En même temps que la banque terminologique se fondera dans une banque linguistique/ontologique et pour permettre cette intégration, on peut parier que, dans un avenir très proche, le dossier terminologique traditionnel éclate ou entre dans un schéma de virtualité d'où le sortiraient, chacun à leur manière et selon leurs besoins, les utilisateurs de

la ressource. Dans cette perspective, les banques terminologiques-langagières seraient tout simplement un dépôt d'entités constituant chacune une relation entre une clé primaire, avec ses attributs et son filtre, et une unité d'information (terminologique, phraséologique ou conceptuelle) indexée par une clé secondaire (ou étiquette), l'entité étant accompagnée de ses attributs intrinsèques, de ses attributs d'état, de tout commentaire utile, et de toute information nécessaire à sa gestion et à son administration. Dans un système de ce type, la clé primaire est une chaîne de caractères, un symbole, un code, un pictogramme, etc. La clé secondaire est une chaîne de caractères, un symbole, un code, un pictogramme, etc. L'unité d'information est une donnée, un matériau, une information proprement dite, un symbole, une chaîne de caractères, etc. accompagnée de ses propres attributs (source, origine, état, etc.). Chaque dossier se construit à partir d'une clé et d'une combinaison d'étiquettes de relations, selon les circonstances et selon les besoins.

Dans cette hypothèse, il n'y a plus de fiches. L'information est consignée et traitée, mais aussi consommée, lorsqu'elle devient disponible, selon son état d'ouvroison (spécifié par des marqueurs adaptés). Tous les problèmes théoriques et pratiques liés à la terminologie

(2) Ressources constituées pour le compte des clients ou constituées par eux-mêmes et concédées à l'exploitant du service linguistique.

(3) La situation est alors celle de la société de service linguistique agissant comme fédérateur et distributeur aux titulaires de droits de tirage de l'ensemble des ressources des divers clients selon des conventions bilatérales ou multilatérales.

s'évaporent et la standardisation est acquise d'emblée, puisqu'il suffit de définir les étiquettes de relations. En même temps, la modularité totale des entités prototerminologiques établit le lien avec les autres ressources ou couches de la banque, puisque ces dernières ne sont rien d'autre que les collections de précurseurs des entités identifiées, au stade final de leur ouvroison, comme des entités terminologiques.

Les points de vue des clients

Les produits terminologiques s'adressent à deux types de clients : le consommateur ponctuel, d'une part, et le consommateur systématique, d'autre part. Si le consommateur ponctuel peut se contenter du répertoire qui répond strictement et complètement à son besoin en termes de métier-fonction-application, le rêve du consommateur systématique serait, bien évidemment, de disposer d'une sorte de banque mondiale de la terminologie, auprès de laquelle il retirerait, selon ses besoins du moment et dans les limites strictes de ces besoins, les données voulues – la contrepartie étant la redevance ou le troc de ses propres données ou de ses propres modules.

Du point de vue du client, la ressource est multiple et éclatée et son objectif est de centraliser l'accès à cette ressource chez lui. On peut aujourd'hui décrire la ressource terminologique des sociétés de service linguistique, mais aussi de professionnels isolés, comme une sorte d'interbanque (*clearing house*) personnalisée exploitant en direct les ressources physiquement maîtrisables et ouvrant les guichets voulus vers les fichiers et banques physiquement distants. Une interbanque terminologique locale standard c'est, par exemple, Termium, BTQ,

Termdat et Lexpro-CD sur CD-ROM (de préférence sur infoserveur avec interrogation par terminal à multifenêtrage), l'accès à Eurodicautom par modem, un serveur local de terminologie sur lequel sont implantées toutes les ressources propres de la maison plus toutes celles des clients⁽²⁾ gérées individuellement ou en pool⁽³⁾ avec, pour faire bonne mesure, la possibilité d'explorer et récupérer sur Internet toutes les ressources utiles disponibles. La situation n'est pas foncièrement nouvelle par rapport à l'époque où le client se constituait une bibliothèque aussi pertinente et complète que possible. Ce qui a changé, ce sont les supports de gestion et les modes de consultation en même temps, bien entendu, que la disponibilité des ressources, en volume et donc en diversité. On ajoutera, pour être complet, que la tendance est à la numérisation des dictionnaires diffusés sur papier afin d'inclure leurs contenus dans les systèmes terminologiques locaux. On peut ainsi craindre, par parenthèse, que la facilité de récupération des données ne signe la fin même de l'activité terminographique par « infocopillage ». La diffusion illicite de copies numérisées de dictionnaires pourrait bien mettre les banques terminologiques, grandes et petites, en faillite.

En attendant ce jour que tous espèrent lointain, le « client » standard des banques terminologiques a pour objectif de consulter ou récupérer ce qui existe et de l'ajouter à son propre trésor pour monter « sa » banque.

Conclusion

Dans le souci d'ouvrir leurs marchés, d'aller vers les utilisateurs, de diversifier les exploitations et supports possibles, de rentabiliser

l'investissement qu'elles nécessitent et de tirer avantage des nouvelles capacités des matériels et logiciels désormais proposés, les banques terminologiques doivent aller jusqu'au bout d'une évolution déjà bien entamée.

L'évolution va dans le sens de la mise en place de banques linguistiques/ontologiques concentrées, réparties ou distribuées, dont la composante terminologique constituera le cœur et, bien entendu, l'ensemble des tables de clés. La composante terminologique sera un conglomérat de modules interconnectés extraits du corpus de base au moyen d'une chaîne terminographique modulaire et mis en œuvre au moyen d'une chaîne aval incluant une chaîne de distribution élaborée. Elle constituera l'une des strates du matériau linguistique auquel elle conférera sa valeur ajoutée et dont elle tirera elle-même l'essentiel de ses plus-values. La ressource terminologique se présentera ainsi à l'état fini ou semi-fini et, probablement, à l'état éclaté, prête à toutes les recombinaisons et élaborations nécessaires pour coller au besoin conjoncturel de chaque client et décourager le piratage, risque mortel auquel sont déjà confrontés les producteurs de données terminologiques et les opérateurs de banques terminologiques. Le guichet devra représenter la banque et le kaléidoscope remplacer la fiche.

*Daniel Gouadec,
Laboratoire CRAIE,
Université de Rennes 2,
France.*

M^{me} Lucie Lépine précise que les 47 centres serveurs de la BIEF répartis dans les pays francophones sont dotés d'équipement informatique adéquat pour assurer les télécommunications. La BIEF est disposée à travailler en ce sens en partenariat.

La discussion s'engage ensuite sur la disponibilité de textes scientifiques et techniques sur support informatique. M. Louis-Jean Rousseau fait remarquer que les terminologues disposent d'outils informatiques pour l'analyse de textes, mais que les textes sont plutôt accessibles sur support imprimé que sur support informatique. Il s'ensuit des opérations coûteuses de numérisation qui n'invitent pas les terminologues à utiliser les systèmes qui sont à leur disposition.

M^{me} Lépine souscrit à ces propos et signale que la BIEF encourage la constitution de répertoires signalant les ressources qui existent sur support électronique et qui sont accessibles, notamment par Internet. Elle cite des exemples pour les textes en anglais et signale l'importance de faire de même pour ce qui est du français.

En réponse à une remarque de M. Meunier sur le fait que la fiche de terminologie actuelle dérive de la fiche documentaire classique, M. Gouadec fait remarquer que le changement imminent relativement aux banques de terminologie, est

l'éclatement de la notion de fiche de terminologie. Il évoque la notion de fiche virtuelle constituée, au moment de la production, à la demande des informations contenues dans les banques de terminologie, dans la mesure où il existe déjà des logiciels permettant aux usagers de les extraire selon des combinaisons variables.

M. Louis-Claude Tremblay fait remarquer que les produits de l'activité terminologique sont exploités par un ensemble donné de clients, ce qui donne lieu à un marché potentiel limité. Le développement et l'expérimentation de nouveaux produits doivent s'accompagner d'une analyse du marché afin de vérifier l'existence d'une clientèle pour chacun de ces produits. Il est d'avis qu'à ce propos, les produits terminologiques sont nettement sous-exploités actuellement, alors que le potentiel de ces produits est probablement beaucoup plus grand que ce que l'on croit généralement.

À propos des aspects commerciaux de la terminologie, M. Meunier estime que l'on peut songer aujourd'hui à diffuser auprès des usagers des logiciels leur permettant d'extraire eux-mêmes les données qui les intéressent à un moindre coût, autant pour l'usager que pour le producteur.

Comptes rendus des débats

En ce qui concerne la diffusion des données terminologiques, on s'entend pour dire que la diffusion de la terminologie sur Internet ne marquera pas la fin de la diffusion sur d'autres supports, tel le disque optique compact, compte tenu de l'existence d'un marché réel pour les différents types de supports.

Cependant, M. Gouadec souligne l'intérêt pour les producteurs de banques de terminologie de confier à des tiers la production de produits dérivés, ciblés pour des clientèles particulières soit en fonction des domaines d'intérêt, soit en fonction de différents environnements informatiques. La demande pour ce type de produits dérivés semble être de plus en plus importante et les producteurs de banques telles que la BTQ et Termium ne sauraient ignorer le potentiel d'un tel marché.

M. Meunier

soutient qu'il est inévitable que les supports de diffusion de l'information terminologique continuent d'évoluer au fil du développement des technologies. Cela signifie que les producteurs de terminologie doivent concevoir leurs bases de données de façon telle qu'ils ne seront pas prisonniers de la structuration de l'information, incapables de s'adapter aux nouveaux systèmes et aux nouvelles fonctions de la terminologie.

M^{me} Michaud

souhaite que l'on multiplie les rencontres interdisciplinaires sur la

question des banques de terminologie afin d'élargir et d'approfondir les réflexions sur le développement des outils et des produits en toute connaissance de leurs divers aspects. M. Meunier abonde dans le même sens en signalant notamment l'intérêt pour les terminologues de se rapprocher des spécialistes des sciences de l'information et des sciences cognitives.

Au terme de ces deux journées riches en idées et en discussions, il serait téméraire de prétendre établir une synthèse ou de tirer une conclusion. Il est plus raisonnable de simplement rappeler quelques idées générales pour la suite des choses.

Tout d'abord, j'avais, à l'ouverture, dressé une liste de questions d'actualité sur le vaste sujet des banques de terminologie. Je me rends compte, au terme de nos débats, que cette liste était à la fois trop longue et trop courte.

Trop longue parce qu'en deux jours, le temps nous a manqué – c'était prévisible – pour traiter de façon approfondie l'ensemble de ces sujets, et trop courte, parce qu'au cours de nos délibérations, les questions sont devenues de plus en plus nombreuses, tant les avenues de développement pour la collecte, le traitement et la diffusion des données terminologiques sont nombreuses, compte tenu de l'expérience passée, du développement accéléré des technologies, de l'émergence de nouvelles fonctions pour la terminologie et de l'apparition de nouveaux marchés. Nous avons assisté en fait à une remise à plat complète de notre métier.

On a parlé de réingénierie de l'ensemble du travail terminologique

et des outils de travail à cause des nouvelles exigences de l'environnement. Nous n'avons fait qu'effleurer certains sujets, je pense qu'il est évident qu'on voit la nécessité de poursuivre, d'approfondir l'ensemble de ces sujets.

Il n'y avait pas eu de réunion sur les banques de terminologie depuis une douzaine d'années. Il nous apparaît désormais nécessaire d'entreprendre un cycle de discussions là-dessus et de créer un forum permanent autour de la question des banques de terminologie. Ce forum devrait rassembler les exploitants, les usagers et tous ceux qui travaillent dans la périphérie des banques de terminologies : c'est un sujet que le Rint a inscrit de façon permanente à son programme d'action.

On a pu constater ici la nécessité de mieux faire connaître les outils informatisés pour la collecte, le traitement et la diffusion de la terminologie. Le Rint s'apprête à mettre en route un inventaire permanent des ressources terminologiques, en complément de l'Inventaire permanent des ressources terminologiques qu'il dresse déjà depuis plusieurs années et auquel est venu se greffer récemment un inventaire des banques de terminologie.

Ces journées ont été également l'occasion de poursuivre les réflexions sur la profession terminologique. Il

Clôture de la table ronde

est devenu inévitable de prendre acte des remises en cause de certaines pratiques actuelles de la terminologie et de repenser la profession de terminologue, et peut-être aussi de repenser les fonctions de la terminologie.

Il nous est enfin apparu nécessaire d'étendre aux pays du Sud les technologies terminotiques et les banques de terminologie en créant des outils adaptés aux besoins, aux langues, aux situations linguistiques, à l'environnement technologique et culturel, et peut-être même politique dans certains cas.

J'ai évoqué au cours de ce colloque le besoin urgent de la création, par les instances de la Francophonie, d'un fonds de dotation en équipement informatique pour les pays du Sud. Dans cette perspective, certains des exposés que nous avons entendus nous ont ouvert des perspectives intéressantes. Ainsi, on a parlé de la mise au point d'ordinateurs à 500 \$ (ordinateurs à distance) et d'ordinateurs jetables, et innovations similaires mettant l'équipement informatique de base à la portée de tous. Alors qu'il y a vingt ans, la création d'une banque de terminologie exigeait des moyens considérables que seuls les gouvernements et les grandes organisations pouvaient investir, aujourd'hui, n'importe qui peut créer une banque de terminologie en achetant un ordinateur personnel et l'un des nombreux logiciels de gestion de données terminologiques existants.

Donc, certains progrès technologiques, dont quelques-uns semblent à première vue utopiques, pourraient permettre de réduire

l'écart entre les pays du Nord et du Sud en matière de capacités technologiques, ce qui permettrait de donner rapidement des suites concrètes à l'exposé de M. Marcel Diki-Kidiri.

On a beaucoup parlé d'Internet et d'inforoutes. En fait 1996 sera pour le Rint l'année de son entrée sur Internet, puisque le réseau se dotera d'un site lui permettant de mieux remplir sa mission d'information et de diffusion. Le Rint pourra élargir la diffusion de ses produits, tels les inventaires et entamera l'implantation du système d'échange de données terminologiques Balnéo⁽¹⁾.

*Louis-Jean Rousseau,
Secrétaire général du Rint.*

(1) Voir à ce sujet l'article de Loïc Depecker et André Le Meur dans le numéro 14 de *Terminologies nouvelles*.

Bibliographie de la néologie

BAAYEN (Harald), 1992 : «Quantitative aspects of morphology productivity», dans BOOIJ et van MARLE (éd.), *Yearbook of Morphology 1991*, Dordrecht/Boston/Londres, Kulwer Academic Publishers, p. 109-149, ISSN 0922-3495 ; ISBN 0-7923-1416-6.

Le but de cet article est de mesurer de façon quantitative la productivité morphologique, surtout par néologie suffixale ; les langues de l'étude sont l'anglais et le néerlandais, mais la démarche serait applicable à d'autres langues. La mesure proposée vise à tenir compte des critères suivants : les résultats qui ne doivent pas contredire les intuitions linguistiques ; la possibilité qu'un élément de s'associer à d'autres dans de nouvelles combinaisons ; tenir compte de propriétés particulières qui feraient baisser la mesure ; expliquer pourquoi la productivité ne se mesure pas simplement en terme de fréquences des formes. Deux types de mesures statistiques complémentaires sont proposées, les deux tenant compte de la fréquence des occurrences. Ces critères sont appliqués à un corpus important numérisé. Les résultats obtenus par ces mesures sont confrontés aux théories du lexique mental. Cette étude, qui ne vise pas la langue de spécialité, a déclenché une discussion nourrie, comme en témoigne l'échange suivant :

MARLE (Jaap van), 1992 : «The relationship between morphological productivity and frequency : a comment on Baayen's performance-oriented conception of morphological productivity», dans BOOIJ et van MARLE (éd.), *Yearbook of Morphology 1991*, Dordrecht/Boston/Londres, Kulwer Academic Publishers, p. 151-163.

BAAYEN (Harald), 1993 : «On frequency, transparency and productivity», dans BOOIJ et van MARLE (éd.), *Yearbook of Morphology 1992*, Dordrecht/Boston/Londres, Kulwer Academic Publishers, 1993, p. 181-208.

Descr. : morphologie ; anglais ; fréquence.

CABRÉ (Maria Teresa) et YZAGUIRRE (Lluís), 1995 : «Stratégie pour la détection semi-automatique des néologismes de presse», dans *Traduction, Terminologie, Rédaction Technolectes et Dictionnaires*, vol. VIII, n°2, p. 89-100, ISBN 2-9803329-5-X.

Les auteurs rappellent succinctement les missions de l'OBNEB, (Observatori de Neologia de Barcelona) la méthodologie et les principes théoriques qui gouvernent la recherche de l'OBNEB dans le domaine de la néologie lexicale (néologismes attestés dans les textes standards et/ou utilisés par un grand nombre de sujets). Depuis peu l'OBNEB a franchi une nouvelle

étape dans ses pratiques de repérage des néologismes en passant du dépouillement manuel à la détection semi-automatique des néologismes ; jusqu'à maintenant seul les néologismes formels sont concernés. Les auteurs décrivent le corpus servant de base à leur recherche (7 000 fichiers - extraits de 119 journaux - transformés en fichiers Ascii), présentent le logiciel utilisé pour repérer et traiter les néologismes et ébauchent plusieurs autres axes de recherches qui la complètent et l'enrichissent, en particulier des travaux syntaxiques permettraient de mettre au point un logiciel de détection de la néologie formelle.

Descr. : analyse lexicale ; corpus d'exclusion ; corpus textuel ; logiciel d'analyse ; néologie fonctionnelle ; néologie formelle ; néologie lexicale ; néonymie.

CAETANO MOCHO (Maria do Céu), 1996 : «Contribution à la base de données textuelles : PORTEXT : processus de néologie formelle de la langue courante, en portugais», dans *CUMFID 17*, Projet PORTEXT (numéro spécial), CNRS-INALF, UPRESA - Bases, corpus et langage, p. 105-114.

L'article de Caetano Mocho reprend le contenu de la communication qu'elle a présentée à la réunion scientifique, qui s'est tenue à Nice en novembre 1993, sur le

Bibliographie de la

néologie

projet PORTEXT (constitution d'une base de données textuelles du portugais largement inspirée de FRANTEXT. À partir d'un corpus constitué par des articles de *O Expresso* (48 numéros dépouillés entre le début 1981 et octobre 1993), l'auteur étudie les processus de néologie formelle et plus spécifiquement les néologismes par suffixation en se limitant aux néologismes relatifs à la vulgarisation des termes scientifiques et techniques. Le corpus, saisi par lecture optique, est traité à l'aide du logiciel *Hyperbase*. Ce qui permet à l'auteur de faire le point sur les apports du logiciel, sur les problèmes qui se posent et de faire des suggestions pour rendre *Hyperbase* plus performant pour le traitement des néologismes du portugais.

Descr. : créativité lexicale ; dérivation ; langue courante ; morphologie dérivationnelle ; néologie formelle ; néologie par suffixation ; vitalité lexicale.

CUSIN-BERCHE (Fabienne), 1994 : «Le management en œuvre(s) : discours lexicographiques, terminographiques et technographiques», dans MOIRAND, BOUACHA, BEACCO et COLLINOT (éd.), *Parcours linguistiques de discours spécialisés : Colloque en Sorbonne les 23-24-25 septembre 1992*. Sciences pour la communication 41, Berne, Peter Lang, p. 145-158, ISBN 3-906751-26-7.

Le sort réservé par différents types de dictionnaire au mot *management* fournit le sujet de cet article. L'auteur examine les dictionnaires de langue (générale), techniques, et terminologiques, ce dernier étant le *Dictionnaire des néologismes officiels*. Elle situe leur traitement de la néologie par rapport à leur statut institutionnel.

Descr. : aménagement linguistique ; dictionnaire ; dictionnaire terminologique ; gestion.

CUSIN-BERCHE (Fabienne), 1995 : *Les stratégies dénominatives et désignatives dans le discours managérial : de la néonymie à la néologie*, thèse de doctorat, Université de Paris X Nanterre, 562 p.

S'appuyant sur une analyse d'un corpus de textes produits par la direction d'une grande entreprise en voie de privatisation, F. Cusin-Berche redéfinit le néonyme comme un terme nouveau, qu'elle oppose à l'archéonyme, le terme ancien qu'il remplace. Elle manie le concept de «paradigme désignationnel», mis au point par M.-F. Mortureux, son directeur de recherche, comme principal outil de repérage de la néologie et met en évidence la créativité lexico-discursive, ce va-et-vient entre discours et langue où se situe la néologie. Elle affine les sous-catégories de la néologie, et montre les liens entre la néologie d'emprunt et celle de création. Le choix même du corpus, enfin, le discours d'une entreprise («usine à mots») doit être salué comme innovateur.

Descr. : entreprise ; analyse du discours ; environnement langagier.

GROUD (claudette), SERNA (Nicole), 1996 : *De ABDOM à ZOO. Regards sur la troncation en français contemporain*, Laboratoire de Linguistique Informatique, Didier érudition, coll. «Études de sémantique lexicale», CNRS-INALF, Paris, 160 p., ISBN 2-86460-276-8.

La troncation, procédé de création lexicale déjà ancien, s'est beaucoup développé depuis une dizaine d'années au point que pratiquement tous les secteurs de la vie (sociale, professionnelle, politique, etc.) sont concernés. Cet essor s'accompagne d'un nouveau statut de la troncation, qui à l'origine

ne concernait que le registre familier et d'une évolution des règles phonologiques et/ou morphologiques qui président à sa création. C'est ce que les auteurs se sont attachés à démontrer et à analyser avant le répertoire qui constitue l'essentiel de l'ouvrage. L'expérience que C. Groud et N. Serna ont acquises dans le domaine de la néologie - elles ont conçu la base de données *BORNÉO* (Base ordonnée de néologismes) de l'INALF et l'alimentent depuis plusieurs années - leur a permis de bien faire ressortir, aussi bien d'un point de vue théorique que pratique, ce qui change dans la troncation et de déterminer le sens de cette évolution. D. Guilbaud, de l'Université de Macquarie de Sydney, qui a collaboré à l'ouvrage, confirme le changement qui s'est opéré dans la communauté linguistique vis à vis de la troncation en étudiant son statut dans trois éditions du *P.R.* (1967, 1977, 1993).

Le répertoire recense plus de cinq cent troncats. La vedette est suivie de l'indication de la catégorie grammaticale, de l'unité qui est à l'origine de la ou des forme(s) abrégée(s), éventuellement d'une définition et surtout d'attestations datées qui font de l'ensemble un ouvrage de référence de base.

Descr. : créativité lexicale ; troncation.

HAMELIN (Louis-Edmond), 1995 : «Le québécoisisme *nordicité* : de la néologie à la lexicalisation», dans *Traduction, terminologie, rédaction. technolectes et dictionnaires*, vol. VIII, n° 2, p. 51-65, ISBN 2-9803329-5-x.

L'histoire de *nordicité* par l'auteur du néologisme. Hamelin replace *nordicité* dans l'ensemble de la «chaîne de dérivation» et montre comment il est pris morphologiquement et sémantiquement entre deux aînés (*nord* et *nordique*) dont les évolutions sémantiques vont lui permettre d'apparaître et vont structurer son

sens. Les premières attestations de *nordicité* («État, niveau, fait et conscience de toutes choses concernant les hautes latitudes boréales») datent des années soixante au Canada et en France. Au départ terme d'une langue de spécialité (géographie), il est emprunté par plusieurs disciplines avant de se diffuser largement dans la langue courante, l'enregistrement par les dictionnaires généraux en est un témoignage.

Descr. : création lexicale; langue de spécialité; langue générale; processus de lexicalisation; technolecte.

HERMANS (Ad), 1994 : «Traduction et néologie. Proposition de coopération», dans *Le langage et l'homme*, vol. XXIX, n° 3-4, p. 299-302, ISBN 2-8041-2002-3.

À partir de la distinction entre la néologie primaire et la néologie traductive, l'auteur rappelle certains problèmes qui se posent spécifiquement aux traducteurs de textes scientifiques et techniques contraints de produire dans la langue d'arrivée un texte ayant les mêmes fonctionnalités que le texte dans la langue de départ. Par manque de temps les traducteurs ont parfois recours à des emprunts ou à des périphrases qui ne les satisfont pas. Ils sont demandeurs d'un «guide pratique de la néologie» qui regrouperait les procédés de formation de mots nouveaux (en relation avec le système morphologique de la néologie) et l'ensemble des règles socio-culturelles qui leur permettrait de décider parmi les différents mots candidats à être des néologismes, ceux qui seront le plus facilement acceptés par la communauté linguistique concernée. De son côté la «terminologie» a élaboré une grammaire de la néologie française (matrices lexicogéniques des disciplines scientifiques et techniques et règles socio-culturelles d'acceptabilité des néologismes). Le

CTB propose une initiative de «Néologie traductive» qui, à partir de la création d'un réseau de coopération entre terminologues et traducteurs, étudierait les pratiques néographiques de ces derniers et mettrait à leur disposition une assistance en matière de néographie et leur fournirait, par domaine de spécialité, la liste des néologismes créés par les traducteurs eux-mêmes.

Descr. : néographie; néologie primaire; néologie traductive; néonymie.

MEJRI (Salah), 1995 : *La néologie lexicale*, Publication de la Faculté des lettres de la Manouba, 378 p., ISBN 99736-900-87.

En 1988, S. Mejri a soutenu à la Faculté des Lettres de la Manouba sa thèse de doctorat de troisième cycle qui a été publiée sous le titre : *la néologie lexicale*. Dans la première partie de l'ouvrage la néologie est abordée d'un point de vue théorique. L'auteur délimite avec précision son objet d'étude : dans un premier temps la néologie est replacée dans le cadre des changements linguistiques, puis elle est étudiée sous l'angle des grandes dichotomies linguistiques (synchronie/diachronie; langue/parole; paradigmaticue/syntagmaticue) et par rapport à la problématique du mot. Le travail de clarification conceptuelle qui fait, avec précision, le tour des différentes théories est d'une aide précieuse, surtout dans une période où chaque auteur qui écrit sur la néologie éprouve le besoin de redéfinir ce concept. Après avoir fait l'analyse critique des approches des différentes écoles linguistiques, l'auteur présente de manière argumentée ses options théoriques. Il se situe dans la lignée des travaux de Guilbert mais intègre aussi dans sa réflexion les recherches linguistiques récentes qu'il développe de façon personnelle; il donne, par exemple, une nouvelle dimension aux aspects pragmatiques, psychologiques

et sociaux, éléments essentiels lors de l'étude de la néologie. Cette réflexion théorique aboutit à un classement des différents procédés de formation des néologismes de forme comme des néologismes de sens. Procédés qui concernent tous les aspects de la langue et qui illustrent la complexité du phénomène. L'auteur teste la validité de ses options théoriques par la description d'un corpus composé de dictionnaires - ce qui lui permet de traiter des questions de l'articulation entre les approches théoriques et les pratiques lexicographiques et d'essayer de réduire la contradiction qui existe entre la néologie (production foisonnante et éphémère) et le dictionnaire soumis à des limites justifiées/justifiables objectivement - et par plusieurs numéros de *La banque des mots*. La description systématique qui en résulte permet de mettre en évidence les tendances générales que la langue privilégie pour enrichir le lexique et de mieux cerner l'importance respective de chaque procédé de formation mis en jeu lors de la création de néologismes.

MORTUREUX (Marie-Françoise), 1994 : «Comment peut-on définir la propriété d'un mot?», dans MOIRAND, BOUACHA, BEACCO et COLLINOT (éd.), *Parcours linguistiques de discours spécialisés: Colloque en Sorbonne les 23-24-25 septembre 1992*. Sciences pour la communication 41, Berne, Peter Lang, p. 3-10, ISBN 3-906751-26-7.

En étudiant la reformulation présente dans des textes qui présentent un phénomène nouveau (du moins pour le public visé), l'auteur s'interroge sur ce qui peut constituer «le mot propre». Elle fait valoir que lexicalement certaines désignations peuvent apparaître comme inappropriées, mais sont justifiées discursivement.

Descr. : aménagement linguistique ; vulgarisation ; reformulation.

PETRALLI (Alessio), 1996 : *Neologismi e nuovi media: verso la «globalizzazione multimediale» della comunicazione?* Bologna, CLUEB, 127 p.

Essai sur l'influence des nouveaux médias (courrier électronique, disque optique compact, réseaux électroniques...) sur les langues, l'anglais et l'italien en particulier, et examen des options néologiques exploitées dans les deux cas. Nombreuses indications pratiques sur l'exploitation possible des ressources disponibles pour le dépouillement néologique et comment y accéder. Aperçu de l'état des études de néologie italienne depuis les 20 dernières années ; indications sur les activités de veille néologique en langue italienne.

Descr. : italien ; médias ; veille néologique.

REBOUL (Sandrine), 1994 : «La terminologie télématique: problèmes de reformulation discursive autour du concept de vidéographie», dans MOIRAND, BOUACHA, BEACCO et COLLINOT (éd.), *Parcours linguistiques de discours spécialisés: Colloque en Sorbonne les 23-24-25 septembre 1992*. Sciences pour la communication 41, Berne, Peter Lang, p. 23-32, ISBN 3-906751-26-7.

Le Minitel français est un exemple très connu de vidéographie. S. Reboul étudie d'abord le travail des commissions de terminologie dont la préoccupation majeure a été la motivation des termes, sans oublier les impératifs commerciaux qui déterminent les noms de marques. Elle étudie enfin la diffusion de cette terminologie et son adaptation dans les textes de vulgarisation.

Descr. : aménagement linguistique ; vulgarisation ; médias.

REBOUL (Sandrine), 1994 : *Le vocabulaire de la télématique: du discours au lexique*, Thèse de doctorat, Université Paris X Nanterre, 484 p.

À partir d'une analyse de l'implantation du Minitel en France, l'auteur examine les répercussions linguistiques du lancement du nom de marque *Minitel* et de tout le vocabulaire qui l'accompagne. On retiendra un exemple particulièrement bien documenté d'aménagement linguistique non étatique, l'exploitation des ressources linguistiques, dont la remotivation, les «truncats», ou fractomorphèmes, et les mots-valises, mais surtout l'ancrage des néologismes dans le discours.

Descr. : aménagement linguistique ; vulgarisation ; médias.

SCHAETZEN (Caroline de), 1994 : «Méthodologie des attestations en néologie terminologique», dans *Le langage et l'homme*, vol. XXIX, n°1, p. 55-75, ISBN 2-8041-1975-0.

L'article reprend le contenu de la communication de C. de Schaezen au séminaire méthodologique sur la néologie du Rint qui s'est tenu à Paris (septembre 1993). L'auteur rappelle les activités de recherche du CTB en néographie et en néologie, ainsi que les enseignements dispensés à l'Institut Marie Haps de Bruxelles. C. de Schaezen expose les options méthodologiques et théoriques du CTB qui sont celles de Guilbert adaptées à la terminographie/terminologie et développe certains axes de réflexions et de recherches propre au CTN (causes du rejet ou de l'adoption de néologismes normalisés par une communauté linguistique, position des institutions et des organismes officiels, politique du Rint vis à vis des emprunts anglais. L'article se clos par quelques indications sur les procédures de dépouillement terminographique du CTB.

Descr. : acronyme ; confixation ; emprunt ; néographie ; néologie allogénique ; néologie endogénique ; néologie lexicale ; néologie de forme ; néologie de sens ; néologisme abrégé ; néologisme syntagmatique ; néonyme.

SCHMITT (Christian), 1996 : «Euromorphologie: Perspektiven einer neuen romanistischen Teildisziplin», dans DAHMEN, HOLTUS, KRAMER, METZELTIN, SCHWEICKARD et WINKELMANN (éd.), *Die Bedeutung der romanischen Sprachen im Europe der Zukunft*, Romanisches Kolloquium IX, Tübingen, Gunter Narr Verlag, p. 119-146, ISBN 3-8233-5073-0.

L'étude des internationalismes, spécialité allemande lancée par Peter Braun et ses collègues, trouve une nouvelle recrue en la personne de Christian Schmitt. Son premier corpus est constitué des nouveaux mots introduits dans l'édition de 1988 du *Petit Robert*, qui se révèlent majoritairement des langues de spécialité et issus de formes gréco-latines. Ces néologismes se retrouvent, sous une forme adaptée, non seulement dans les autres langues latines examinées, mais aussi en allemand et en anglais.

Descr. : aménagement linguistique ; internationalisme.

TETET (Christiane), 1995 : «Les sports de montagne. L'alpinisme/l'escalade. II (1976-1989)», *Matériaux pour l'histoire du vocabulaire français. Mots nouveaux contemporains*, Klincksieck, Paris, 393 p., ISBN 2-252-02986-2.

Analyse, selon la présentation habituelle de la collection, de 2 400 créations néologiques du vocabulaire de l'alpinisme et de l'escalade (en milieu naturel et sur structures artificielles) relevées pendant une vingtaine d'années dans la presse spécialisée (*Alpinisme et randonnée*, *Montagne magazine*, *Vertical* - depuis

sa création en 1985, *La Montagne et l'alpinisme* - revue du club alpin français) et qui témoigne bien du foisonnement des créations néologiques du domaine.

VOLZ (Norbert), 1996 :
«CORDON - A joint Venture Case Study», dans RETTIG (éd.),
Proceedings of the first European seminar «Language resources for language technology», Tihany, 15-16 septembre 1995, Budapest, Hungarian Academy of Sciences, p. 159-167.

Description d'un projet de mise au point d'un système d'extraction de néologismes dans un certain nombre de langues européennes, dont le français et l'anglais, en terminologie comme en langue générale. Le projet, qui a commencé en 1995, vise l'extraction automatique de candidats néologismes de corpus déjà constitués de textes balisés. Après une définition opératoire du néologisme, qui tient compte plus particulièrement des néologismes sémantiques, l'auteur décrit les modules du système : le balisage du corpus, la mise au point d'un module de détection de néologismes basé sur les contextes, et un module de détection basé sur le temps et la structure. Afin de réduire le bruit, le système construit des listes de noms propres et de sigles. Le but est d'élaborer un outil qui permet à l'utilisateur, surtout industriel, d'extraire les néologismes d'un corpus tel qu'un disque optique compact.

Descr. : extraction automatique; corpus.

*Chantal Girardin,
John Humbley,
Centre de terminologie et de néologie,
Laboratoire de linguistique
informatique,
Université Paris XIII,
France*

La normalisation

Il a été décidé, lors de la dernière réunion du Rint, en janvier 1996 à Québec, de mettre en place cette nouvelle rubrique qui doit traiter des activités de normalisation. Nous la souhaitons vivante, protéiforme et reflet de l'activité de normalisation terminologique dans les différents modules.

Nous demandons donc à tout un chacun de contribuer avec son expérience propre et nous demandons à chaque responsable de module d'entrer en contact avec la responsable de cette rubrique, Élisabeth Blanchon, pour lui fournir leurs contributions. Ses coordonnées sont données en fin d'article.

1 Introduction

Dans un inventaire à la Pérec, quel point commun peut-on trouver entre un moteur, du papier à lettre et des enveloppes, des grues, des cercueils, des machines à traire, des seringues, des fruits secs, des ventilateurs industriels, des peaux non traitées, du lait, des aiguilles pour machine à coudre, les fluides pour systèmes spatiaux, des préservatifs, des peintures, du beurre, des gilets pare-balles, des pneus, des plastiques, des fixations de ski, des types de coutures ou de points, des lentilles de contact, des tournevis, des cartes bancaires, des lasers, des grains oléagineux, des implants chirurgicaux, des disques optiques et des bouchons d'oreille ? Ils font tous l'objet d'une normalisation, et généralement, d'une normalisation terminologique, tout au moins dans la partie « Définitions » des normes techniques, parfois aussi dans des normes de vocabulaire.

Il n'est peut-être pas inutile de faire le point sur la normalisation, ses acteurs et ses processus, car c'est un monde bien particulier, qui est aussi ignoré du grand public qu'il lui est fondamentalement utile dans au quotidien, dans la mesure où la normalisation influe à notre insu sur

tous les aspects de notre vie quotidienne, de la nourriture aux communications en passant par l'environnement, l'informatique ou les transports, en établissant au passage, pour chacun de ces domaines, une terminologie.

2 Les structures de la normalisation

- L'Iso (Organisation internationale de normalisation), fondée en 1946, travaille dans tous les domaines de la normalisation.
- La CEI (Commission électrotechnique internationale), fondée en 1906, se consacre au domaine électrotechnique.
- L'UIT (Union internationale des télécommunications) s'occupe des télécommunications.

Les membres de ces organisations sont des organisations nationales de normalisation (une par pays). Des liens étroits, appelés liaisons, sont entretenus avec d'autres organisations importantes, groupements de fabricants, d'utilisateurs ou sociétés savantes sous la forme d'échanges de documents et de possibilités de participation aux réunions.

Les travaux sont répartis au sein de comités techniques (ou comités d'études) sectoriels, eux-mêmes

Normalisation

subdivisés en sous-comités et groupes de travail qui réunissent les experts.

Au niveau européen, ces organisations ont des miroirs, sous la forme du :

- CEN (Comité européen de normalisation);
- Cenélec (Comité européen de normalisation électrotechnique);
- Etsi (European Telecommunications Standard Institute).

Ces organismes regroupent les organisations nationales des pays européens.

Au niveau national, on trouve entre autres comme miroir de l'Iso : Afnor (Association française de normalisation) en France ;

- CCN (Conseil canadien des normes);
- IBN (Institut belge de normalisation);
- Snima (Service de normalisation industrielle marocaine);
- SNV (Organisation suisse de normalisation);
- Innorpi (Institut national de la normalisation et de la propriété industrielle).

Chaque organisation nationale dispose idéalement d'un comité miroir de chaque comité technique, réunissant des experts du domaine. Ce comité étudie les documents diffusés et fournit un vote et des commentaires ou des commentaires simples, selon le niveau de document, à chaque étape d'élaboration des documents. Ces commentaires sont traités, intégrés ou rejetés, de manière motivée et argumentée, par les groupes de travail internationaux réunissant des experts de chaque pays.

Ce mode de fonctionnement sous-entend qu'avant d'arriver à la publication d'une norme internationale, il faut un long travail de préparation, d'harmonisation, d'échanges d'informations entre experts et entre pays, de votes nationaux, etc.

Cela se traduit par l'élaboration de toute une gamme de documents

intermédiaires. Chaque norme passe successivement par les étapes suivantes :

- Nouveau sujet de travail (généralement abrégé en NWI pour New Work Item);
- Document de travail, (WD pour Working Draft);
- Projet de comité (CD pour Committee Draft);
- Projet de norme internationale (DIS pour Draft International Standard);
- Norme internationale (IS pour International Standard).

Ces différents documents peuvent eux-mêmes avoir plusieurs versions successives avant de faire l'objet d'un quasi-consensus. Les normes, une fois achevées, sont à nouveau soumises à un vote international tous les cinq ans, pour préciser si elles peuvent être confirmées, annulées ou doivent être révisées.

3 La normalisation terminologique

3.1 Le CT 37

Le CT 37 «Terminologie (Principes et coordination)», fondé en 1946, présidé par la Finlande, avec un secrétariat autrichien (Infoterm, Vienne), est divisé en trois sous-comités (SC), eux mêmes divisés en groupes de travail responsables d'une ou plusieurs normes.

Le CT 37 a élaboré un certain nombre d'outils destinés à guider les terminologues dans leur travail, et qui font le point de la théorie dans le domaine, ou fournissent des outils pratiques :

- *Principes et méthodes de la terminologie* (Iso 704), en cours de révision;
- *Harmonisation internationale des notions et des termes* (Iso 860);
- *Vocabulaire de la terminologie* (Iso 1087-1), en cours de révision;

- *Références bibliographiques pour les travaux de terminologie*, (Iso 12615);
- *Travail terminographique orienté traduction* (Iso 12616-2);
- *Codes de langues* (Iso 639);
- *Symboles lexicographiques et conventions typographiques à utiliser en terminographie* (Iso 1951);
- *Organisation alphabétique des données multilingues* (Iso 12199);
- *Gestion de projet de normalisation de la terminologie*, en projet;
- *Élaboration et la présentation des normes de terminologie*;
- *Répertoire d'éléments de données* (Iso 12620);
- *Vocabulaire - Partie 2: Aides à apporter par les ordinateurs dans les travaux terminologiques* (Iso 1087-2);
- *Aides informatiques dans les travaux terminologiques: Constitution et utilisation de données terminologiques et de corpus textuels* (Iso 12618);
- *Aides informatiques dans les travaux terminologiques - Format d'échange terminologique - Applications SGML (Martif)* (Iso 12200).

3.2 Iso/CEI JTC 1/SC 1

Le JTC1 (*Joint technical Committee 1*) est un comité mixte mis sur pied, en 1986, par l'Iso et la CEI pour normaliser tout ce qui relève du secteur des technologies de l'information. Ses 20 sous-comités normalisent tous les aspects des technologies de l'information, des langages de programmation (SC 22), à la sécurité (SC 27), en passant par le multimédia (SC 29) en passant par la terminologie de l'ensemble du domaine des technologies de l'information, apanage du SC 1, sous-comité à vocation transversale.

Les travaux du SC 1 ont commencé avant même la création du JTC 1, dès 1968, puisque le SC 1 faisait alors partie du TC 97 de l'Iso, incorporé par la suite au JTC 1 (la partie 1 de la norme Iso/CEI 2382 fut publiée pour la première fois en 1970).

Le SC1 élabore une norme terminologique unique, la norme Iso/CEI 2382, bilingue anglais-français, divisée actuellement en plus de 35 parties correspondant à autant de sous-domaines répartis entre quatre groupes de travail et qui vont des «Termes fondamentaux» (partie 1) à la terminologie de l'«Hypermédia et multimédia» (partie 33) en passant par la «Vision artificielle» (partie 30), le «Courrier électronique» (partie 32), ou encore la «Sécurité informatique» (partie 8).

3.3 Le Vocabulaire électrotechnique international (VEI)

La CEI (Commission électrotechnique internationale) fut le premier organisme de normalisation à se soucier de terminologie. *Le Vocabulaire électrotechnique international (VEI)* fut en effet commencé en 1906 et a sans cesse été révisé et complété depuis. Il existe sous la forme d'épais volumes et, désormais, également sous la forme d'une banque de données terminologiques consultable en ligne. Les entrées sont multilingues et contiennent les équivalents dans les langues suivantes: anglais, français, russe, allemand, espagnol, italien, néerlandais, polonais et suédois.

Comment y accéder: sur Internet, via Telnet (ou autre outil similaire): vax36.iec.ch ou [193.134.248.40](tel:193.134.248.40). Nom d'utilisateur IEV, mot de passe IEV.

Publications de 1996:

- Classe 1, Notions générales, Chapitre 111, Physique et chimie, deuxième édition;
- Classe 4, Matériels électriques, Chapitre 411, Machines tournantes, deuxième édition;
- Classe 4, Matériels électriques, Chapitre 481, Piles électriques, Première édition.

3.4 La terminologie des télécommunications

Dans le domaine des télécommunications, l'UIT a également élaboré une très importante terminologie normalisée de référence, disponible sur papier mais aussi en ligne sous la forme de la banque de donnée Termite. Cette base contient quelque 59 000 entrées essentiellement du domaine des télécommunications mais aussi d'autres domaines techniques ou administratifs et financiers se rapportant à la structure et au fonctionnement de l'UIT. Les entrées sont principalement en anglais, français, espagnol et parfois en russe (transcrit), mais un certain nombre d'entre elles peuvent contenir des équivalents en italien, allemand et portugais. L'intégration de l'arabe, du russe et du chinois est également prévue.

Comment y accéder: sur Internet: Serveur TIES WorldWideWeb: <http://www.itu.ch>, choisir *ITU General Secretariat, puis ITU Databases (ITU Global Directory, ITU-T/ITU-R Catalogues...)* puis *ITU Telecommunication Terminology Database (TERMITE/Gopher)*.

3.4 Autres normes Iso

Le TC 22 «Véhicules routiers» vient de publier un amendement à l'Iso 611: 1994 *Véhicules routiers - Freinage des véhicules automobiles et de leurs remorques - Vocabulaire* et de mettre au vote le DIS 6813 *Véhicules routiers - Classification des collisions - Terminologie* (Révision de l'Iso 6813: 1981).

Le TC 23 «Tracteurs et matériels agricoles et forestiers» a mis au vote le DIS 3918 *Installations de traite - Vocabulaire* (révision de la norme de 1977).

Le TC 35 «Peintures et vernis» a mis au vote le DIS 4618-2 *Peintures et vernis - Termes et définitions pour*

produits de peinture - Partie 2: Termes particuliers relatifs aux caractéristiques et aux propriétés des peintures (Révision de l'Iso 4618-1: 1984, de l'Iso 4618-2: 1984 et de l'Iso 4618-3: 1984) et le DIS 4618-3 *Peintures et vernis - Termes et définitions pour produits de peinture - Partie 3: Préparation de surfaces et méthodes d'application* (Révision de l'Iso 4618-1: 1984, de l'Iso 4618-2: 1984 et de l'Iso 4618-3: 1984).

Le TC 38 «Textiles» a mis au vote le DIS 8388 *Étoffes à mailles - types - Vocabulaire*.

Le TC 45 «Élastomères et produits à base d'élastomères» a publié la norme Iso 1382: 1996 *Caoutchouc - Vocabulaire*.

3.5 Sources utiles

L'Iso Bulletin, publié par l'Iso, fournit mensuellement la liste des normes en cours de traitement. Au lecteur de trouver les normes de terminologie.

TermNet News inclut depuis quelques mois un supplément STT (*StandardTerm*), auparavant diffusé indépendamment, qui fournit la liste des normes de terminologie publiées dans le monde entier, en les classant par domaine.

Élisabeth Blanchon,
Centre de terminologie et de néologie,
CNRS - LLI - UMR 195,
Université Paris XIII,
France.

Adresse:
Élisabeth Blanchon,
Centre de terminologie et de néologie,
CNRS - LLI - UMR 195,
Université Paris XIII,
Avenue J.B. Clément,
F-93430 VILLETANEUSE,
FRANCE
Téléphone: + 33 1 49 40 38 66
Télécopie: + 33 1 48 26 30 19
blanchon@dialup.francenet.fr

Publications

Manual of Specialised Lexicography

Après avoir défini les concepts fondamentaux de la lexicographie, les auteurs analysent la question fondamentale pour le lexicographe : qui consultera le dictionnaire technique ? Les utilisateurs ne sont guère faciles à décrire. S'agit-il de l'expert d'un domaine technique, du semi-expert (l'étudiant, par exemple), du profane (le journaliste, le traducteur) ? Quel niveau de connaissances linguistiques ont-ils atteint ? Maîtrisent-ils parfaitement les deux langues d'un dictionnaire bilingue ?

Les réponses à ces questions guideront le lexicographe dans ses stratégies : l'expert n'a pas besoin d'explications de nature technique (encore appelées *connaissances encyclopédiques*), tandis que le profane attend une note explicative. Le lecteur de langue maternelle étrangère espérera des informations grammaticales, voire une aide à la prononciation des termes techniques. Pour les domaines liés à une civilisation en particulier, le droit par exemple, le lexicographe devra également noter le degré d'équivalence des termes. De plus, si le dictionnaire sert non seulement à la lecture-réception, mais encore à la production de textes, il devra être très complet et fournir un maximum de collocations.

Tout auteur de dictionnaire doit sélectionner les entrées. Il peut soit

limiter le domaine couvert et enregistrer un maximum de termes, soit embrasser un ensemble plus vaste et réduire les entrées de chaque sous-ensemble. S'il répertorie le vocabulaire d'un domaine bien délimité du savoir ou de la technique, il sera bien avisé de dessiner un graphe représentant la hiérarchie notionnelle, afin de s'assurer qu'il n'exclut aucun terme important. Choisira-t-il la méthode introspective liée à sa compétence linguistique et technique ? Il s'exposera alors aux risques de choix arbitraires. Les fichiers de citations et de lecture ne dépouillent jamais exhaustivement un domaine et leurs auteurs se focalisent souvent sur des mots exceptionnels davantage que sur les termes les plus généraux (et les plus utiles). Cherchera-t-il ses sources dans des encyclopédies, des dictionnaires, des manuels et des livres de référence ? Il devra alors mentionner ses sources. Enfin, s'il se constitue un corpus indexé et analysé par un programme informatique de concordance, il pourra extraire les termes les plus fréquents, mais le travail préparatoire sera très fastidieux.

Le nombre de langues traitées influencera également la présentation et déterminera l'usage : un dictionnaire unilingue est le plus apte à couvrir les besoins de réception et de production de textes. À l'autre extrême, les dictionnaires multilingues

présentent de nombreux inconvénients : la recherche est plus fastidieuse, les informations nécessairement moins détaillées à volume égal et le niveau d'équivalence n'est pas toujours le même entre les diverses langues.

Une fois les entrées sélectionnées, il faudra les classer. Le système le plus simple, le plus pratique et le plus répandu est le classement alphabétique. Toutefois, un coup d'œil dans les dictionnaires permet de repérer de grandes divergences dans l'ordre alphabétique, selon que les espaces ou les morphèmes (marquant la flexion, le genre et le nombre) sont pris en compte ou ignorés dans le tri des syntagmes. Les auteurs comparent aussi la présentation sous forme de liste et en niche, cette dernière économisant de l'espace grâce à la décomposition des lexèmes en éléments constitutifs.

En bref, ce petit manuel est à la fois clair, agréable à lire et très complet. Il permettra aux débutants de s'initier au métier et aux utilisateurs de juger les dictionnaires. Un seul petit regret toutefois : si les techniques informatiques ne sont pas ignorées, elles ne se voient peut-être pas accorder l'importance qu'elles méritent. Certes les logiciels de saisie de l'information sont trop nombreux pour être analysés et comparés en détail, mais il importe de soupeser leurs avantages et leurs inconvénients

En bref

respectifs et d'évaluer les problèmes liés à leur utilisation (formatage d'écran et de fenêtres, choix des rubriques, opérations de tri, de fusion...) En effet, la lexicographie ne se conçoit plus sans l'ordinateur. Il est vrai que la bibliographie analytique fournie à la fin de chaque chapitre permet à chacun de se perfectionner.

André Dussart,
Section de langue allemande,
Institut supérieur de traducteurs et
interprètes,
Bruxelles.

Bergenholtz (Henning) - Tarp (Sven)
éd., *Manual of Specialised Lexicography,
The Preparation of Specialised Dictionaries*,
Amsterdam/Philadelphie, John Benjamins
Publishing Company, 1995, 254 p.

Les dictionnaires bilingues

À une époque où s'intensifient les échanges entre les communautés linguistiques, les travaux de recherche sur les dictionnaires bilingues et multilingues généraux et spécialisés revêtent une importance toute particulière. Les *Dictionnaires bilingues*, premier ouvrage en français consacré à ce sujet, présente les réflexions inédites d'une douzaine d'universitaires, européens et nord-américains. Dirigé par Henri Béjoint et Philippe Thoiron, de l'Université Lumière Lyon 2, il s'adresse aux linguistes, étudiants et enseignants, aux chercheurs et à tous ceux qui s'intéressent aux dictionnaires.

Contributeurs: Claude Boisson, Teresa Cabré, André Clas, Cristina Gelpi, Valery Grundy, Frank Knowles, Marcel Lemmens, Brigitte Lépinette, Eliott Macklovitch, Carla Morello, Roda Roberts et Thomas Szende.

Béjoint (Henri) et Thoiron, (Philippe),
dir., *Les dictionnaires bilingues*, Louvain-la-Neuve, Aupelf-Uref et Duculot, 1996
(Champs linguistiques — Recueil).

Lèxic de materials de la construcció

Ce lexique contient environ 800 termes en catalan et en espagnol concernant les matériaux utilisés pour construire la structure portante des bâtiments, tels que la pierre, la brique, le bois, le fer et l'acier laminé, le ciment, la chaux, le plâtre, les carreaux, la vitre et les pièces préfabriquées de ciment qui sont d'un usage plus habituel.

Il ne prétend pas être un inventaire exhaustif de la terminologie de la construction mais un outil de consultation et d'aide à la résolution des doutes terminologiques les plus fréquents que rencontrent les professionnels de ce secteur.

TERMCAT, Centre de Terminologia,
Lèxic de materials de la construcció.
Barcelona, Generalitat de Catalunya.
Departament d'Indústria, Comerç i
Turisme, 1996 (La indústria a Catalunya).

Vocabulari d'hidrologia subterrània

Le *Vocabulari d'hidrologia subterrània*, qui présente plus de 2 600 termes en catalan, espagnol et anglais, et un index pour chacune de ces langues, réunit la terminologie la plus usuelle du domaine de l'hydrologie souterraine et des sciences et techniques qui s'y rapportent, telles que la géologie, la physique ou la chimie.

Élaboré par la Fondation internationale d'hydrologie souterraine à l'occasion de la célébration de son 30^e anniversaire, l'ouvrage a compté sur l'appui de Termcat et du département de l'Aménagement du territoire et des Travaux publics de la Generalitat de Catalunya dans la phase finale de son élaboration.

VALVERDE FERREIRO (M.B.),
ANGUITA BAROLOME (F.),
Vocabulari d'hidrologia subterrània.
Barcelona, Fundació Centre Internacional
d'Hidrologia Subterrània, 1996.

Vocabulari de fotografia

Dans la ligne de production de matériaux pour la diffusion de la terminologie catalane, Termcat et le Consortium pour la normalisation linguistique ont élaboré le *Vocabulari de fotografia*. Présenté sous forme d'un tryptique, le vocabulaire contient 61 termes catalans avec définition, regroupés en trois sous-domaines: appareils et accessoires, traitement et techniques de réalisation.

Ce vocabulaire a été élaboré avec la collaboration de plusieurs institutions du secteur de la photographie avec le but de faciliter l'usage du catalan aux professionnels et aux amateurs de la photographie.

CONSORCI PER A LA
NORMALITZACIO LINGUISTICA,
CENTRE DE NORMALITZACIO
LINGÜÍSTICA VALLÈS
OCCIDENTAL 3, *Vocabulari de
fotografia* [Triptic], [Barcelona], Centre
de Normalització Lingüística Vallès
Occidental 3, [1996].

Présentation des tapuscrits

1 Principes généraux

- Les auteurs qui souhaitent publier un article dans *Terminologies nouvelles* sont priés de le proposer au responsable du module dont ils relèvent pour obtenir l'accord du comité de lecture. Les auteurs qui ne relèvent d'aucun module s'adresseront au secrétariat de rédaction.
- Seuls seront publiés les textes répondant aux conventions ici mentionnées.
- Le texte sera original, inédit et rédigé en langue française. Par la suite, il ne pourra pas être publié ailleurs sans l'accord du Rint.
- En fin d'article, on mentionnera en italiques :
 - Prénom (s) et nom de l'auteur,*
 - Département,*
 - Organisme,*
 - Localisation.*
- Le *tapuscrit*, ou manuscrit dactylographié, sera présenté en double interligne sur une feuille de format A4 avec une marge de deux centimètres sur les quatre bords et ne dépassera pas une longueur de 20 pages. Toute illustration est la bienvenue.
- L'auteur proposera obligatoirement un résumé de l'article en 80 mots maximum. Ce résumé servira de chapeau et devra être suivi de l'énoncé de 3 à 6 mots-clés.
- Si cela lui est possible, l'auteur enverra une disquette contenant l'article sous un format Ascii et un format de traitement de texte.

2 Typographie

- Le texte sera présenté en caractères romains. Seuls les titres (ouvrages, colloques, programmes, etc.), les autonymes et les mots étrangers figureront en *italiques* (à défaut, ils seront soulignés). Les caractères **gras** ne serviront qu'à signaler tout premier usage d'un terme spécialisé figurant dans une liste explicative en fin d'article (glossaire, lexique, etc.).
- Pour l'usage général des majuscules, on se conformera aux conseils de Hanse (1994 : 534-536). Les titres d'ouvrages seront toujours écrits avec une majuscule au premier mot cité et aux éventuels noms propres.
- Les guillemets utilisés sont les doubles chevrons « ». Ils encadrent les citations, les traductions et tout premier emploi d'un mot utilisé de manière inhabituelle ou inventé. Si des guillemets sont utilisés à l'intérieur d'une citation, il convient de les remplacer par des guillemets simples « ».

3 Titre et sous-titres

- Le titre de l'article sera concis et attirant et le texte sera organisé de manière à ne pas dépasser deux niveaux de sous-titres, numérotés sous la forme 1 et 1.1 :
 - 1 Politique linguistique
 - 1.1 Les incitants
 - 1.2 Les obstacles
 - 2 Bilan et perspectives

4 Sigles et acronymes

- Les abréviations seront expliquées dans des parenthèses lors de leur premier emploi, à moins que leur signification ne soit supposée connue d'un large public.
- Les noms d'organismes dont l'abréviation est épelée (sigles) seront écrits en majuscules, sans points abrégatifs : BTQ, CEE, DGTSL, OLF, etc. Ceux dont l'abréviation est prononcée comme un mot (acronymes) ne prendront la majuscule qu'à l'initiale et n'auront pas de points abrégatifs : Cifl, Eurodicautom, Rint, etc. Si un choix est possible, l'auteur adoptera la règle qui correspond à sa manière de prononcer l'abréviation : Onu ou ONU, Urss ou URSS, etc.

5 Énumérations

- On évitera tout usage abusif de l'énumération, ce procédé étant réservé à la citation de points relativement brefs. Chaque élément énuméré :
 - Sera précédé d'un tiret ;
 - Commencera par une majuscule ;
 - Se terminera par un point virgule, le dernier élément étant suivi d'un point.

6 Exemples

- Les énoncés utilisés comme exemples dans le texte seront précédés d'un numéro entre crochets. S'ils sont en langue étrangère, ils figureront en italiques.

- [1] *Nagize scandale* (= « escale ») à *Nairobi*.
[2] *Muganga yasanze mfise affection* (= « infection ») *mu ryinyo*.

7 Notes

- Les appels de note se placent entre parenthèses selon une numérotation continue. On regroupera toutes les notes en fin de tapuscrit. Aucune référence bibliographique ne pourra figurer en note (*cf.* bibliographie).

8 Bibliographie

- Dans le corps du texte, on mentionnera uniquement le nom de l'auteur, suivi entre parenthèses de la date d'édition et de la page concernée. Si l'auteur a publié plusieurs ouvrages la même année, on les identifiera par l'ajout d'une lettre :

Comme le fait remarquer Muller (1968a : 149), « L'histoire de la langue peut créer une distinction entre polysémie et homonymie ».

Par *équivalent*, il faut entendre « chacun des termes de langues différentes qui désignent des notions correspondantes. » (Boutin-Quesnel *et alii* 1985 : 20.)

- La bibliographie proprement dite sera placée en fin d'article. Elle sera classée selon l'ordre alphabétique des auteurs, conformément aux exemples figurant ci-après.

Boutin-Quesnel (Rachel), Bélanger (Nicole), Kerpan (Nada) et Rousseau (Louis-Jean), 1985 : *Vocabulaire systématique de la terminologie*, Québec, Les publications du Québec (Les cahiers de l'Office de la langue française).

Hanse (Joseph), 1994 : *Nouveau dictionnaire des difficultés du français moderne*, 3^e édition, Louvain-la-Neuve, De Boeck - Duculot.

Muller (Charles), 1968a : *Initiation à la statistique linguistique*, Paris, Larousse (Langue et langage).

Terminologies nouvelles, 1990a : *Harmonisation des méthodes en terminologie. Actes du séminaire (Talence, juin 1989 – Hull, décembre 1989)*, Bruxelles, ACCT et Communauté française de Belgique, n° 3.

Vernet (Pierre), 1990 : « Problématique de la recherche terminologique en Haïti », dans *Terminologies nouvelles*, juin 1990, n° 3, p. 61-67.

Coordinatrice

Martine Garsou, Ministère de la Culture et des Affaires sociales: Service de la langue française, Communauté française de Belgique.

Secrétaire de rédaction

Marc Van Campenhoudt, Institut supérieur de traducteurs et interprètes de la Communauté française de Belgique (Bruxelles).

Correspondants

AFRIQUE CENTRALE ET DE L'EST: Ntita Nyembwe.

AFRIQUE DE L'OUEST: Cherif Mbodj.

CANADA: Diane Michaud, Travaux publics et Services gouvernementaux Canada — Terminologie et Normalisation.

COMMUNAUTÉ FRANÇAISE DE BELGIQUE: Martine Garsou, Service de la langue française.

FRANCE: John Humbley, Centre national de la recherche scientifique.

HAÏTI: Pierre Vernet, Centre de linguistique appliquée.

MADAGASCAR: Bakoly D. Ramiamana.

MAROC: Saadia Aït Taleb, Institut d'études et de recherches pour l'arabisation.

QUÉBEC: Dominique Dos Ghali, Office de la langue française.

RÉPUBLIQUE CENTRAFRICAINE: Michel-Marie Koyt.

SUISSE: Françoise Parc, Chancellerie fédérale suisse, Services linguistiques centraux, Section de terminologie.

TUNISIE: Zouhaier Marrakchi, Institut national de la normalisation et de la propriété industrielle.

UNION LATINE: Daniel Prado, Union latine — Direction du II^e programme.

Comité de lecture

Giovanni Adamo (Assiterm, Rome)

Saadia Aït Taleb (IERA, Rabat)

Daniel Blampain (Isti, Bruxelles)

Teresa Cabré i Castellví (Iula, Universitat Pompeu Fabra, Barcelone)

Ion Coteanu (Termrom et Académie roumaine, Bucarest)

Loïc Depecker (DGLF, Paris)

Mário Augusto de Quinteiro Vilca (Universidade de Porto — Faculdade de Letras, Porto)

Michel Trousson (Communauté française - Service de la langue française, Bruxelles)

Marcel Diki-Kidiri (Coordination Afrique)

Thierry Fontenelle (CCE - Service de traduction)

John Humbley (CNRS-CTN, Paris)

Alzouma Oumarou Issoufi (Indrap, Niamey)

Ahmed Lakhdar Ghazal (IERA, Rabat)

Gina Mamavi (DGLF, Paris)

Diane Michaud (TPSGC — Bureau de la traduction, Hull (Québec))

Françoise Parc (Chancellerie de la Confédération suisse — Section de terminologie, Berne)

Louis-Jean Rousseau (OLF, Québec)

Clau Solèr (Ligue romanche, Coire)

Michèle Valiquette (TPSGC — Bureau de la traduction, Hull (Québec))

Andrée Vansteelandt (Institut libre Marie Haps, Bruxelles)

Terminologies nouvelles est la revue du Réseau international de néologie et de terminologie (Rint). Le n° 16 de *Terminologies nouvelles* paraîtra en juin 97.

Je soussigné souhaite recevoir gratuitement la revue *Terminologies nouvelles*.

Nom : _____

Entreprise, organisme: _____

Fonction: _____

Adresse: _____

Ce bulletin d'abonnement est à adresser au module dont vous relevez (adresse au dos de la revue)

ISSN: 1015-5716

© Tous droits de traduction de reproduction et d'adaptation réservés pour tous pays.

Édit. resp. :
M. Garsou, 44 Boulevard Léopold II,
1080 Bruxelles, Belgique.

Conception :
Junius, Alternatives théâtrales

Photocomposition et impression :
Édition & Imprimerie

Adresses des organismes membres du Rint

Afrique centrale et de l'Est

Coordination: Centre de linguistique
théorique et appliquée
BP 4956
Kinshasa/Gombe
Zaire.

Afrique de l'Ouest

Coordination: Centre de linguistique
appliquée
Université Cheikh Anta Diop
Dakar — Fann
Sénégal.

Canada

Terminologie et Normalisation
Bureau de la traduction
Travaux publics et
Services gouvernementaux
Portage II, 3^e étage
165, rue Hôtel-de-Ville
Hull (Québec)
K1A 0S5
tél.: 1 (819) 994-5934

Communauté française de Belgique

Ministère de la Culture
et des Affaires sociales
Service de la langue française
44, Bd Léopold II
B-1080 Bruxelles
tél.: 32 (2) 413 22 95

France

Délégation générale à la langue
française
1, rue de la Manutention
F-75116 Paris
tél.: 33 (1) 40 69 12 00

Haïti

Faculté de linguistique
Université d'État d'Haïti
38, Rue Dufort
(Quartier Bois-Verna)
Port-au-Prince
tél.: (509) 45 12 33

Madagascar

Centre des langues de l'Académie
malgache
BP 6217
Antananarivo 101.

Maroc

Institut d'études et de recherches
pour l'arabisation
BP 6216
Rabat — Instituts
tél.: 212 (7) 77 30 05

Québec

Office de la langue française
200, chemin Sainte-Foy,
Québec (Québec)
G1R 5S4
tél.: 1 (418) 634-4144

République centrafricaine

Conseil national d'aménagement
linguistique
BP 888
Bangui.

Suisse

Chancellerie fédérale suisse
Services linguistiques centraux
Section de terminologie
Gurtengasse 2-4, 4^e étage
CH 3003 Berne
tél.: 41 (31) 324 11 49

Tunisie

Innorpi
10bis, rue Ibn el Jazzar
1012 Tunis — Belvédère
tél.: 216 (1) 785 922

Modules associés

Union latine
Bureau de Paris
131, rue du Bac
F-75007 Paris
tél.: 33 (1) 45 49 60 60



Coédité par:
L'Agence de la francophonie
et la Communauté française de Belgique
(Service de la langue française
du ministère de la Culture
et des Affaires sociales
et Commissariat général
aux relations internationales)

Secrétariat du Rint:
Office de la langue française
200, chemin Sainte-Foy,
Québec (Québec)
G1R 5S4 Canada

Le Rint sur Internet:
[http://www.olf.gouv.qc.ca/
techno/pages/f08.html](http://www.olf.gouv.qc.ca/techno/pages/f08.html)